Non-technical summary

This paper demonstrates that cooperation in international environmental negotiations can be explained by preferences for equity. The last two decades have shown an increasing number of negotiations and protocols on subjects of environmental concern such as global warming. Since the impact of a single country on the global pollution level is rather small, each country has an incentive to free-ride on (possible) emission reductions made by the rest of the world. Therefore, the incentive structure of such an emission game is similar to a prisoner's dilemma. The question is how the cooperation that is observed in prisoner's dilemma situations as well as in some international agreements on the abatement of global pollutants can be explained.

In this paper, we rely on a preference structure in which countries are not solely interested in their absolute payoff but also in equity. We first show that within a symmetric N-country prisoner's dilemma in which agents can either cooperate or defect, in addition to the standard non-cooperative equilibrium, partial or even full cooperation can also result as a Nash equilibrium. If not all countries defect, then the fraction of cooperating countries is rather large. Equity preferences, however, cannot improve upon the standard inefficient Nash-equilibrium in an emission game where countries have a continuous choice of the abatement level: Countries behave as if they would unilaterally maximise their payoff. The last part of this paper is dedicated to the study of a two stage game on coalition formation. Here, the presence of equity-interested countries increases the coalition size and the total abatement effort. A stable international environmental agreement with full cooperation, i.e. an efficient outcome, can be reached if all countries' interests in equity are strong enough.

Cooperation in International Environmental Negotiations due to a Preference for Equity

Andreas Lange[a,b] *and Carsten Vogt*[a]

[a]Centre for European Economic Research (ZEW)
P.O.Box 10 34 43
D-68034 Mannheim, Germany
tel. +49 (0)621 / 1235-208
fax +49 (0)621 / 1235-226
e-mail: lange@zew.de, vogt@zew.de

[b]Interdisciplinary Institute for Environmental Economics, University of Heidelberg

IINS Research Paper No. 10

IINS ist eine von der Deutschen Forschungsgemeinschaft finanzierte Forschergruppe.
The research group IINS is financed by the Deutsche Forschungsgemeinschaft (DFG).

Abstract: This paper demonstrates that cooperation in international environmental negotiations can be explained by preferences for equity. Within a N-country prisoner's dilemma in which agents can either cooperate or defect, in addition to the standard non-cooperative equilibrium, cooperation of a large fraction or even of all countries can establish a Nash equilibrium. In an emission game, however, where countries can choose their abatement level continuously, equity preferences cannot improve upon the standard inefficient Nash-equilibrium. Finally, in a two stage game on coalition formation, the presence of equity-interested countries increases the coalition size and leads to efficiency gains. Here, even a stable agreement with full cooperation can be reached.

JEL classification: C7, D63, H41, Q00

Keywords: international environmental negotiations, cooperation, equity preference, coalition formation

1 Introduction

It is well known that international cooperation is needed in order to deal effectively with environmental problems, such as global warming and the depletion of the ozone layer. Since pollution crosses borders and causes negative effects on a global scale, any country that reduces emissions supplies a public good, e.g. the reduction of climate change.

The impact of a single country on the global pollution level, however, is rather small. Hence, when reducing emissions, a country must bear abatement costs but benefits only little, i.e. the incentives for a country to reduce emissions are small. Rather, the country has an incentive to free-ride on (possible) emission reductions made by the rest of the world. Consequently, although it pays if all parties cooperate, each agent unilaterally has an incentive to defect from cooperation. Thus, the incentive structure can be described as a prisoner's dilemma, clearly leading to an undesired outcome. This incentive problem is rather severe, because – contrary to locally bounded pollution – there is no supranational authority that can enforce cooperative behavior.

However, the last two decades have shown an increasing number of negotiations and protocols on subjects of (global) environmental concern. Perhaps most prominent are the Montreal Protocol on Substances that deplete the Ozone Layer, the Basel Convention on the Control of Transboundary Movements of Hazardous Waste, and the U.N. Framework Convention on Climate Change, which established a process leading to the 1998 Kyoto Protocol which specifies targets for the reduction of greenhouse gases. However, it is not absolutely clear whether these agreements go substantially beyond actions that countries would have done in their self-interest anyway. For example, Murdoch and Sandler (1997) argue that the Montreal protocol does not prove cooperative behaviour. The Kyoto protocol has not yet been ratified by most countries, although it specifies substantial emission reduction targets. Rather, there are attempts by some (but not all) countries to water down (renegotiate) the whole agreement. This became evident when countries tried to obtain the management of sinks for CO_2 accredited as an abatement effort at COP6 in The Hague in November 2000. Clearly, whatever the result of the ongoing negotiations will be, any adopted policy must be in the interest of each party involved.

In attempting to explain cooperation and coalition formation, most theoretical models use a two-period structure as introduced by Barrett (1992, 1994). Here, countries must first decide whether or not to join a coalition. In a second step, both the coalition and the remaining countries choose their emission levels non-cooperatively. A coalition is stable if no country wants to join the coalition and no country has an incentive to leave. Simulations by Barrett (1992, 1994, 1997), Carraro and Siniscalco (1993), and Hoel (1992) have shown that – although there is cooperation – the coalition size is rather small. Most recently, Borek and Rutz (2000) have demonstrated that this result is directly caused by the fundamental incentive structure of this game: If one allows for a continuous number of small countries within the coalition, all positive effects of coalition vanish and the non-cooperative outcome emerges.

The question remains how the cooperation that is observed in prisoner's dilemma situations as well as in some international agreements can be explained. In this paper, we rely on a preference structure given by the ERC-theory (Bolton and Ockenfels 2000, 2001).[1] This theory explains most of the behaviour of agents observed in diverse experiments[2] but deviates little from the traditional utility concept. The utility of an agent is not solely based on the absolute payoff but also on the relative payoff compared to the overall payoff to all agents. Given a certain relative payoff share, the utility is strictly increasing in the own absolute payoff of the agent. Given a fixed absolute payoff, the agent is best off when receiving just the equal (fair) share. To both sides of this equal share, i.e. when receiving less or more than the fair amount, utility is lower, even if the absolute payoff does not change.[3]

Although this behaviour was found on an individual basis, it can also be motivated for countries (governments). Essentially, any government must consider the preferences of its voters. A decision not to reduce emissions can probably be "sold" to these voters

[1] ERC hereby stands for *E*quity, *R*eciprocity, and *C*ompetition.

[2] As already noted by Bolton and Ockenfels, this theory can generate cooperation in the standard prisoner's dilemma.

[3] Note that such a preference for equity is self-centred only and is distinct from altruism (Bolton and Ockenfels 2001). A country's utility is determined solely by its own absolute and relative payoff.

more easily the more other countries also refuse to cooperate. If, however, the rest of the world is within a coalition and substantially reduces emissions, it might be hard for a government to explain free-riding, even though this certainly maximises the "payoff". Whereas this argument (qualitatively) requires similar proportional abatement levels, another equity argument is based on the per capita emission level. Specifically developing countries claim that developed countries with high per capita emission levels of greenhouse gases are responsible for global warming; for example the leader of the Indian delegation at the negotiations on the Rio Convention 1991 demanded that, within the negotiations, "the principle of equity should be the touchstone for judging any proposal" (Dagupta 1994: 133).

In this paper, we apply the ERC-theory to the problem of international cooperation in reduction of some global pollutant like the greenhouse gas CO_2. We concentrate on the case of countries, which are identical with respect to both their abatement costs and their benefits from abatement. We first study a symmetric N-country prisoner's dilemma game in which the agents have only two options available – cooperate or defect. We analyse Nash-equilibria when agents' preferences can be described by ERC, i.e. players value both their absolute and their relative payoff. In particular, we look at the number of countries who play cooperatively. We show that non-cooperation is always an equilibrium, since – if no other country cooperates – a country would maximise its absolute payoff and receive the equal share by choosing to defect. Additionally, however, there may be Nash-equilibria in which countries cooperate: If, for example, the rest of the world plays cooperatively, a country can get the equal share by choosing to cooperate as well. Hence, if it values its relative payoff being close to the equal share more than its absolute payoff, it will choose to complete the grand coalition. Clearly, partial cooperation can also occur; some countries cooperate while others defect. For such equilibria, we show that the number of cooperating countries is rather large: Since cooperation leads to a lower absolute payoff, for a country to choose to cooperate, playing cooperatively must move it closer to the equal share than defecting would. As we show, this can only be the case if at least half of the countries cooperate. This result contrasts with the standard result by Barrett and others that the coalition size is rather small.

Note, however, that in the prisoner's dilemma, the countries have only the discrete choice of cooperating or defecting, but with respect to the global warming problem, countries can choose their emission level continuously. We therefore analyse a symmetric emission game. Here, ERC alone cannot improve upon the non-cooperative Nash-equilibrium with standard preferences in which only the absolute payoff matters to a country. The reason is that as long as a country receives less than the equal share and has the possibility to increase its absolute payoff, it will do so. Since this holds for all countries, none agree to get less than the fair share. The unique Nash equilibrium therefore is identical to the symmetric Nash-equilibrium with standard preferences.

This result changes, however, if a stage of coalition formation is introduced. To illustrate this, we study a standard two stage coalition formation game with ERC preferences. Here, in contrast to the traditional models by Barrett (1994), Carraro and Siniscalco (1993), etc., coalitions that involve a rather large fraction of countries can be stabilised. If each country puts enough weight into getting close to the equal payoff share, even the grand coalition can be obtained in equilibrium. In general, the presence of (some) equity-interested countries increases the incentive to join a coalition and, therefore, leads to a larger coalition size in equilibrium. The efficiency gains are maximal if countries that are interested in equity stay outside the coalition and reward a higher abatement effort of the coalition with an expansion of their own abatement level.

The paper is organised in the following way: In section 2, the basic features of the ERC theory are introduced. We then study the discrete version of a symmetric N-country prisoner's dilemma in section 3, whereas the emission game is analysed in section 4. Here, we look at the Nash-equilibria of the one shot game first and then study stable coalitions in the two-stage coalition formation game. The final section – as always – concludes.

2 The preference structure

The analysis in this paper relies on a preference structure in which players do not solely draw utility from their absolute payoff. Rather, along with their own payoff, they are motivated (non-monotonously) by the relative payoff share they receive, i.e. how their standing compares to that of others. With this, we rely on the ERC-model by Bolton and Ockenfels (2000) but use a full information framework. This theory explains the behaviour of people observed in a rather large group of experiments better than the standard theory. As already mentioned, it can also be rationalised for countries (governments) whose citizens have some preference for fair sharing abatement efforts among countries.

Let the (non-negative) payoff to agent i be denoted by y_i, $i, \ldots, N$, the relative share by $\sigma_i = y_i / (\sum_j y_j)$. The utility function is given by:

$$a_i u(y_i) + b_i r(\sigma_i)$$

where $a_i, b_i \geq 0$, $u(\cdot)$ is strictly increasing and concave, and $r(\cdot)$ is concave and has its maximum at $\sigma_i = 1/N$. Throughout this paper we assume that $r(\cdot)$ is symmetric around $1/N$, i.e. $r(1/N - x) = r(1/N + x)$ for all $x > 0$. The types of countries are characterised by the relative weights a_i/b_i.

3 The Prisoner's Dilemma

As already noted, many problems of providing public goods have the basic incentive structure of a prisoner's dilemma. Although every agent (country) prefers the provision of this good, she (it) has an incentive to free-ride on the efforts of the other players and, hence, the public good will not be provided, or at least not in an efficient amount. This, in particular, holds true for any problem of transboundary pollution, e.g. the problem of climate change caused by CO_2 emissions, where reducing emissions can be interpreted as providing such a public good.

In this section we study a simple symmetric N-country prisoner's dilemma (PD game) where each country has two actions available. It can cooperate, “c”, or defect, “d”. In terms of the climate change problem: The country either reduces emissions by a (pre-) specified amount or it doesn't. This situation may occur when countries have to decide whether or not to ratify an international agreement which already specifies the emission levels that must be abated.

3.1 The payoff structure

The total number of cooperating countries is denoted by k. For any given k, the payoff to a country is given by $B(k)$ if the country defects (tries to free-ride). If a country plays cooperatively, it must bear some additional costs $C(k)$. Its payoff is therefore given by $B(k) - C(k)$. We assume decreasing marginal benefits for a country if the number of contributers rises, i.e. $B(\cdot)$ is increasing and concave. Further, the total cost of cooperation, $kC(k)$, increases in k.

In order to generate the standard incentive structure of a PD game, we assume that $B(k+1) - B(k) < C(k+1)$, i.e. playing cooperatively reduces the absolute payoff, given an arbitrary number of “c”-countries. To make more cooperation attractive from both the social and the individual point of view, we make the following assumption:

(i) $NB(k+1) - (k+1)C(k+1) \geq NB(k) - kC(k)$. “socially desirable”

(ii) $B(k+1) - C(k+1) \geq B(k) - C(k)$. “individually desirable”

Further, we assume that payoffs for both cooperating and defecting countries are non-negative for all k.

3.2 The Nash equilibria

In the following, we analyse Nash equilibria in the one shot PD game where countries choose simultaneously. Assume that k countries, aside from country i, play cooperatively. Then country i chooses to play "c" if and only if:

$$
\begin{aligned}
& a_i u(B(k+1) - C(k+1)) + b_i r\left(\frac{B(k+1) - C(k+1)}{NB(k+1) - (k+1)C(k+1)}\right) \\
\geq \;& a_i u(B(k)) + b_i r\left(\frac{B(k)}{NB(k) - kC(k)}\right).
\end{aligned}
$$

This is equivalent to country i playing "c" if

$$
a_i/b_i \leq \delta(k) := \frac{r\left(\frac{B(k+1)-C(k+1)}{NB(k+1)-(k+1)C(k+1)}\right) - r\left(\frac{B(k)}{NB(k)-kC(k)}\right)}{u(B(k)) - u(B(k+1) - C(k+1))}. \tag{1}
$$

In other words, in order to choose "c" the country must be overcompensated for the loss in absolute payment by moving closer to the average payment.

Note that in (1) the denominator of $\delta(k)$ is positive, since playing "d" always maximises the absolute payoff. The sign of the numerator, however, depends on the number k of cooperating countries. For $k = 0$, playing "d" gives the equal share and, hence, maximises the motivation by relative payoff. Thus, $\delta(0)$ is negative and no country will unilaterally play cooperatively, since a_i/b_i is non-negative. For $k = N-1$, however, the numerator is positive, because playing cooperatively and, hence, completing the grand coalition gives the equal payoff share here. Thus, what matters for the choice of action is the type of the country. If it values the absolute payoff relatively high, i.e. a_i/b_i being relatively large, it plays "d". If, however, the relative payoff is given more weight ($a_i/b_i \leq \delta(N-1)$), it chooses to play "c". Thus, all countries playing "c" can establish an equilibrium, provided that all countries' types are smaller than $\delta(N-1)$.

Hence, we can obtain cooperative behaviour for certain types of countries. In general, there may also be equilibria where only a certain fraction of countries plays cooperatively. The general conditions for a Nash equilibrium of this ERC-PD game are given by

$$
a_i/b_i \leq \delta(k^* - 1) \text{ for } k^* \text{ countries (playing "c")}, \tag{2}
$$

$$
a_i/b_i \geq \delta(k^*) \text{ for the remaining } N - k^* \text{ countries (playing "d")}. \tag{3}
$$

Since $\delta(0) < 0$ and $\delta(N-1) > 0$, we immediately obtain the following result:

Proposition 1 (Nash equilibria) *All countries playing "d" is always a Nash equilibrium. Additionally – depending on the types of the countries – there may be equilibria in which some or all countries play "c".*

Two extreme cases may serve to illustrate the proposition: If all agents value only the absolute payoff ($b_i = 0$), in other words, if we are back to standard preferences, it merely results the non-cooperative equilibrium. If, however, agents are solely motivated by their relative payoff share, $a_i = 0$, we have two equilibria – one in which all countries play "d", the other where all countries cooperate. Thus, if we look at the question of whether or not to ratify an already negotiated international agreement, we see that no country would like to be the first if it expects the others not to ratify. If, however, all except of one country have already ratified, then this country may decide to ratify as well.

In the following, we have a closer look at the number of countries that may possibly cooperate in a Nash equilibrium. On the one hand, as long as $\delta(k^* - 1) < 0$, there is no chance of having a coalition of size k^*. Here, $a_i/b_i > \delta(k^* - 1)$ for all types and condition (2) cannot hold for any country. On the other hand, the conditions for a Nash equilibrium given by (2) and (3) immediately imply that if $\delta(k^* - 1) > 0$ then there are types $((a_i/b_i)_{i=1,...,N})$ of countries such that k^* countries cooperate and $N - k^*$ countries free-ride. These types – for example – could be given by $a_i/b_i = \delta(k^* - 1)$ for $i = 1, ...k^*$, and $a_i/b_i = \min\{\delta(k^* - 1), \delta(k^*)\}$ for $i = k^* + 1, ..., N$.

In order to find feasible coalition sizes, we must therefore study conditions in which $\delta(\cdot)$ is positive. Note again that the denominator of (1) is positive. The crucial point is therefore whether or not the numerator is positive. Remember that we assumed that the motivation drawn from the relative payoff, $r(\cdot)$, is symmetric around (its maximum in) $1/N$. Therefore, we have $\delta(k) > 0$ if and only if, by choosing "d", a country further deviates from the equal share $(1/N)$ than by playing "c", i.e.

$$\frac{B(k)}{NB(k) - kC(k)} - \frac{1}{N} > \frac{1}{N} - \frac{B(k+1) - C(k+1)}{NB(k+1) - (k+1)C(k+1)}.$$

Straightforward calculus shows that this is equivalent to

$$0 < B(k+1)C(k)Nk + B(k)C(k+1)N[k+1-N] + C(k)C(k+1)[Nk - 2k(k+1)]$$

or

$$\begin{aligned}
0 &< B(k+1)C(k)\frac{N}{k+1} + B(k)C(k+1)N\frac{k+1-N}{k(k+1)} + C(k)C(k+1)\left(\frac{N}{k+1} - 2\right) \\
&= \left(B(k+1)C(k)\frac{N}{k+1} - B(k)C(k+1)\frac{N}{k}\right) \\
&\qquad + \left(\frac{NB(k)}{k} - C(k)\right)C(k+1)\left(2 - \frac{N}{k+1}\right).
\end{aligned} \tag{4}$$

We can use this inequality to study the number k^* of agents that play cooperatively in equilibrium. First, note that we assumed payoffs to be non-negative and therefore $NB(k) - kC(k) > 0$. Thus, the second summand is negative for $k < N/2 - 1$. For payoff functions that satisfy the requirement that total costs proportionally increase more than the total benefits, i.e.

$$\frac{(k+1)C(k+1)}{kC(k)} > \frac{NB(k+1)}{NB(k)}, \tag{5}$$

the first bracket in (4) is negative as well. As a consequence, inequality (4) cannot hold and $\delta(k) < 0$ for $k < N/2 - 1$. Thus, for any given vector of types, if a country plays "c" in equiibrium, then, in total, at least half of the countries cooperate. For PD games where inequality (5) does not hold, we obtain $\delta(N/2 - 1) > 0$, suggesting that the minimal coalition size can be smaller than $N/2$. As long as the number of countries exceeds 8, however, this is not the case:

Proposition 2 (minimal coalition size) *For any given payoff structure of the PD game, there is a lower bound of the coalition size that cannot be undercut: If any country plays cooperatively, then – at least –* $N/2$ *countries cooperate if* $N \geq 8$ *or total costs of cooperation proportionally increase more than the benefits, i.e. inequality (5) holds.*

This proposition shows that if there is a coalition of cooperating countries, then it is rather large. The intuition is that in order to make cooperation attractive for a country, it has to be closer to the equal payoff share than when playing “d”. This can only the case if the number of countries that receive the (smaller) cooperative payoff is already large. This result contrasts with the standard results from a coalition game á la Barrett (1994), in which the coalition size is small.[4]

In order to prove the proposition, we derive an upper bound for $\delta(k)$ which is negative for $k \leq N/2-2$. Therefore, $\delta(k^*-1) > 0$ can only be the case if $k^* \geq N/2$. The proof is given in the appendix.[5]

The following example illustrates proposition 2.

Example 1 *Let $B(k) = km$, $C(k) = c$, where $c > m$. Then, using inequality (4), $\delta(k-1) > 0$ if and only if $[Nm - c]c[2 - N/k] > 0$. The minimal coalition size in case of cooperation, therefore, equals $N/2$. It arises as an equilibrium, when exactly $N/2$ countries have type $a/b = 0$, and the others value the absolute payment positively ($a/b > \delta(N/2)$). Figure 1 illustrates the function $\delta(k)$ for the special case in which $r(\sigma) = -(\sigma - 1/N)^2/2$, $u(y) = y$, $m = 1$, $c = 1.5$, and $N = 10$. Here, $\delta(N-1) = 3.1 \cdot 10^{-4}$. Therefore, if countries weight the utility from the relative payoff share by more than 3225 times the absolute payoff ($3225a_i < b_i$), the grand coalition would result as an equilibrium.*

Note that proposition 2 is based on the assumption that the valuation of relative payoff is symmetric around the equal share. If, however, agents value a downward deviation less than an upward bias from the equal share, i.e. $r(1/N - x) < r(1/N + x)$ for $x > 0$, the minimal coalition size is even larger, and $N/2$ is still a lower bound. If the valuation is skewed to the left, then the minimal coalition size shrinks.

[4]Note, however, that in the coalition model by Barrett, cooperation was generated by a two stage structure of the game: Countries first decide whether or not to join a coalition and then, in a second step, choose the emission level. We study the equilibria based on ERC preferences in such a game in section 4.2.

[5]Note that although the derived upper bound is negative only for $N \geq 8$, we have not been able to construct examples for a smaller number of agents where the proposition does not hold.

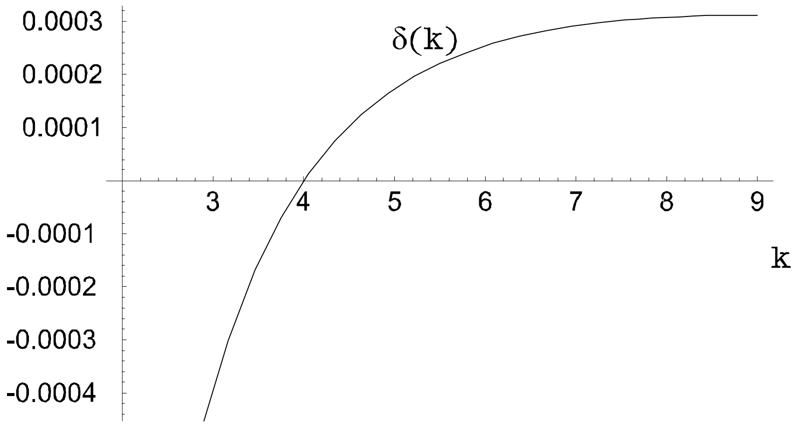

Figure 1: *In order to obtain k^* as the equilibrium number of cooperating countries, the ERC-type a_i/b_i must be below the line at $k = k^* - 1$ for k^* and above the line at $k = k^*$ for the remaining $N - k^*$ countries.*

4 The emission game

In the previous section, we assumed that countries have only the discrete choice whether or not to cooperate. Now, we look at an emission game where countries can choose their emission levels continuously. It will become obvious that, here, ERC alone cannot improve upon the non-cooperative Nash-equilibrium with standard preferences where only the absolute payoff matters. However, introducing more structure to the game, i.e. if countries play a coalition game as in Barrett (1994), ERC may yield a rather large coalition size or even support the grand coalition.

Let the number of countries again be denoted by N. Each country must choose its abatement level q_i ($i = 1, \ldots, N$). Abatement induces some costs $C(q_i)$ that are assumed to be increasing and convex in the abatement level ($C'(\cdot) > 0$, $C''(\cdot) > 0$). Reducing emissions creates some benefits $B(Q)$ in terms of abated damage from climate change, where $Q = \sum_i q_i$ denotes the aggregate abatement level. Benefits from abatement are increasing and concave. The payoff to a country is therefore determined by $B(Q) - C(q_i)$.

4.1 Nash equilibria in the one shot emission game

We again analyse Nash equilibria when countries act simultaneously. Taking q_j $(j \neq i)$ as given, country i chooses q_i to maximise:

$$a_i u(y_i) + b_i r(\sigma_i)$$

where $y_i = B(\sum_{j \neq i} q_j + q_i) - C(q_i)$ denotes the absolute, and $\sigma_i = y_i / \sum_j y_j$ denotes the relative payoff to country i. By choosing q_i, a country directly influences its own abatement costs and the benefits from abatement. It thereby also has an impact on the payoff to the rest of the world, which enters its own utility through the relative payoff. The first order condition is therefore given by:

$$0 = \left[a_i u'(\cdot) + b_i r'(\cdot) \frac{\sum_{j \neq i} y_j}{[\sum_j y_j]^2} \right] [B'(Q) - C'(q_i)] - b_i r'(\cdot) \frac{y_i}{[\sum_j y_j]^2} (N-1) B'(Q) \tag{6}$$

$$= a_i u'(\cdot)[B'(Q) - C'(q_i)] + b_i r'(\cdot) \left[\frac{\sum_j (y_j - y_i)}{[\sum_j y_j]^2} B'(Q) - \frac{\sum_{j \neq i} y_j}{[\sum_j y_j]^2} C'(q_i) \right]$$

$$= a_i u'(\cdot)[B'(Q) - C'(q_i)] + b_i r'(\cdot) \left[\frac{1 - N\sigma_i}{\sum_j y_j} B'(Q) - \frac{\sum_{j \neq i} \sigma_j}{\sum_j y_j} C'(q_i) \right]. \tag{7}$$

The reaction of country i to a given abatement policy for the rest of the world can be calculated from this first order condition. Let us first study the two extreme cases, $a_i = 0$ and $b_i = 0$, respectively. For $b_i = 0$, i.e. an absolute payoff maximiser, the first order condition reduces to $B'(Q) - C'(q_i) = 0$. For $a_i = 0$, the country is solely interested in getting the equal payoff share. Hence it would choose q_i to satisfy $NC(q_i) = \sum_j C(q_j)$. For $a_i, b_i \neq 0$, the chosen abatement level is between the levels for those extreme cases.

In a Nash equilibrium, the first order condition must be satisfied for all countries simultaneously. Since $r'(1/N) = 0$, it follows immediately that for all types a_i/b_i, $i = 1, \ldots, N$ there is a symmetric equilibrium where all countries choose the same abatement level, i.e. $\sigma_i = 1/N$ for all i. Here, the resulting abatement level q^* is given by the first order condition $B'(Nq^*) - C'(q^*) = 0$. It corresponds to the Nash equilibrium when agents are only interested in their absolute payoff and therefore resembles the Nash equilibrium in the PD game where all countries defect.

This equilibrium is the only one assuming that for at least one country a_i is greater than 0. To see this, assume that there is an asymmetric equilibrium, i.e. some countries receive less, others more than the equal share. In this case, on the one hand, $\sigma_i < 1/N$ implies that $r'(\sigma_i) > 0$, and hence from equation (6), we obtain $B'(Q) - C'(q_i) > 0$. On the other hand, for $\sigma_i > 1/N$ we have $r'(\sigma_i) < 0$, and therefore equation (7) implies $B'(Q) - C'(q_i) < 0$. These two inequalities, however, would imply that a country which takes more than the equal share has larger marginal abatement costs than countries that receive less. This, however, implies that the abatement cost of countries that get more than the equal share are larger, which clearly contradicts the assumed payoff distribution. Hence, only symmetric equilibria exist. Here, if $a_i > 0$ for at least one country, we get $B'(Nq) - C'(q) = 0$ from equation (6). Only in the extreme (unlikely) case that all countries are solely interested in equal payoff could any arbitrary abatement level be implemented as a symmetric equilibrium.

We can summarise the result in the following proposition:

Proposition 3 (emission game) *In the emission game for ERC preferences, the equilibrium is given by $B'(Nq*) = C'(q*)$. It is unique as long as at least one country draws utility from its absolute payoff ($a_i > 0$).*

Introducing ERC preferences, therefore, does not increase the abatement effort chosen by the countries. It does not even change the equilibrium emission levels. In contrast to the (discrete) prisoner's dilemma, ERC does not add any equilibria in which there is more abatement effort (cooperation). The existence of equilibria in the PD game that mimicks cooperative behaviour, therefore, only arises in the presence of discrete action sets. Having a continuous decision variable, ERC does not change the set of equilibria. Apparently, the reason is that ERC does not establish a preference for being cooperative, but for being similar to other countries with respect to the payoff.

4.2 Coalition formation in the emission game

Whereas in the last section we analysed the Nash equilibria in the one shot emission game, we now study a two-stage game of international negotiations as developed by Barrett (1994). Let us again assume that all countries are identical with respect to their payoff function. In a first stage, countries decide whether or not to join the coalition. Here, each country takes the decisions of the other countries as given. Each country anticipates, however, that the abatement levels, which are chosen in the second stage, depend on whether it does or does not enter the coalition. In stage 2, countries inside and outside the coalition simultaneously select their abatement levels. The coalition thereby maximises its collective benefits but plays Nash against the non-signatory countries which simultaneously maximise their individual utility.[6]

We first study the case of countries that have identical ERC-types. We demonstrate that within the coalition formation game, ERC-preferences can enforce cooperation and even result in the grand coalition. In the second step, we then look at the case of heterogenous ERC-types. By studying the extreme scenario of countries that are solely interested either in their absolute payoff or in equity, we will explore the effects of the existence of some equity-oriented countries.

4.2.1 Cooperation of identical ERC-types

We solve the coalition formation game backwards. That is, for any coalition size k, we first study the first order conditions for the choice of the abatement level inside and outside the coalition. Then, in the second step, the equilibrium coalition size is determined by a stability condition. This means that in the equilibrium, k must satisfy the condition that there is an incentive to neither leave nor join the coalition. For standard preferences (here resulting as the special case in which b equals 0), the traditional literature shows that the coalition size is rather small. For ERC preferences, however, the number of countries within a coalition can be much higher in equilibrium.

[6]Note that, in contrast to Barrett (1994), we assume that the coalition does not behave as a Stackelberg leader. Rather, it takes the abatement levels of the rest of the world as given, i.e. it plays Nash.

Instead of solving the game in general, we will show that if countries only value the relative payoff high enough, i.e. a/b is below a certain bound, then even the grand coalition can be stable.

For countries outside the coalition, the first order condition is again given by (6), whereas the coalition S maximises the utility of a representative member by choosing the abatement policy for all its members, $i \in S$. It is plain that all countries within the coalition S must have the same abatement level, q^S, since all countries are assumed to be of the same type. Hence, the absolute ($y^S = B(Q) - C(q^S)$) and the relative payoffs ($\sigma^S = y^S/(ky^S + \sum_{j\notin S} y_j)$) are equal for all members of the coalition. This leads to the following condition:

$$
\begin{aligned}
0 &= [au'(\cdot) + br'(\cdot)] \frac{\sum_{j\notin S}\sigma_j}{\sum_j y_j}[kB'(Q) - C'(q^S)] - br'(\cdot)\frac{\sigma^S}{\sum_j y_j}(N-k)kB'(Q) && (8)\\
&= au'(\cdot)[kB'(Q) - C'(q^S)] + br'(\cdot)\left[-\frac{\sum_{j\notin S}\sigma_j}{\sum_j y_j}C'(q^S) + kB'(Q)\frac{1 - N\sigma^S}{\sum_j y_j}\right]. && (9)
\end{aligned}
$$

From the last section we already know that if $\sigma_j < (>) 1/N$ for $j \notin S$, then $B'(Q) > (<) C'(q_j)$. Analogously, for the coalition, we obtain from (8) and (9) that if $\sigma^S < (>) 1/N$ then $kB'(Q) > (<) C'(q^S)$. Since $B'(Q) < kB'(Q)$, the first order conditions therefore imply that for countries within a coalition $\sigma^S < 1/N$ and, thus,

$$kB'(Q) \geq C'(q^S).$$

This inequality can now be used to show the following proposition:

Proposition 4 (coalition game) *In the symmetric coalition game for identical ERC preferences (type a/b), the grand coalition is stable if a/b is sufficiently small, i.e. countries are interested enough in being close to the equal share.*

Note first, that within the grand coalition, the emission level satisfies the condition $NB'(Nq^*) = C'(q^*)$, independently of the ERC-types. Here, countries clearly receive the equal share. If country i leaves the coalition ($k = N - 1$), then from the first order conditions we obtain:

$$(N-1)B'((N-1)q^S + q_i) \geq C'(q^S) \geq C'(q_i) \geq B'((N-1)q^S + q_i).$$

Let us now look at the abatement levels that would result if the ERC-type a/b goes to zero. In this case, countries get more and more interested in getting their equal share, and their abatement levels will converge. Hence, in the limit $q = q^S = q_i$, but still $(N-1)B'(Nq) \geq C'(q)$. Hence, in the limit the absolute payoff of a country leaving the coalition is smaller than within the grand coalition, whereas the relative payoff is the same. Therefore, as long as a/b is small enough, the absolute payoff remains lower, and – due to the asymmetric payoff share – the utility derived from the relative payoff is also smaller than in the grand coalition. Thus, no country has an incentive to leave the grand coalition if a/b is small enough.

4.2.2 Coalition of heterogenous ERC-types

Apparently, if one allows for heterogenous ERC-types, starting in the grand coalition, countries that have the largest a_i/b_i will have the greatest interest to leave the coalition in order to obtain a larger absolute payoff. In the following we concentrate on the extreme case in which countries are either interested in their absolute payoff ($b_i = 0$) or in equity ($a_i = 0$). The former are referred to as A-countries, the latter as B-countries. In total, there are N_a A-countries and N_b B-countries; k_a of these A-countries and k_b B-countries form the coalition. The abatement levels are denoted by q_{as}, q_{bs} for signatory countries, q_{an} and q_{bn} for countries outside the coalition.

Let us first look at the behaviour of B-countries. Outside the coalition, any B-country can arrive at the equal share by choosing the average abatement cost level. Thus,

$$C(q_{bn}) = \frac{1}{N_a + k_b}\left[k_a C(q_{as}) + k_b C(q_{bs}) + (N_a - k_a)C(q_{an})\right].$$

Therefore, for a B-country inside a coalition to have no incentive to leave, it also must receive the equal share:

$$C(q_{bs}) = \frac{1}{N - k_b}\left[k_a C(q_{as}) + (N_b - k_b)C(q_{bn}) + (N_a - k_a)C(q_{an})\right]. \tag{10}$$

Thus, in any equilibrium, all B-countries choose the same abatemenent level, $q_b := q_{bn} = q_{bs}$, and receive the equal share:

$$C(q_b) = \frac{1}{N_a}\left[k_a C(q_{as}) + (N_a - k_a)C(q_{an})\right].$$

k_a	1	2	3	4	5	6
π^{an}	0.113	0.1322*	0.170	0.227	0.301	0.390
π^{as}	0.113	0.118*	0.1318	0.155	0.188	0.229
Q	0.012	0.014*	0.018	0.023	0.031	0.040
k_a	7	8	9	10	11	12
π^{an}	0.495	0.612	0.741	0.879	1.024	-
π^{as}	0.279	0.336	0.400	0.471	0.548	0.629
Q	0.051	0.063	0.077	0.092	0.109	0.126

Table 1: *Payoffs to non-cooperating (cooperating) countries, π^{an} (π^{as}), and aggregate abatement, Q, as a function of k_a ($N_a = 12$, $N_b = 0$). The unique equilibrium is marked with *.*

A-countries outside the coalition again maximise their absolute payoff, $B(Q) - C(q_{an})$. The first order condition is given by

$$B'(Q) = C'(q_{an}), \tag{11}$$

whereas the coalition maximises the utility of a representative A-type-member by guaranteeing its B-members the equal share, i.e. equation (10). The first order condition for choosing q_{as} is therefore given by:

$$\begin{aligned} 0 &= B'(Q)\left(k_a + k_b \frac{\partial q_{bs}}{\partial q_{as}}\right) - C'(q_{as}) & (12)\\ &= B'(Q)k_a\left(1 + \frac{k_b}{N - k_b}\frac{C'(q_{as})}{C'(q_{bs})}\right) - C'(q_{as}). & (13)\end{aligned}$$

By construction, for any given k_a and k_b, every B-country is indifferent to being either inside or outside the coalition. For a coalition to be stable, an A-country must not have an incentive to join or leave the coalition. In general, for any k_b there will be a certain number of A-countries, k_a, that will join the coalition. We have multiple equilibria.

The properties of the equilibria are best understood by considering an example. Consider the following quadratic payoff functions: $B(Q) = 10(Q - Q^2/2)$ and $C(q) = 5000q^2$.

Table 1 presents the simulation results for a total of 12 countries when all are of type

$k_a \rightarrow$ $k_b \downarrow$		0	1	2	3	4	5	6
0	π^{an}	0.113	0.113	0.156	0.240	0.354	0.490	-
	π^{as}	-	0.113	0.142	0.202	0.284	0.382	0.492*
	Q	0.012	0.012	0.016	0.025	0.037	0.051	0.067*
1	π^{an}	0.113	0.115	0.168	0.264	0.388	0.534*	-
	π^{as}	-	0.114	0.148	0.215	0.302	0.404*	0.518
	Q	0.012	0.012	0.017	0.027	0.040	0.055*	0.073
2	π^{an}	0.113	0.118	0.185	0.293	0.430*	0.587	-
	π^{as}	-	0.115	0.157	0.230	0.321*	0.427	0.544
	Q	0.012	0.012	0.019	0.030	0.044*	0.061	0.080
3	π^{an}	0.113	0.122	0.206	0.330	0.480*	0.650	-
	π^{as}	-	0.117	0.166	0.245	0.341*	0.451	0.571
	Q	0.012	0.013	0.021	0.034	0.050*	0.068	0.088
4	π^{an}	0.113	0.130	0.233	0.375*	0.540	0.726	-
	π^{as}	-	0.120	0.176	0.260*	0.361	0.475	0.597
	Q	0.012	0.014	0.024	0.039*	0.056	0.076	0.097
5	π^{an}	0.113	0.142	0.269	0.432*	0.618	0.821	-
	π^{as}	-	0.122	0.184	0.273*	0.378	0.494	0.619
	Q	0.012	0.015	0.028	0.045*	0.064	0.086	0.110
6	π^{an}	0.113	0.161	0.318*	0.507	0.718	0.943	-
	π^{as}	-	0.124	0.188*	0.279	0.385	0.503	0.629
	Q	0.012	0.017	0.033*	0.053	0.075	0.100	0.126

Table 2: *Payoffs to non-cooperating (cooperating) countries, π^{an} (π^{as}), and aggregate abatement, Q, as a function of k_a and k_b ($N_a = 6$, $N_b = 6$). The equilibria are marked with *.*

A, i.e. interested solely in their absolute payoff ($N_a = 12$). Here, in equilibrium, two countries cooperate. If, however, half of the 12 countries are interested in receiving the equal share ($N_a = N_b = 6$), the coalition size in equilibrium becomes larger. Table 2 presents both the total abatement level and the payoffs to cooperating and non-cooperating A-countries as a function of the number of A- and B-countries inside the coalition. Here, if no B-country enters the coalition, all A-countries cooperate. If all B-countries are within the coalition, then only two A-countries enter. In general, when B-countries join the coalition, they successively drive out A-countries. This can be illustrated by the equilibria in Table 2: The first, second, fourth, and sixth entering B-country reduces the number of A-countries in the coalition and thereby lowers the payoffs to both, signatory and non-signatory countries. The payoffs to all countries increase, however, when a third or a fifth B-country enters the coalition and the number of cooperating A-countries stays at 4 and 3, respectively. In all equilibria, however, the payoffs and the aggregate abatement level are larger than in the case where $N_a = 12$, i.e. when all countries are solely interested in their absolute payoff. Therefore, in this example, the presence of equity-interested countries leads to efficiency gains which are largest if all B-countries stay outside the coalition.

The results of the simulations – which appear to be robust to variations of the payoff functions and the total number of countries – can be summarised as follows:

Result 5 *The larger the total number of equity-oriented countries (N_b) is, the higher the incentives are for A-countries to join the coalition. In other words, for a given k_b, the number of cooperating A-countries k_a increases in N_b.*

Result 6 *The more B-countries join the coalition, the smaller the incentives are for A-countries to do so. In other words, in equilibrium, k_b and k_a are negatively correlated.*

Result 7 *The total abatement level increases with the number of B-types outside the coalition. A joining B-country improves the payoffs only if it does not drive out an A-country.*

The rationale of results 5 and 6 is the following: If an A-country enters the coalition and the coalition increases its abatement efforts, B-countries outside the coalition increase their abatement activities as well and thereby additionally reward the entering country. If the number of such equity-oriented B-countries outside the coalition gets larger, this external reward for joining a coalition increases and, therefore, the equilibrium coalition size increases. Analogously, if B-countries join the coalition, less countries outside the coalition reward the entering A-country by an increase of their abatement activities. Hence, the incentives for A-countries to enter the coalition decrease and the number of A-countries that are inside the coalition in equilibrium gets smaller. Result 7 reflects the fact that, in general, the more countries cooperate, the higher the efficiency gains are and the closer the aggregate abatement level is to the efficient one. The impact of A- and B-countries on the decision of the coalition, however, differs in the following way: A joining A-country is interested in the absolute payoff and, consequently, the re-optimising coalition increases its abatement effort because the positive effect on one more country is now taken into account. A joining B-country, however, is not primarily interested in the absolute payoff, but in the equal share. Therefore, the coalition will not increase the total abatement level that much because the B-country refrains from deviating from the abatement level of non-cooperating countries. Consequently, the efficiency gains are larger if an A-country enters the coalition than if a B-country joins. Therefore, B-countries are welcome inside a coalition only if their entering does not drive out an A-country.

5 Conclusion

In this paper, we studied two different games that model basic features of international cooperation to provide a public good such as environmental quality. Instead of countries that are solely interested in their absolute payoff, we studied equilibria based on ERC-preferences. This theory explains the behaviour of people observed in experiments by assuming that a part of utility depends (non-monotonously) on the received relative payoff share.

We demonstrated that for a discrete prisoner's dilemma, cooperation can result. If it does, then the fraction of cooperating countries is rather large (above 50 per cent). This result, however, cannot be confirmed for the one-shot emission game, which better describes the action set with respect to environmental problems like global change. Here, in fact, ERC does not change the equilibrium at all. Countries (in equilibrium) still behave as if they were exclusively maximising their own payoff.

We then studied the process of international environmental negotiations in the standard two stage coalition formation game. If countries' preferences can properly be described by ERC, the coalition game can generate cooperative behaviour. The result of the traditional literature on coalition formation that the coalition size is rather small therefore differs substantially: Even the grand coalition can be stable if all countries put enough weight on getting close to the equal payoff share. In general, the presence of countries that are highly motivated through obtaining the equal share leads to efficiency gains. In particular, for countries that are interested exclusively in their absolute payoff, the incentive to join a coalition increases with the number of equity-oriented countries that stay outside the coalition, but it becomes smaller when those countries enter the coalition. Thus, by entering a coalition, equity-oriented countries may drive out countries that are motivated by their own absolute payoff and thereby worsen the payoffs for all countries. Aggregate abatement and, therefore, efficiency gains are maximal if the countries that are interested in getting their equal share stay outside the coalition.

In conclusion, observing international environmental agreements that are signed by a large number of countries may indicate that (at least some) countries are not solely interested in their absolute payoff, and their behaviour could be explained by the ERC-theory.

Note, however, that, in this paper, we exclusively analysed the case of countries which are symmetric with respect to their payoffs. In general, countries clearly are heterogenous (as, for example, the U.S. and the European Union, and some developing countries are with respect to their benefit and costs of abatement of CO_2), and a preference for equity would probably not regard an equal payoff structure. Instead, equity might be preferred with respect to the abatement targets as percentages of the business-asusual-emissions or in terms of CO_2 emissions per capita. The implications of such a preference for equity in the case of heterogenous countries remain subject to further research.

6 Appendix

Proof of proposition 2:

We have to show that $\delta(k) < 0$ for $k \leq N/2 - 2$. Using (4), this is equivalent to

$$\left(\frac{B(k+1)}{B(k)} - \frac{(k+1)C(k+1)}{kC(k)}\right) + (k+1)C(k+1)\left(\frac{1}{kC(k)} - \frac{1}{NB(k)}\right)\left(2 - \frac{N}{k+1}\right) < 0.$$

From the monotonicity and concavity of $B(\cdot)$ it follows that $B(k+1)/(k+1) < B(k)/k$. Further, total cost of cooperation $kC(k)$ increase in k. Therefore,

$$\frac{B(k+1)}{B(k)} - \frac{(k+1)C(k+1)}{kC(k)} \leq \frac{k+1}{k} - 1 = \frac{1}{k}.$$

Since we assumed payoffs to be non-negative, $B(k) \geq C(k)$. Thus,

$$(k+1)C(k+1)\left(\frac{1}{kC(k)} - \frac{1}{NB(k)}\right) \geq \frac{(k+1)C(k+1)}{kC(k)}\frac{N-k}{N} \geq \frac{N-k}{N}.$$

For $k < N/2 - 1$ we therefore obtain

$$\begin{aligned} & \left(\frac{B(k+1)}{B(k)} - \frac{(k+1)C(k+1)}{kC(k)}\right) + (k+1)C(k+1)\left(\frac{1}{kC(k)} - \frac{1}{NB(k)}\right)\left(2 - \frac{N}{k+1}\right) \\ \leq & \ \frac{1}{k} + \frac{N-k}{N}\left(2 - \frac{N}{k+1}\right) \\ = & \ \frac{N(k+1) + 2(N-k)k(k+1) - (N-k)Nk}{Nk(k+1)}. \end{aligned}$$

The numerator equals $-2k^3 + (3N-2)k^2 - N(N-3)k + N$ which can easily be shown to be negative for $1 \leq k \leq N/2 - 2$ as long as $N \geq 8$. Hence, for $N \geq 8$ we have $\delta(k-2) < 0$ for $k \leq N/2$.

7 References

Barrett, S. (1992), International environmental agreements as games, in: Pethig, R.(ed.), Conflicts and cooperation in managing environmental resources, Springer, Berlin, 11-37.

Barrett, S. (1994), Self enforcing international environmental agreements, *Oxford Economic Papers* 46, 878-894.

Barrett, S. (1997), Heterogeneous international environmental agreements, in: Carraro, C. (ed.), International environmental agreements: strategic policy issues, E. Elgar, Cheltenham, 9-25.

Bolton, G.E. and A. Ockenfels (2000), ERC: a theory of equity, reciprocity, and competition, *The American Economic Review* 90, 166-193.

Bolton, G.E. and A. Ockenfels (2001), Self-centered fairness in games with more than two players, in: Charlie Plott and Vernon Smith (eds.), Handbook of Experimental Economics Results, North Holland: Amsterdam.

Borek, Th. and S. Rutz (2000), International environmental negotiations: does coalition size matter?, Paper presented at annual conference of Verein fuer Socialpolitik 2000, Berlin.

Carraro, C. and D. Siniscalco (1993), Strategies for the international protection of the environment, *Journal of Public Economics* 52, 309-328.

Dasgupta, C. (1994), The climate change negotiations, in: Mintzer, I.M. and J.A. Leonard (eds.), Negotiating climate change – the inside story of the Rio convention, Cambridge University Press, Cambridge, 129-148.

Hoel, M. (1992), International environmental conventions: the case of uniform reductions of emissions, *Environmental and Resource Economics* 2, 141-59.

Murdoch, J.C. and T. Sandler (1997), The voluntary provision of a pure public good: the case of reduced CFCs emissions and the Montreal protocol, *Journal of Public Economics* 63: 331-349.