RUHR
ECONOMIC PAPERS

Marco Rogna
Carla Vogt

Coalition Formation with Optimal Transfers when Players are Heterogeneous and Inequality Averse

#865

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger
Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, e-mail: W.Leininger@tu-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Ronald Bachmann, Prof. Dr. Roland Döhrn, Prof. Dr. Manuel Frondel, Prof. Dr. Ansgar Wübker
RWI, Phone: +49 (0) 201/81 49-213, e-mail: presse@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #865

Responsible Editor: Manuel Frondel

All rights reserved. Essen, Germany, 2020

ISSN 1864-4872 (online) – ISBN 978-3-96973-002-7

The working papers published in the series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editors.

Ruhr Economic Papers #865

Marco Rogna and Carla Vogt

Coalition Formation with Optimal Transfers when Players are Heterogeneous and Inequality Averse

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/96973002
ISSN 1864-4872 (online)
ISBN 978-3-96973-002-7

Marco Rogna and Carla Vogt[1]

Coalition Formation with Optimal Transfers when Players are Heterogeneous and Inequality Averse

Abstract

Obtaining significant levels of cooperation in public good and environmental games, under the assumption of players being purely selfish, is usually prevented by the problem of freeriding. Coalitions, in fact, generally fail to be internally stable and this cause a serious underprovision of the public good together with a significant welfare loss. The assumption of relational preferences, capable of better explaining economic behaviors in laboratory experiments, helps to foster cooperation, but, without opportune transfers scheme, no substantial improvements are reached. The present paper proposes an optimal transfers scheme under the assumption of players having Fehr and Schmidt (1999) utility functions, whose objective is to guarantee internal stability and to maximize the sum of utilities of coalition members. The transfers scheme is tested on a public good contribution game parameterized on the data provided by the RICE model and benchmarked with other popular transfers scheme in environmental economics. The proposed scheme outperforms its benchmarking counterparts in stabilizing coalitions and sensibly increases cooperation compared to the absence of transfers. Furthermore, for high but not extreme values of the parameter governing the intensity of dis-utility from disadvantageous inequality, it manages to support very large coalitions including three quarters of all players.

JEL-Code: C72; D63; H41; Q54

Keywords: Climate policy; coalitions; inequality aversion; RICE model; transfers scheme

September 2020

1 Both Bochum University of Applied Sciences. – All correspondence to: Carla Vogt, Bochum University of Applied Sciences, Am Hochschulcampus 1 , 44801 Bochum, Germany, e-mail: carla.vogt@hs-bochum.de

1 Introduction

In the long-lasting debate over the possibility to achieve significant levels of cooperation in international environmental agreements (IEAs) and, more generally, in public good games, it is well recognized that an important role is played by transfers (Carraro et al., 2006). In real-world environmental treaties, such as the Kyoto protocol or the Paris agreement, different typologies of transfers have been adopted with the aim of fostering the participation of developing countries. Strong, when not total, deductions in the emissions reduction burden, technology transfers and different forms of monetary incentives can all be enumerated as examples of transfers with the mentioned objective. The academic literature has also acknowledged the importance of this instrument to partially limit the detrimental effects of free-riding. Several authors, in fact, have proposed transfers schemes apt at stabilizing relatively large coalitions and tested them with encouraging results, e.g. Chander and Tulkens (1995), Botteon and Carraro (1997), Carraro et al. (2006) and Eyckmans and Tulkens (2006).

Despite some remarkable exceptions such as Chander and Tulkens (1995), Carraro et al. (2006) and McGinty et al. (2012), the transfers schemes analyzed in the domain of IEAs have been rarely derived by a game theoretical analysis of the game at hand. Several proposed transfers schemes are simply solution concepts derived from cooperative game theory – see Rogna (2016) for a review –, others, instead, are based on principles and fairness considerations originating outside the realm of game theory: e.g. sovereignty principle, polluters pay rule, capability rule, etc. (Finus, 2008). Both types of transfers schemes suffer the problem of being ill-suited for tackling the most prominent obstacle to coalition formation in IEAs: free riding. Cooperative solution concepts, being based on the cooperative paradigm, disregard by nature this aspect, and the other type of transfers schemes, totally detached from game theory, is even more problematic on this regard.

The present paper proposes an endogenous transfers scheme – the Rogna-Vogt (RV) transfers scheme – whose double objective is to maximize the global welfare of a coalition while incentivizing, as much as possible, participation. The main assumption upon which the proposed transfers rests is that the players involved in the public good game under consideration are endowed with Fehr and Schmidt (1999)[1] utility functions. This implies that players derive utility not solely from the material part of their payoff, but also in consideration of the relative amount of such payoff in comparison with that of the other players. This relational component of the utility function can be seen, therefore, as a desire for fairness. Together with the work of Bolton and Ockenfels (2000), Fehr and Schmidt (1999) preferences originated from the failure of standard economic theory to explain persistent phenomena, such as rejection of offers in ultimatum games or positive offers in dictator games, arising in economic experiments. Both works, by hypothesizing the presence of a relational component in the utility function of economic actors, have offered plausible explanations to these phenomena that have been confirmed in a series of subsequent experiments: e.g. Güth et al. (2003), Bolton and Ockenfels (2005), Fehr and Schmidt (2006) and Fischbacher and Gachter (2010).

Despite F&S preferences having been envisaged with individuals in mind, the possibility to extend this type of utility function to actors such as countries is not theoretically unfeasible. By hypothesizing the existence of a median voter endowed with F&S preferences, it follows that the negotiators bargaining over an IEA on her behalf should take into consideration such preferences. Furthermore, studies as Lange et al. (2007) and Dannenberg et al. (2010) have found empirical confirmations for the importance of fairness considerations in international negotiations related to environmental matters.

[1] Since now on simply F&S.

Resting on the F&S preferences assumption, the paper considers a public good game with heterogeneous players, where heterogeneity stems from the marginal benefit of consuming the public good and the marginal cost of producing it. The game is then solved for the optimal contribution of public good of each player given each possible coalition. Considering then a coalition of players able to attain a positive surplus necessary to prevent free-riding, given that the way in which this surplus is divided impacts the final wealth of the same coalition, the transfers scheme here proposed aims at maximizing this value. It is therefore collectively optimal and it constitutes a stationary equilibrium in a hypothetical bargaining game for sharing the surplus. Furthermore, the maximization of the coalition's wealth is operated through F&S preferences, thus conciliating efficiency with fairness. In addition to solving explicitly the maximization problem, that involves participation constraints taking the form of inequalities, we offer a relatively simple algorithmic solution resting on analytic expressions.

Finally, we test the performance of the transfers scheme through an emissions reduction game parameterized on data derived from the popular RICE model (Nordhaus and Yang, 1996). The test is operated in a comparative fashion, benchmarking the proposed transfers scheme with others, already present in the literature, derived from Finus (2008). The focus is placed on the ability to obtain large stable coalitions and on the welfare improvement granted by their formation. The RV transfers scheme outperforms the others, and the case of absence of transfers, and this holds in various sensitivity analyses. Furthermore, compared to standard preferences, there is a clear improvement in coalition size and welfare attainment.

An important aspect to be underlined is the non-monotonic relation between the degree of disadvantageous inequality aversion – represented by the parameter α – and the number and size of stable coalitions allowed by the RV transfers scheme. An increase of this parameter, in fact, strongly enlarges the number and size of stable coalitions, but, after a certain threshold, it reduces both of them. For a particular value of α, the RV transfers scheme allows for stable coalitions of nine members over twelve players.

Our results clearly show that, although a positive surplus is a necessary prerequisite for sustaining cooperation, the way in which it is distributed is a key component to allow for stable coalitions. In particular, its distribution must be perceived as fair. The proposed transfers scheme is generally strongly re-distributive, implying positive transfers from "strong" to "weak" players, but the comparison with other transfers schemes that are also re-distributive in nature but less successful in stabilizing coalitions implies the need to carefully balance transfers in order to foster cooperation.

Section 2 proposes a brief literature review, Section 3 presents the public good game and its solution, while Section 4 describes the RV transfers scheme. Section 5 is dedicated to the simulation and the comparison between transfers schemes and Section 6 concludes.

2 Literature Review

The present paper originates from two important strands in the economic literature. From one side, the game theoretical analysis of public good games and, more specifically, of environmental games. Particularly relevant for our purposes is the non-cooperative approach to coalition formation. On the other side, the behavioral literature criticizing the standard theory of purely self-interest on the ground of its inability to explain several experimental results. The outcome of this critique, namely the proposal to include a relational component into the utility function of economic agents, is central to this paper, as mentioned in the introduction. Finally, this is not the first attempt that tries to combine these approaches and it is therefore necessary to mention the previous works that have laid down such path.

The analysis of IEAs started in the early nineties with the seminal papers of Carraro and Siniscalco (1993), Barrett (1994) and Chander and Tulkens (1995). While the latter follows the cooperative approach, thus partially circumventing the problem of free-riding, the former two papers initiate the non-cooperative strand, whose results are rather pessimistic. Relying on the equilibrium concept of stability derived from the literature on cartel formation (d'Aspremont et al., 1983), they find that stable coalitions include only a very tiny fraction of all players. Furthermore, Barrett (1994) underlines the cooperation paradox according to which the higher are the gains from it, the higher the incentives to free ride, so that cooperation is easier to attain when it is less needed. In the following three decades, are countless the proposed variants of these models. Several authors, among which McGinty (2007), Pavlova and De Zeeuw (2013) and Bakalova and Eyckmans (2019), take into consideration heterogeneous countries; Weikard et al. (2015) add the test of a minimum participation constraint, while Diamantoudi and Sartzetakis (2015) and Breton and Garrab (2014) examine the role of, respectively, farsightedness and evolutionary farsightedness. Despite significant improvements in cooperation attainments reached by some of these variants, their theoretical and\or empirical foundations are not always commonly accepted and rest on specific assumptions.

Note that the just mentioned literature on coalition formation is almost always based on standard assumptions with regard to preferences, namely pure self-interest. As mentioned, this assumption has been strongly criticized by the behavioralist school, that, on the base of vast experimental evidence, has shown how self-interest often fails in explaining actual behavior. In particular, positive contributions in the dictator game, e.g. Kahneman et al. (1986), as well as rejections of offers in the ultimatum game, e.g. Güth et al. (1982), are at odd with rationality and self-interest. The attempts to explain these results holding the assumption of rational behavior have lead to consider a relational component into the utility function of agents: Fehr and Schmidt (1999) and Bolton and Ockenfels (2000). In particular, both advantageous and disadvantageous inequality seem to be disliked, with the aversion for the latter being stronger. Güth et al. (2003), Bolton and Ockenfels (2005), Fehr and Schmidt (2006) and Fischbacher and Gachter (2010) have confirmed, through experiments, the explanatory ability of this relational component in predicting actual behaviors.

An attempt to link these two literature strands has already been initiated. One way, undertaken by van der Pol et al. (2012), considers players as having a degree of pure altruism, consequently leading to higher levels of cooperation. The other way, to which we are more indebted, assumes instead that the actors involved in a public good game are endowed with utility functions featuring a relational component. Lange and Vogt (2003) consider identical players with F&S preferences finding an improvement in cooperation compared to the case of standard preferences. A similar result, although more multifaceted, is obtained by considering heterogeneous players (Lange, 2006). Vogt (2016) not only considers heterogeneous players, whose characterizing parameters are derived from the RICE model, as in the present paper, but she also notices the need for transfers schemes in order to achieve substantial improvements in cooperation. The present paper draws heavily on this last work substantially improving its results. In fact, it proposes an optimal transfers scheme for players endowed with F&S preferences committed in a public good game entailing the formation of coalitions.

3 The Public Good Game

The game here considered is a rather standard coalitional public good contribution game. It features two stages with the first describing the decision of players to join or not a coalition and the second representing their choice of the contribution for the production of a public good. Coalition members are assumed to act in the interest of the whole coalition, by maximizing the joint utility of its members, whereas outsiders simply maximize their own utility. The game is solved backwards by firstly considering the last step and finding the optimal contribution to the production of the public good of each player. Following is a description of the game.

There is a set $N = \{1, 2, ..., n\}$ of players, each of which having a strictly positive endowment z: $z_i > 0, \forall i \in N$. Assume, *w.l.o.g*, that players are ordered in decreasing[2] order of endowment: $z_1 > z_2 > ... > z_n$. Player i can use her endowment z_i, or part of it, for private consumption (x_i) or providing it as input (q_i) for the production of the public good y. There is a simple linear production function describing the process of generating the public good: $y = \gamma \sum_{j \in N} q_j$, with γ being an efficiency parameter assumed to be strictly positive. A generic player i derives her payoff by the consumption of the private and of the public good: $\pi_i = p_i x_i + a_i y$, where p_i and a_i, strictly positive, are simply the values attributed by i to the consumption of the associated goods. Finally, there is a non-linear transformation function linking the consumption of the private good with the provision of the input q for producing the public good:

$$x_i = z_i - \frac{1}{z_i} q_i^2, \forall i \in N. \tag{1}$$

Note that equation (1) is concave – $\frac{dx_i}{dq_i} = -\frac{2q_i}{z_i}$ and $\frac{d^2 x_i}{dq_i} = -\frac{2}{z_i}$ – and that it implies $x_i = 0 \iff q_i = z_i$ and $q_i = 0 \iff x_i = z_i$. Once considering this transformation function, the payoff of player i can be rewritten as a function of q_i alone:

$$\pi_i = \left(z_i - \frac{1}{z_i} q_i^2 \right) p_i + a_i \gamma \sum_{j \in N} q_j, \forall i \in N. \tag{2}$$

Given the concavity of the transformation function, the players' payoff is very similar, in its behavior, to the one used by Barrett (1994) for describing an abatement game, with the public good being there represented by the decrease in pollution.

Players' utility is modeled as in Fehr and Schmidt (1999):

$$U_i(\pi_i, \pi_{-i}) = \pi_i - \frac{\alpha_i}{n-1} \sum_{j \in I^+} (\pi_j - \pi_i) - \frac{\beta_i}{n-1} \sum_{k \in I^-} (\pi_i - \pi_k), \forall i \in N. \tag{3}$$

Beyond the material component (π_i), given by the consumption of the private and of the public good as in (2), the utility of player i is negatively affected by disadvantageous inequality, with α_i representing its intensity, and by advantageous inequality, with β_i having an analogous role as α_i. I^+ and I^- indicate[3] the sets of players with a payoff grater, the former, and lower, the latter, than the one of player i: $j \in I^+ \iff \pi_j > \pi_i$ and $k \in I^- \iff \pi_k < \pi_i$. In general, it is assumed, and it has been

[2] We use "decreasing" instead of "non-increasing" since, in our simulation, z is represented by countries' per capita GDP, making perfect coincidence extremely unlikely.

[3] Depending on the subscript used to identify a player, the identifiers of these sets will change accordingly: e.g. J^+ and J^- for the sets referring to player j.

confirmed in several experiments (Fehr and Schmidt, 1999; Blanco et al., 2011), that $\alpha_i > \beta_i, \forall i \in N$.

By substituting the payoffs (π) in (3) with the expressions obtained in (2) and solving the maximization problem of each player, having only q has variable, we can find the analytic solution for the optimal level of q for both the members of a coalition and for the outsiders. The difference simply stays in the maximization problem to be solved, with an outsider, say i, solving $\max_{q_i} U_i$, for $i \in N \setminus C$, whereas a member of a coalition C, say j, solving $\max_{q_j} \sum_{k \in C} U_k$, for $j \in C$. Once solved the resulting system of equations – we have a system of equations since the derivative of the utility function of a player for her own level of q will necessary have the level of q of the other players in it – we get the analytic solutions for all qs. Since the game is the same as the one in Vogt (2016), we report the solutions without proofs, reminding the interested reader to the mentioned paper.

$$q_i^* = \frac{z_i a_i \gamma}{2p_i} - \frac{z_i}{2p_i} \frac{\alpha_i \sum_{j \in I^+} a_j \gamma - \beta_i \sum_{k \in I^-} a_k \gamma}{n - 1 + |I^+|\alpha_i - |I^-|\beta_i}, \forall i \notin C. \tag{4A}$$

$$q_i^* = \frac{\mathrm{MB}_i}{\mathrm{MC}'_i} - \frac{1}{\mathrm{MC}'_i} \frac{\alpha_i \sum_{j \in I^+} \mathrm{MB}_j - \beta_i \sum_{k \in J^-} \mathrm{MB}_k}{n - 1 + |I^+|\alpha_i - |I^-|\beta_i}, \forall i \notin C. \tag{4B}$$

Note that equation (4B) is just the same as (4A) with $MB_i = a_i \gamma$ representing the marginal benefit of an additional unit of q_i and $MC_i = \frac{2p_i q_i}{z_i}$ its marginal cost. Finally, $|I^+|$ and $|I^-|$ represent the cardinality of, respectively, set I^+ and I^-. This being the optimal level of q for an outsider, let us see the one of a coalition member:

$$q_i^* = \frac{z_i}{2p_i} \frac{(n-1) \sum_{j \in C} a_j \gamma - \sum_{j \in C} \alpha_j \left(\sum_{k \in J+} a_k \gamma - |J^+| a_j \gamma \right) - \sum_{j \in C} \beta_j \left(|J^-| a_j \gamma - \sum_{k \in J^-} a_k \gamma \right)}{n - 1 + \alpha_i (|I^+| - |C_i^-|) - \beta_j (|I^-| - |C_i^+|)}, \forall i \in C. \tag{5A}$$

$$q_i^* = \frac{1}{MC'_i} \frac{(n-1) \sum_{j \in C} MB_j - \sum_{j \in C} \alpha_j \left(\sum_{k \in J+} MB_k - |J^+| MB_j \right) - \sum_{j \in C} \beta_j \left(|J^-| MB_j - \sum_{k \in J^-} MB_k \right)}{n - 1 + \alpha_i (|I^+| - |C_i^-|) - \beta_j (|I^-| - |C_i^+|)}, \forall i \in C. \tag{5B}$$

Here, C_i^+ and C_i^- represent the sets of players belonging to coalition C and having a payoff, respectively, higher and lower than player i[4]. Once having determined the optimal level of q for all players in each possible coalition, it is then possible to compute their payoffs and, consequently, their utilities. Payoffs, however, can be transfered among coalition members in order to increase the possibility of cooperation. The next section is dedicated to describe our proposed transfer scheme to achieve this objective.

4 The Transfer Scheme

In order to understand the proposed transfer scheme, it is firstly necessary to have a clear idea of the equilibrium concept used in the game: coalition stability. Derived from the literature on cartel stability and formally defined in d'Aspremont et al. (1983), stability can be seen as a translation of the Nash equilibrium to a coalitional setting. It requires two conditions to hold simultaneously: internal and external stability. Following is a formal definition.

Definition 4.1. (Coalition Stability) Given a coalition $C \subseteq N$, such coalition is said to be stable if it is both internally and externally stable, with internal stability requiring

$$U_i(C) \geq U_i(C \setminus i), \forall i \in C;$$

[4] Alternatively, $C_i^+ = I^+ \cap C$. Analogously, we can define set C_i^-.

and external stability requiring

$$U_j(C) \geq U_j(C \cup j), \forall j \in N \setminus C.$$

In simple words, a coalition is said to be stable if each one of its members has no incentive to leave it and no one of the outsiders can gain by joining it. In public good and in environmental games, the formation of a coalition generally entails an increase of the global welfare since the coalition members, maximizing their joint utility, will provide a higher level of public good or will reduce the production of a public bad (e.g. polluting emissions). Therefore, the larger a coalition, the larger is the total welfare, given a greater amount of public good, but also the higher is the incentive to be an outsiders, since this allows to reap the cooperative benefits without paying its costs (increasing own contribution of q). This further implies that, while lack of external stability may be a problem for low levels of cooperation, the lack of internal stability is the key concern that generally prevents cooperation in public good games. This is well documented in Barrett (1994) and in most of the game theoretical literature dealing with environmental games: e.g. Carraro and Siniscalco (1993), Diamantoudi and Sartzetakis (2006) and Finus et al. (2017), just to name a few.

Ideally, therefore, a transfer scheme should solve the internal stability problem: $U_i(C) \geq U_i(C \setminus i), \forall i \in C$. There are, however, two limitations to take into consideration. The first relates to the object of the transfer. From a theoretical point of view, it seems safer to allow for transfers in terms of payoffs rather than in terms of utils. The formers, in fact, either are expressed in monetary amounts, as in our simulation, or they can be easily converted into them, whereas the latter, particularly in the case of F&S preferences, would require a potentially complex mapping function to translate utils into monetary amounts. Our choice, therefore, is to allow only for transfers in terms of payoffs. The second limitation is that internal stability cannot be always guaranteed, even by making use of transfers. This can be easily seen if we consider the implementation of a transfer scheme as a "cake division problem". Borrowing notation from cooperative game theory, let us define the worth of a coalition as $v(C) = \sum_{i \in C} \pi_i(C)$. Then, the post-transfer payoff of player i will be $\bar{\pi}_i = \theta_i v(C), \forall i \in C$, with $\sum_{i \in C} \theta_i = 1$ and $\theta_i \geq 0, \forall i \in C$.[5] We can then formally state the mentioned limitation of a transfer scheme.

Proposition 4.1. *If the worth of a coalition is lower than the sum of the payoffs coalition members could get by abandoning, one per time, the coalition, there is no feasible transfers scheme able to solve the problem of internal stability:* $v(C) - \sum_{i \in C} \pi_i(C \setminus i) < 0 \Rightarrow \nexists \boldsymbol{\theta} : \bar{\pi}_i = \theta_i v(C) \geq \pi_i(C \setminus i), \forall i \in C$.

Proof. Let us suppose to have $v(C) - \sum_{k \in C} \pi_k(C \setminus k) < 0$ and to have implemented a transfers scheme such that $\bar{\pi}_j = \pi_j(C \setminus j), \forall j \in C \setminus i$. This means that we have implemented a transfers scheme able to solve the internal stability problem for all players in C except that for player i and that this has been done by granting to each of them the minimum possible payoff. Since $\bar{\pi}_j = \theta_j v(C), \forall j \in C \setminus i$, we then have $\theta_i v(C) = (1 - \sum_{j \in C \setminus i} \theta_j) v(C) \Rightarrow \theta_i = 1 - \sum_{j \in C \setminus i} \theta_j$. If $\sum_{j \in C \setminus i} \theta_j \geq 1$, then $\theta_i \leq 0$, but this would necessary fail to guarantee internal stability since $\bar{\pi}_i \leq 0 < \pi_i(C \setminus i)$. Assume, instead, $\sum_{j \in C \setminus i} \theta_j < 1$ and that $\theta_i v(C) \geq \pi_i(C \setminus i)$. But then, $\sum_{k \in C} \theta_k v(C) \geq \sum_{k \in C} \pi_k(C \setminus k)$ and, since $\sum_{k \in C} \theta_k = 1$, we have $v(C) - \sum_{k \in C} \pi_k(C \setminus k) \geq 0$, contradicting our starting assumption. ■

Once having established that payoffs and not utils can be transferred and that transfers are helpful in solving the internal stability problem only if there is enough to distribute, we can write a preliminary

[5]This way of expressing a transfer may appear odd. A more canonical way would in fact be: $\bar{\pi}_i = \pi_i + \tau_i v(C), \forall i \in C, \sum_{i \in C} \tau_i = 0$. However, the two ways are perfectly equivalent, with the following equation relating $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$: $\theta_i = \frac{\pi_i}{\sum i \in \pi_i} + \tau_i, \forall i \in C$.

version of our transfers scheme:

$$\bar{\pi}_i = \begin{cases} \pi_i(C \setminus i) + \theta_i \left(v(C) - \sum\limits_{j \in C} \pi_j(C \setminus j) \right), \text{ if } v(C) - \sum\limits_{j \in C} \pi_j(C \setminus j) > 0; \\ \pi_i(C), \text{ otherwise;} \end{cases} \tag{6}$$

with $\theta_i \geq 0, \forall i \in C$ and $\sum_{i \in C} \theta_i = 1$. From (6), it is clear that internal stability, at least at payoff level, is always respected whenever it is possible. Each player, in fact, receives her outsider payoff plus a non-negative share of the remaining, positive, surplus. However, the transfers scheme is not yet fully determined since the vector $\boldsymbol{\theta}$ has not been defined.

Although there are infinite possibilities to share the surplus $v(C) - \sum\limits_{j \in C} \pi_j(C \setminus j)$, the fact that the way in which it is distributed affects not only the utility of single players but also their joint utility helps us in establishing a valid criterion for characterizing $\boldsymbol{\theta}$. Given that coalition members act for maximizing their joint utility, it seems natural that the selected transfers scheme serves this purpose as well. For simplifying reasons, let us define the surplus as $s(C) = v(C) - \sum\limits_{j \in C} \pi_j(C \setminus j)$ and assume $s(C) > 0$. The utility of a player, member of the coalition C, once implemented the transfers will then be:

$$\begin{aligned} U_i = \pi_i + \theta_i s(C) - \frac{\alpha_i}{n-1} \sum_{j \in C_i^+} \Big(\pi_j + \theta_j s(C) - \pi_i - \theta_i s(C) \Big) + \\ - \frac{\beta_i}{n-1} \sum_{k \in C_i^-} \Big(\pi_i + \theta_i s(C) - \pi_k - \theta_k s(C) \Big) + \\ - \tfrac{\alpha_i}{n-1} \sum_{r \in I^+ \setminus C_i^+} \Big(\pi_r - \pi_i - \theta_i s(C) \Big) - \tfrac{\beta_i}{n-1} \sum_{l \in I^- \setminus C_i^-} \Big(\pi_i + \theta_i s(C) - \pi_l \Big). \end{aligned} \tag{7}$$

As mentioned, the objective is to maximize the joint utility of coalition members. Therefore, we define the vector $\boldsymbol{\theta}$ by solving the following optimization problem: $\max\limits_{\boldsymbol{\theta}} \sum_{i \in C} U_i$.

Before analyzing this maximization problem and adding the necessary constraints, it is opportune to reformulate equation (7) in order to have a more concise form. In the third row of (7) we have the dis-utility of player i for the disadvantageous and advantageous inequality suffered with respect to outsiders. In fact, the two summations are made over the sets $I^+ \setminus C_i^+$ and $I^- \setminus C_i^-$. Being outsiders, the members of these sets clearly do not receive any surplus transfer. However, just for simplifying purposes, we can enlarge the vector $\boldsymbol{\theta}$ from being $|C|$-dimensional to be n-dimensional and set the value of its elements corresponding to outsiders equal to zero: $\theta_i = 0, \forall i \in N \setminus C$. Clearly, nothing has changed in practical terms, since outsiders will still have their base payoff without any addiction, but this modification allows to rewrite equation (7) as follows:

$$\begin{aligned} U_i = \pi_i + \theta_i s(C) - \frac{\alpha_i}{n-1} \sum_{j \in I^+} \Big(\pi_j + \theta_j s(C) - \pi_i - \theta_i s(C) \Big) + \\ - \frac{\beta_i}{n-1} \sum_{k \in I^-} \Big(\pi_i + \theta_i s(C) - \pi_k - \theta_k s(C) \Big). \end{aligned} \tag{8}$$

Once having defined a more compact form for the post-transfer utility of a coalition member, it is time to discuss the constraints of the optimization problem to be solved. The non-negativity of each element of vector $\boldsymbol{\theta}$ has already been mentioned. Each coalition member should have, at least, the payoff she would get by leaving the coalition: $\theta_i \geq 0, \forall i \in C$. Furthermore, the sum of the elements of $\boldsymbol{\theta}$ cannot be greater than one, since this would imply to distribute more than what is available. However, given that efficiency is a reasonable and obvious objective and given that the maximization program would necessarily lead to efficiency, we can turn this inequality constraint into an equality: $\sum_{i \in N} \theta_i = 1$. Finally, we need to add a last constraint. A transfer scheme must be unanimously accepted by all coalition members to be implemented. This implies that, in order to be accepted, each member must have a post-transfer utility that is at least equal to the utility that she will obtain if no transfers scheme was implemented. Therefore, if we name U_i^{pt} the post-transfer utility of player i as defined by equation (8) and U_i^{nt} as the utility that i would get as coalition member in absence of any transfer, defined in equation (3), we then have the condition $U_i^{pt} \geq U_i^{nt}$. By some manipulation, it is possible to restate this constraint into the following form:

$$\theta_i \geq \frac{\alpha_i \sum_{j \in I^+} \theta_j - \beta_i \sum_{k \in I^-} \theta_k}{n - 1 + \alpha_i |I^-| - \beta_i |I^+|}, \forall i \in C. \tag{9}$$

Finally, we can write the optimization problem with all constraints:

$$\begin{aligned} &\max_{\theta_i, \forall i \in C} U_i^{pt} \\ &\text{s.t.} \\ &\sum_{i \in N} \theta_i = 1, \\ &\theta_i \geq \frac{\alpha_i \sum_{j \in I^+} \theta_j - \beta_i \sum_{k \in I^-} \theta_k}{n - 1 + \alpha_i |I^-| - \beta_i |I^+|}, \forall i \in C, \\ &\theta_i \geq 0, \forall i \in C, \\ &\theta_i = 0, \forall i \in N \setminus C. \end{aligned} \tag{10}$$

By solving this optimization problem we obtain a transfers scheme vector that, once inserted into (6), guarantees to each coalition member a payoff at least as great as the one obtained by leaving the coalition and that maximizes the worth of the same coalition. Furthermore, it also guarantees that no coalition member suffers a loss by accepting the transfers scheme rather than simply participating to the coalition without any transfer being implemented.

4.1 An algorithm to find the optimal transfers vector

It is immediate to see that the constraints of the optimization problem (10) form a convex and compact polytope and, therefore, the set of solutions is always non-empty. The problem is not particularly complex, being a linear optimization problem, and, numerically, it can be solved by using either the simplex algorithm or an interior-point method. Solving the Lagrangean is another possibility, but it is tedious and quite cumbersome. However, we are going to show a method to solve the problem that relies entirely on analytic solutions whose derivation will be here displayed. Let us exclude, for mere convenience, the last set of equality constraints ($\theta_i = 0, \forall i \in N \setminus C$), being it inconsequential since we can treat the θs related to outsiders as simple parameters. The Lagrangean will then be:

$$\mathcal{L} = \sum_{i \in N} \left\{ \pi_i + \theta_i s(C) - \frac{\alpha_i}{n-1} \sum_{j \in I^+} \Big(\pi_j + \theta_j s(C) - \pi_i - \theta_i s(C) \Big) + \right.$$
$$\left. - \frac{\beta_i}{n-1} \sum_{k \in I^-} \Big(\pi_i + \theta_i s(C) - \pi_k - \theta_k s(C) \Big) \right\} - \lambda_1 \left(\sum_{i \in N} \theta_i - 1 \right) +$$
$$\sum_{i \in N} \lambda_{1+i} \left(\theta_i - \frac{\alpha_i \sum_{j \in I^+} \theta_j - \beta_i \sum_{k \in I^-} \theta_k}{n - 1 - \beta_i |I^-| + \alpha_i |I^+|} \right) + \sum_{i \in N} \lambda_{n+1+i} \theta_i.$$

By taking the derivative of $\mathcal{L}$ for θ_1, assuming 1 being the first player in C, we have:

$$\frac{\partial \mathcal{L}}{\partial \theta_1} = 1 - \beta_1 - \sum_{j \in C \setminus 1} \frac{\alpha_1}{n-1} - \sum_{j \in C \setminus 1} \frac{\lambda_{j+1 \alpha_j}}{n - 1 + |J^+| \alpha_j - |J^-| \beta_j} + \lambda_{n+2}.$$

By equating this derivative to zero, it is immediately clear that λ_{n+2} must necessarily be positive for the equality to hold, being all other therms strictly non-positive. Since λ_{n+2} is the multiplier associated with the constraint $\theta_1 \geq 0$, this necessarily implies that $\theta_1 = 0$. Remembering that players are ordered in decreasing level of endowment and assuming this mirrors the payoff ordering, we have that the strongest coalition member – the one with the highest payoff – will not get any share of the surplus. We have a first optimal element of vector $\boldsymbol{\theta}$.

From simplex theory, we know that an optimal vector must lie at the intersection of n hyperplanes, with these lasts being defined by the constraints of the optimization problem. As mentioned, the $|N \setminus C|$ equality constraints relative to the outsiders are necessary binding. Then we have the upper bound defined by the other equality constraint, the sum of the elements of $\boldsymbol{\theta}$ must be equal to one, implying that there is one $\boldsymbol{\theta}$ element such that $\theta_i = 1 - \sum_{j \in C \setminus i} \theta_j$ and $\theta_i > 0$. As seen, this element cannot be the one associated to the first player in C. Therefore, we can say that there is always a coalition member whose associated $\boldsymbol{\theta}$ element is positive, whereas the $\boldsymbol{\theta}$ elements of the players preceding her are all equal to zero: $\theta_i = 1 - \sum_{j \in I^-} \theta_j$, $\theta_i > 0$ and $\sum_{k \in I^+} \theta_k = 0$. Let us call i the pivot player. For all the coalition members following i, one of the other two lower bound constraints must be binding: $\theta_j \geq 0$ or constraint (9). Note that the two constraints, except for the case in which the RHS of (9) is equal to zero, are mutually exclusive. Therefore, for all j players in C_i^- or $\theta_j = 0$ or $\theta_j = \frac{\alpha_j \sum_{q \in J^+} \theta_q - \beta_j \sum_{k \in J^-} \theta_k}{n - 1 + \alpha_j |J^-| - \beta_j |J^+|}$.

Let us assume the pivot player is the last one in C. By the definition we have given of the pivot player, since $C_i^- = \emptyset$, we have $\theta_i = 1$ and $\theta_j = 0, \forall j \in N \setminus i$. It is easy to check that this vector represents a feasible solution since it satisfies all constraints. Assume now that the pivot player i is the one before the last. Consider than the inequality constraint (9) of player j, the only coalition member following i. It will look as follows:

$$\theta_j \geq \frac{\alpha_j (1 - \theta_j)}{(n-1)(1 - \alpha_j)}.$$

Since $0 \leq \theta_j < 1$, the RHS is necessarily positive, implying that constraint (9) is binding since $\theta_i \neq 0$. We then get $\theta_j = \frac{\alpha_j}{n - 1 + n\alpha_j}$, $\theta_i = \frac{(n-1)(1+\alpha_j)}{n-1+n\alpha_j}$ and all other $\boldsymbol{\theta}$'s elements equal to zero. This is another feasible solution. We can continue supposing the pivot player is the third last member of coalition C. The constraint (9) for the last player remains identical. However, a more general formulation of constraint (9) for all coalition members following i is the following:

$$\theta_j \geq \frac{\alpha_j(1-\theta_j) - (\alpha_j+\beta_j)\sum_{k\in J^-}\theta_k}{n-1+|J^+|\alpha_j - |J^-|\beta_j}.$$

Once moving θ_j to the LHS, we have that the condition for the positiveness of the RHS is the following: $\frac{\alpha_j}{\alpha_j+\beta_j} > \sum_{k\in J^-}\theta_k$. Whenever this condition holds, constraint (9) is binding, and then we can use this expression to find the value of θ_j. Otherwise, this last will be equal to zero. By starting from the last player, all the values of the $\boldsymbol{\theta}$ elements can be determined once having defined a pivot player. In particular, if we define the set P_j as the set of coalition members following member j for which the just mentioned condition holds, we have the following analytic solution for all θs relative to coalition members coming after the pivot player.

$$\theta_j = \begin{cases} 0, \quad \text{if} \quad \frac{\alpha_j}{\alpha_j+\beta_j} \leq \sum\limits_{k\in P_j}\theta_k, \\ \dfrac{\alpha_j \prod\limits_{k\in P_j}\big(n-1+(|K^+|+1)\alpha_k - |K^-|\beta_k\big) - (\alpha_j+\beta_j)\left(\sum\limits_{k\in P_j}\alpha_k \prod\limits_{q\in P_j\setminus k}(n-1+|Q^+|\alpha_q - |Q^-|\beta_q)\right)}{\prod\limits_{k\in P_j}(n-1+(|K^+|+1)\alpha_k - |K^-|\beta_k)(n-1+(|J^+|+1)\alpha_j - |J^-|\beta_j)}, \text{ oth.} \end{cases}$$

We finally have all analytic expressions for finding the value of each $\boldsymbol{\theta}$ element once having selected a pivot player. Since the first coalition member cannot be a pivot player due to the fact that her associated $\boldsymbol{\theta}$ value must be equal to zero, as seen before, we have $|C|-1$ feasible transfers vectors. At this point we simply need to compute the utilities of all players for each one of the feasible transfers using (6) and see which one gives the highest sum of coalition members' utilities. This one will be the optimal transfer vector.

5 Simulation and Benchmarking of Transfer Schemes

The objective of this section is to evaluate the performances of the proposed transfers scheme in terms of stability outcomes and welfare gains. In order to do so, it is required to formulate a proper game and to have a touchstone for benchmarking the RV scheme. Given the extreme difficult of doing so through analytic results, due to the complexity of the task, a numerical simulation is a second best. With regard to the game, our choice is to propose the same setting as in Vogt (2016). Since the present paper proposes an optimal transfers scheme for the model presented in Vogt (2016), it is naturally interesting to compare the results with the ones reached in the mentioned paper. Furthermore, the game in Vogt (2016) is parameterized on the popular RICE model of Nordhaus and Yang (1996), one of the most widespread models evaluating the impact of global warming. More specifically, the RICE model aggregates the world into 12 macro-regions, some of which coinciding with countries, and predicts the trajectories of several key variables, among which GDP, pollution and environmental damages for future time periods. By using the same 12 regions as the players for our game, we can retrieve from the RICE model all the elements to fill equations (2) and for deriving the optimal levels of public good contribution: equations (4B) and (5B). The level of endowment z is represented by the per-capita GDP of each region and the contribution to the public good, q, is the fraction of per-capita GDP used to abate CO_2 emissions. The marginal benefit (MB) of abatement are derived from Nordhaus (2011), whereas the procedure to compute the marginal cost (MC) is identical to the one described in Vogt (2016). One difference is that this last work considers only the data of 2015 to perform the simulation, whereas we will additionally take into consideration years 2025 and 2035 as a sensitivity analysis. In Table 1, all the relevant parameters used in the simulation, except the values of α and β, are reported.

Table 1: Countries with Relevant Parameters

		2015			*2025*			*2035*	
Countries	**z**	**MC**	**MB**	**z**	**MC**	**MB**	**z**	**MC**	**MB**
USA (US)	49018	330.75	3.6	58846	330.71	4.38	68690	271.56	5.28
Japan (JPN)	36671	514.45	0.78	43106	511.02	0.95	51364	430.48	1.11
OHI (OHI)	35093	404.23	1.37	44885	408.34	1.77	53307	339.21	2.06
EU (EU)	32003	514.45	4.11	39724	522.21	5.2	47777	431.96	6.29
Russia (RUS)	16287	220.52	0.51	20834	222.70	0.79	25749	185.37	0.95
Mid. East (MEA)	11550	367.49	3.36	15046	368.23	5.04	18900	304.18	6.48
Latin Am. (LAT)	11531	251.29	2.6	15497	306.12	3.97	20025	285.78	5.41
Eurasia (EUA)	8604	220.52	0.48	9646	219.11	0.87	12863	183.46	1.24
China (CHN)	6931	257.27	10.4	11692	257.35	23.92	15256	214.66	31.7
Others (OTH)	3866	233.54	6.29	5703	289.85	11.62	8013	281.03	19.97
India (IND)	3672	404.23	7.98	5345	404.32	16.91	7419	336.24	26.03
Africa (AFR)	2596	203.74	7.83	3831	264.53	13.87	5383	254.51	24.75

Once having determined all the required elements for computing the payoffs, we need to establish the values of α and β in order to compute the utilities and to describe the alternative transfers scheme that will be used as a benchmark. We start by this last task, postponing the discussion about the inequality aversion parameters. First of all, note that equation (6) is still used as the base for computing the alternative transfers. What varies, therefore, is the definition of the vector $\boldsymbol{\theta}$. As mentioned in the introduction, different transfers schemes, often inspired by antithetical principles, have been proposed in the literature. Finus (2008) collects an exhaustive sample whose merit is to feature transfers whose implementation is already translated into mathematical terms. Furthermore, several of these transfers schemes have been largely debated during the negotiations of real-world IEAs (Finus, 2008). They seem therefore well suited for our purposes. Table 2 shows the selected alternative transfers scheme together with a description of their inspiring principles. Note that, for the computation of some of these transfers, additional parameters not included in Table 1, such as the level of CO_2 emissions or the population, are required. Such data have been retrieved from the RICE model as well.

Table 2: Definition of Alternative Transfers Schemes

Sovereignty principle	
Current GDP (C_GDP)	$\theta_i = \frac{GDP_i}{\sum_{j \in C} GDP_j}$
Current Emissions (C_EM)	$\theta_i = \frac{e_i}{\sum_{j \in C} e_j}$
Ability to pay	
Inverse of per capita GDP (I_GDP)	$\theta_i = \frac{\left(\frac{GDP_i}{POP_i}\right)^{-1}}{\sum_{j \in C}\left(\frac{GDP_j}{POP_j}\right)^{-1}}$
Polluters pay	
Inverse of emissions (I_EM)	$\theta_i = \frac{e_i^{-1}}{\sum_{j \in C} e_j^{-1}}$
Egalitarian	
Equal sharing (EQ_SH)	$\theta_i = \frac{1}{\lvert C \rvert}$
Equal per capita sharing (EQ_pcSH)	$\theta_i = \frac{POP_i}{\sum_{j \in C} POP_j}$

$e = CO_2$ emissions, POP = Population

Once defined six benchmarking transfers schemes, the last elements still missing are the values for α and β. Lacking studies providing their estimates for subjects other than single persons, and remembering the median voters argument previously proposed, we rely on the data provided by the same Fehr and Schmidt (1999)[6] and used also in Vogt (2016): $\alpha = 0.833$ and $\beta = 0.288$. As a sensitivity analysis, we also double and halve the two parameters in order to investigate the effect on stability. Note that, since we lack any evidence for regional differences in the degree of inequality aversion, both parameters are always kept identical for all players.

5.1 Results

Table 3 reports the results obtained for the years 2015, 2025 and 2035, with $\alpha = 0.833$ and $\beta = 0.288$. It is immediately visible how the RV transfers scheme outperforms the others in terms of fully stable coalitions. In fact, both the number and the dimension of the supported coalitions are larger. The other transfers schemes may even be worst than the case of absence of transfers, with this being particularly true for the year 2035, when the benefits of pollution abatement are greater and, according to Barrett's paradox, cooperation is more difficult to achieve. Another important element to notice is that, among the alternative transfers schemes, the ones that are more re-distributive performs generally better, with re-distribution here implying positive transfers from players with high levels of z to players with lower levels. The Inverse per-capita GDP transfers scheme – remembering that z is per-capita GDP – is, in fact, the most successful way of distributing the surplus after the RV scheme in guaranteeing stability. An exception is the Inverse per-capita emissions scheme that, despite the positive correlation between per-capita GDP and emissions, rendering it a rather re-distributive scheme, performs very poorly. Although we have not underlined it before, the RV scheme is also very re-distributive. In fact, given the intrinsic nature of the F&S utility function, the maximization of the joint utility of coalition members requires to reduce as much as possible

[6] Fehr and Schmidt (1999) actually propose an interval for both parameters retrieved from experimental results. The values here adopted are the median values of such intervals.

inequality, thus favoring weak players.

Table 3: Stable Coalitions under Different Transfers Schemes ($\alpha = 0.833, \beta = 0.288$)

Year	RV	C_GDP	C_EM	I_GDP	I_EM	EQ_SH	EQ_pcSH	No Transfer
				Fully Stable Coalitions				
2015	US, CHN, IND, AFR	CHN, IND	–	EUA, OTH, IND	–	EUA, OTH, IND	CHN, IND	CHN, IND
	OHI, CHN, IND, AFR							IND, AFR
	JPN, CHN, IND, AFR							
	EU, CHN, IND, AFR							
	RUS, CHN, IND, AFR							
	LAT, CHN, IND, AFR							
	LAT, OTH, IND, AFR							
	MEA, CHN, IND, AFR							
	MEA, OTH, IND, AFR							
	CHN, EUA, IND, AFR							
	CHN, OTH, IND, AFR							
				Number of Internally Stable Coalitions				
	137	35	36	100	55	66	48	11
				Number of Externally Stable Coalitions				
	315	224	224	224	224	224	224	225
				Fully Stable Coalitions				
2025	US, CHN, IND, AFR	–	–	CHN, IND, AFR	–	CHN, OTH, IND	CHN, OTH, IND	CHN, IND
	OHI, CHN, IND, AFR							IND, AFR
	JPN, CHN, IND, AFR							
	EU, CHN, IND, AFR							
	RUS, CHN, IND, AFR							
	LAT, CHN, IND, AFR							
	MEA, CHN, IND, AFR							
	MEA, OTH, IND, AFR							
	CHN, EUA, IND, AFR							
	CHN, OTH, IND, AFR							
				Number of Internally Stable Coalitions				
	140	34	39	93	47	65	48	9
				Number of Externally Stable Coalitions				
	313	213	213	213	213	213	213	215
				Fully Stable Coalitions				
2035	EU, IND, AFR	–	–	–	–	–	–	IND, AFR
	OHI, CHN, IND, AFR							
	OHI, EUA, IND, AFR							
	JPN, CHN, IND, AFR							
	RUS, CHN, IND, AFR							
	RUS, OTH, IND, AFR							
	LAT, CHN, IND, AFR							
	MEA, CHN, IND, AFR							
	MEA, OTH, IND, AFR							
	CHN, EUA, IND, AFR							
	CHN, OTH, IND, AFR							
	EUA, OTH, IND, AFR							
				Number of Internally Stable Coalitions				
	145	38	38	72	55	56	42	11
				Number of Externally Stable Coalitions				
	301	202	202	202	202	202	202	204

Table 3 further shows the number of internally and externally stable coalitions. From this it is possible to have a confirmation that the problem of internal stability is much more serious and detrimental to cooperation than the one of external stability, given that the numbers of internally stable coalitions are far lower. Despite not having being envisaged to tackle the problem of external stability, the RV scheme outperforms the others in this field too. A possible reason is its re-distributive nature that improves the payoffs of the outsiders as well. However, its real strength is in fostering internal stability, generally allowing a number of stable coalitions that is more than double the one allowed by the other transfers schemes. Another interesting fact to note is that the results are quite similar in all the considered time periods. This implies that the game and the RV transfers schemes are quite robust to significant perturbations in the values of the payoffs parameters.

Despite the several mentioned desirable features shown by the RV transfers scheme in this comparative analysis, that render this scheme ideal for an eventual IEA, there is also a major drawback. The number, and particularly the size, of fully stable coalitions is not so impressive. Even the number of internally stable coalitions, considering that, given 12 players, there are 4095 possible non-empty coalitions, is far from being astonishing. The main problem rests on the possibility to have a positive surplus in terms of payoffs. Without having it, no transfers scheme can solve the cooperative dilemma. Although the RV transfers scheme does not allow for stable coalitions with more than one third of the total players, still this result should not be underestimated. A coalition among USA, China, India and Africa, for example, that under the RV transfers scheme is stable both in 2015 and in 2025, brings together the two most polluting countries in the world, along with a fast rising polluter as India and the most damaged region in the world. When compared with the highest attainable level of public good provision, this coalition can still provide the 35% of it.

Table 4: Transfers Payments and $\boldsymbol{\theta}$ Vector Elements for coalition {EU, CHN, IND, AFR}

Countries	**C_GDP**	**C_EM**	**I_GDP**	**I_EM**	**EQ_SH**	**EQ_pcSH**	**RV**
			Transfers of Money (billions of \$)}				
EU	-1.545	-1.906	-2.121	-1.890	-1.834	-2.004	-2.174
CHN	0.548	0.855	0.286	0.189	0.429	0.550	0.090
IND	0.663	0.681	0.946	0.877	0.826	0.901	0.486
AFR	0.334	0.370	0.889	0.823	0.579	0.553	1.597
			Elements of the $\boldsymbol{\theta}$ *vector*				
EU	0.463	0.197	0.039	0.209	0.250	0.125	0
CHN	0.338	0.564	0.144	0.073	0.250	0.339	0
IND	0.130	0.143	0.338	0.288	0.250	0.306	0
AFR	0.069	0.096	0.478	0.430	0.250	0.231	1

$\alpha = 0.833, \beta = 0.233$, year=2025

Table 4 shows, for exemplifying purposes, the transfers among the coalition members and the elements of the transfers vector related to coalition $\{EU, CHN, IND, AFR\}$. This is a stable coalition under the RV scheme but not under the other transfers schemes. The transfer is here defined as the difference with the payoff a player would obtain in absence of any transfer, and, therefore, it includes both the distribution of the surplus and the difference between the cooperative and the outsider payoff. Note that the RV scheme is clearly the most re-distributive, granting all the surplus to Africa. This happens for several other coalitions of Table 3. Furthermore, under the RV scheme, EU is obliged to pay the most, whereas Chine remains an an almost neutral condition. We further have the confirmation that the second most re-distributive scheme is the inverse per-capita GDP.

As a further sensitivity analysis, also the values of α and β have been varied, using the data related to year 2025 for determining the payoffs parameters. Table A1, in Appendix, reports the results, from which it is possible to observe a substantial similarity with the results of Table 3. In the first simulation, the value of α has been kept constant and 0.833, whereas the one of β has been doubled to 0.567. This has reduced the number of internally stable coalitions for all transfers schemes and the number of fully stable coalitions for the RV scheme. Notice, instead, how the number of externally stable coalitions is increased. Another interesting fact to note is that a higher level of dis-utility towards advantageous inequality seems to favor that transfers schemes that are less re-distributive such as Current GDP and Current emissions. The opposite occurs in the next simulation, where β is kept equal to 0.288 and the value of α is halved. The number of internally stable coalitions is slightly increased and the re-distributive transfers schemes gain in ability to stabilize coalitions. The RV transfers scheme, instead, is scarcely sensitive to these modifications. In the last simulation, the value of α is slightly increased to 0.95 and the one of β is halved to 0.144. Again, the re-distributive transfers schemes improve their performance whereas there is a general drop in the number of externally stable coalitions. From this short analysis emerges that, for increasing values of the aversion towards advantageous inequality, less re-distributive transfers schemes perform better than the more re-distributive ones, while the opposite holds for increasing values of α. The RV transfers scheme, despite surely belonging to the set of re-distributive schemes, appears to be quite insensitive to changes in these two parameters.

In the set of simulations described in the previous paragraph, taking as reference the median values of α and β, we have varied them significantly in order to test the sensitivity of the various transfers schemes. By halving and doubling the value of β, we have covered almost the whole spectrum of possible values as provided in Fehr and Schmidt (1999), that, in fact, ranges from 0 to 0.6. The domain of values provided by the same authors for α, instead, is much wider, being included between 0 and 4. Whereas β can be considered as the real “altruistic” parameter, since it governs the intensity of dis-utility for advantageous inequality, α is more “selfish” since it causes dis-utility when a player is in an inferior position towards the others. It is significant, and quite plausible, therefore, that the latter may be greater than the former. In our previous analysis we have missed the investigation of more extreme values of α. Filling this gap is the objective of this last set of simulations, in which the value of β is kept constant at 0.233 and the one of α varies: 1.5, 2, 3 and 4. Table 4 reports the results of the analysis, with only the fully stable coalitions under the RV transfers scheme being shown[7] and using the data related to year 2015 for the other parameters. The other transfers schemes, in fact, are not very sensitive to large increases in the values of α, and the number and composition of the fully stable coalitions they allow are not dramatically different from the ones portrayed in Table 3 and A1[8]. The changes for the RV transfers scheme, instead, are very significant and interesting. Table 5 shows the fully stable coalitions, or a sample of them when their number is too large, along with the percentage of the public good contribution over the best attainable level reached by their formation. The ranking value of each coalition – in ascending order – in terms of public good contribution and in terms of total utility are further shown.

The most immediate and surprising result that emerges from Table 5 is that the RV transfers scheme, for certain values of α, is able to stabilize several coalitions comprising a very high number of players, namely 9. This happens for a value of α equal to three, that, among the considered values, is the one that maximizes both the number of stable coalitions (104) and their maximal size[9]. From

[7]Results related to the other transfers schemes are available from the authors upon request.

[8]Under $\alpha = 3$, when the RV scheme allows for 104 fully stable coalitions, the best of the other schemes, namely, the Inverse per-capita GDP, allows for 7 stable coalitions with maximal size of 3 players.

[9]There are 32 fully stable coalitions with 9 players for $\alpha = 3$, of which only half are reported in Table 5 due to reasons of space.

Table 5: Stable Coalitions under the RV Transfers Scheme for Varying Levels of α

Fully stable coalitions	**Percentage of public good production (%)**	**Ranking in public good production**	**Ranking in total utility**
Year = 2015; $\beta = 0.233; \alpha = 1.5$			
RUS, EUA	35.6	738	227
JPN, CHN, IND, AFR	38.6	1553	672
EU, CHN, IND, AFR	38.5	1545	563
RUS, CHN, IND, AFR	39.2	1760	740
MEA, CHN, IND, AFR	38.5	1528	487
MEA, OTH, IND, AFR	38.0	1382	432
LAT, CHN, IND, AFR	38.7	1594	531
LAT, OTH, IND, AFR	38.2	1438	466
CHN, OTH, IND, AFR	38.8	1629	442
US, JPN, EU, RUS, LAT, EUA	47.7	3154	3117
US, JPN, RUS, MEA, LAT, EUA	47.2	3126	3071
Year = 2015; $\beta = 0.233; \alpha = 2$			
45 fully stable coalitions (only the largest, with 6 players, are here reported)			
OHI, JPN, EU, RUS, LAT, EUA	50.8	3436	3252
OHI, JPN, RUS, MEA, LAT, EUA	50.4	3400	3222
OHI, EU, RUS, MEA, LAT, EUA	53.8	3651	3398
OHI, RUS, MEA, LAT, EUA, OTH	58.8	3840	3572
Year = 2015; $\beta = 0.233; \alpha = 3$			
104 fully stable coalitions (only half of the largest, with 9 players, are here reported)			
US, JPN, EU, RUS, MEA, LAT, CHN, EUA, OTH	95.4	4006	3612
US, JPN, EU, RUS, MEA, LAT, EUA, OTH, IND	95.3	4000	3566
US, JPN, EU, RUS, MEA, LAT, EUA, OTH, AFR	95.3	3998	3562
US, JPN, EU, RUS, MEA, LAT, EUA, IND, AFR	95.4	4007	3618
US, JPN, EU, RUS, MEA, EUA, OTH, IND, AFR	95.8	4021	3704
US, JPN, EU, RUS, LAT, EUA, OTH, IND, AFR	95.7	4018	3692
US, JPN, EU, MEA, LAT, CHN, EUA, OTH, IND	96.0	4043	3770
US, JPN, EU, MEA, LAT, CHN, EUA, OTH, AFR	96.0	4041	3765
US, JPN, EU, MEA, LAT, CHN, EUA, OTH, AFR	96.1	4056	3807
US, JPN, EU, MEA, LAT, CHN, EUA, IND, AFR	96.7	4081	3875
US, JPN, EU, MEA, LAT, CHN, OTH, IND, AFR	95.8	4022	3733
US, JPN, EU, MEA, LAT, EUA, OTH, IND, AFR	96.4	4074	3854
US, JPN, EU, MEA, CHN, EUA, OTH, IND, AFR	96.4	4070	3844
US, JPN, EU, LAT, CHN, EUA, OTH, IND, AFR	95.4	4003	3592
OHI, JPN, EU, RUS, MEA, LAT, CHN, EUA, OTH	95.5	4011	3647
OHI, JPN, EU, RUS, MEA, LAT, CHN, EUA, IND	95.5	4010	3643
OHI, JPN, EU, RUS, MEA, LAT, CHN, EUA, AFR	95.2	3996	3544
Year = 2015; $\beta = 0.233; \alpha = 4$			
US, MEA	46.8	2204	490
US, LAT	46.8	2054	467
US, CHN	47.2	3369	1019
US, OTH	47.0	2833	639
US, IND	47.1	3204	835
US, AFR	47.1	3183	819
IND, AFR	46.8	1851	119

The ranking is in ascending order, with the maximum being 4095 the statistics reported in Table 5, it is also possible to observe that these 9 players stable coalitions allows for substantial achievements in terms of public good production and total utility (this last defined as the sum of the utilities of all players in the game). It is also possible to observe that the number and the size of stable coalitions supported by the RV transfers scheme appear to be a growing function of the α parameter. For $\alpha = 1.5$, in fact, we have a number of stable coalitions equal to the case of $\alpha = 0.833$, but their maximal size is increased of two players. When switching to $\alpha = 2$, the maximal size remains the same as when $\alpha = 1.5$, but it is the number of stable coalitions to increase significantly till 45. The passage to $\alpha = 3$ entails a substantial growth on both sides. The mentioned relation, however, is not linear, since the extreme value of $\alpha = 4$ significantly reduces both the number (7 coalitions) and the size (2 players) of stable coalitions supported by the RV scheme.

It is not an easy task to understand the reasons behind the non-monotonic relation between the value of α and the number and size of stable coalitions allowed by the RV scheme. The disadvantageous inequality parameter, in fact, is present on the analytic solution from which the RV transfers is derived, on the analytic solution for determining the optimal q and on the final utility of players that is used to determine stability. The result of their interaction, therefore, is quite complicated to disentangle. At first sight, the result is rather counterintuitive, particularly if compared with observation 1 in Vogt (2016), according to which increasing values of α are detrimental for cooperation, in opposition to increasing values of β (observation 2). However, also in Vogt (2016) it is recognized that the relation is rather complex with some coalitions being internally stable for high values of α but not for lowers, everything else being equal. The condition for a non-zero transfer, namely $\frac{\alpha_j}{\alpha_j+\beta_j} > \sum_{k \in P_j} \theta_k$, is likely to play a major role. Given a constant β, in fact, it is more likely to have non-zero transfers. This allows for more balanced transfers rather than having just the weakest player to get all the surplus, but still leaving the RV scheme as the most re-distributive. This allows to sustain large coalitions, but, for very high levels of α, its detrimental effect described in observation 1 of Vogt (2016) prevails.

In conclusion, the RV transfers scheme is quite robust to perturbations of the parameters shaping players' payoffs and to variations of β. It is also robust to mild modifications to α, allowing to obtain significant, but not astonishing improvements in cooperation. It is, instead, rather sensitive to large variations of α, differently from the other transfers schemes. In particular, the number and the maximal size of stable coalitions is, till certain levels, a growing function of α. Only at very extreme levels its role becomes deleterious, sensibly reducing both the number and the size of stable coalitions. Despite values of α above 2 may be unlikely, there is still the possibility to obtain substantial levels of cooperation provided an opportune transfers scheme such as the one here proposed is implemented.

6 Conclusions

Obtaining significant levels of cooperation in public good and environmental games is a fascinating but often frustrating task. Under-provision of the public good due to free-riding incentives and, consequently, instability of large coalitions, is the most common result of this typology of games. This is particularly true when standard assumptions about players' behavior, namely perfect rationality and selfishness, are adopted. The latter assumption, however, has found several critiques by the behavioralist school on the base of a vast amount of experimental results, leading to the formulation of alternative utility functions that incorporate fairness and relational considerations. The use of the behavioral insights and, specifically, of F&S utility functions into public good and environmental games has contributed to improve the results by increasing the number of stable coalitions, but without causing a great leap forward.

A possibility for further improving the level of cooperation in the mentioned games has been seen in adopting opportune transfers schemes, in order to minimize the detrimental effects of free riding and, thus, fostering internal stability. The present paper has proposed the RV transfers scheme and it has analyzed its performance by benchmarking it with other transfers schemes popular in the domain of IEAs, through an environmental game parameterized on the RICE model. The scheme is based on payoffs transfers and is envisaged under the assumption that players have F&S utility functions. From one side it tries to guarantee internal stability by granting to each coalition member, whenever it is possible, the payoff she would obtain by being the only one to leave the coalition. On the other side, it distributes the eventual remaining surplus among the coalition members in such a way that the sum of their utilities is maximized.

The simulation exercise has confirmed the usefulness ot the RV transfers scheme in improving the possibility of cooperation by stabilizing coalitions, compared to the absence of transfers. It has also shown the superiority of the proposed scheme compared to the other candidates used as touchstone, that rarely provide significant improvements in cooperation. As a sensitivity analysis, payoffs have been computed using parameters relative to different time periods, namely 2015, 2025 and 2035. Furthermore, the key parameters of the F&S utility function regulating the intensity of suffered dis-utility due to disadvantageous and advantageous inequality have been varied. For mild variations in both of them around their median values, the RV scheme has confirmed a substantial lack of variation, offering improvement in cooperation that are significant but far from astonishing. However, for larger values of the disadvantageous inequality parameter, coalitions including even three quarters of all players have resulted as stable, providing significant improvements in total welfare. The performances of the RV transfers scheme seems to increase for growing values of the disadvantageous inequality parameter, but this is true only till a certain point over which the performances are drastically reduced. There appears to be, therefore, a non linear relation between the two. Once accepted the assumption that players have F&S preferences, the paper has shown how the RV transfers scheme is a valid instrument to foster cooperation. This may be an important result, especially for the current debate about climate change and emissions reduction.

Apart from the possibility of adopting the proposed transfers scheme in negotiations that, however, would require an estimation of the inequality aversion parameters for all participant countries, a more general and easy to use lesson that can be learned from the present paper is the better capability of re-distributive transfers scheme to foster stability. This resulted in almost all our simulations where, after the RV scheme, the second best performing transfers have been guaranteed by the inverse per-capita GDP. The only exception is, counter-intuitively, for a high value of the parameter representing the dis-utility for advantageous inequality. Besides calling for the need of empirical estimations of the parameters describing the intensity of dis-utility due to inequality in international negotiations, the paper suggests a preference for re-distributive transfers schemes in order to foster stability.

References

Bakalova, I. and Eyckmans, J. (2019). Simulating the impact of heterogeneity on stability and effectiveness of international environmental agreements. *European Journal of Operational Research*, 277(3):1151–1162.

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

Blanco, M., Engelmann, D., and Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, 2(72):321–338.

Bolton, G. E. and Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American economic review*, 90(1):166–193.

Bolton, G. E. and Ockenfels, A. (2005). A stress test of fairness measures in models of social utility. *Economic Theory*, 25(4):957–982.

Botteon, M. and Carraro, C. (1997). Environmental Coalitions with Heterogeneous Countries: Burden-Sharing and Carbon Leakage. *Fondazione Eni Enrico Mattei Working Paper*, (24.98).

Breton, M. and Garrab, S. (2014). Evolutionary farsightedness in international environmental agreements. *International Transactions in Operational Research*, 21(1):21–39.

Carraro, C., Eyckmans, J., and Finus, M. (2006). Optimal transfers and participation decisions in international environmental agreements. *The Review of International Organizations*, 1(4):379–396.

Carraro, C. and Siniscalco, D. (1993). Strategies for the International Protection of the Environment. *Journal of public Economics*, 52(3):309–328.

Chander, P. and Tulkens, H. (1995). A Core-theoretic Solution for the Design of Cooperative Agreements on Transfrontier Pollution. *International Tax and Public Finance*, 2(2):279–293.

Dannenberg, A., Sturm, B., and Vogt, C. (2010). Do equity preferences matter for climate negotiators? An experimental investigation. *Environmental and Resource Economics*, 47(1):91–109.

d'Aspremont, C., Jacquemin, A., Gabszewicz, J. J., and Weymark, J. A. (1983). On the stability of collusive price leadership. *Canadian Journal of economics*, pages 17–25.

Diamantoudi, E. and Sartzetakis, E. S. (2006). Stable international environmental agreements: An analytical approach. *Journal of public economic theory*, 8(2):247–263.

Diamantoudi, E. and Sartzetakis, E. S. (2015). International environmental agreements: coordinated action under foresight. *Economic Theory*, 59(3):527–546.

Eyckmans, J. and Tulkens, H. (2006). Simulating coalitionally stable burden sharing agreements for the climate change problem. In *Public goods, environmental externalities and fiscal competition*, pages 218–249. Springer.

Fehr, E. and Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly journal of Economics*, pages 817–868.

Fehr, E. and Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism–experimental evidence and new theories. *Handbook of the economics of giving, altruism and reciprocity*, 1:615–691.

Finus, M. (2008). Game theoretic research on the design of international environmental agreements: insights, critical remarks, and future challenges. *International Review of environmental and resource economics*, 2(1):29–67.

Finus, M., Cooper, P., and Almer, C. (2017). The use of international agreements in transnational environmental protection. *Oxford Economic Papers*, 69(2):333–344.

Fischbacher, U. and Gachter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American economic review*, 100(1):541–56.

Güth, W., Kliemt, H., and Ockenfels, A. (2003). Fairness versus efficiency: An experimental study of (mutual) gift giving. *Journal of Economic Behavior & Organization*, 50(4):465–475.

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of economic behavior & organization*, 3(4):367–388.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of business*, pages S285–S300.

Lange, A. (2006). The impact of equity-preferences on the stability of international environmental agreements. *Environmental and Resource Economics*, 34(2):247–267.

Lange, A. and Vogt, C. (2003). Cooperation in International Environmental Negotiations Due to a Preference for Equity. *Journal of Public Economics*, 87(9):2049–2067.

Lange, A., Vogt, C., and Ziegler, A. (2007). On the Importance of Equity in International Climate Policy: An Empirical Analysis. *Energy Economics*, 29(3):545–562.

McGinty, M. (2007). International environmental agreements among asymmetric nations. *Oxford Economic Papers*, 59(1):45–62.

McGinty, M., Milam, G., and Gelves, A. (2012). Coalition stability in public goods provision: testing an optimal allocation rule. *Environmental and Resource Economics*, 52(3):327–345.

Nordhaus, W. D. (2011). Estimates of the social cost of carbon: background and results from the RICE-2011 model. Technical report, National Bureau of Economic Research.

Nordhaus, W. D. and Yang, Z. (1996). A regional dynamic general-equilibrium model of alternative climate-change strategies. *The American Economic Review*, pages 741–765.

Pavlova, Y. and De Zeeuw, A. (2013). Asymmetries in international environmental agreements. *Environment and Development Economics*, 18(1):51–68.

Rogna, M. (2016). Cooperative Game Theory Applied to IEAs: A Comparison of Solution Concepts. *Journal of Economic Surveys*, 30(3):649–678.

van der Pol, T., Weikard, H.-P., and van Ierland, E. (2012). Can altruism stabilise international climate agreements? *Ecological Economics*, 81:112–120.

Vogt, C. (2016). Climate coalition formation when players are heterogeneous and inequality averse. *Environmental and Resource Economics*, 65(1):33–59.

Weikard, H.-P., Wangler, L., and Freytag, A. (2015). Minimum participation rules with heterogeneous countries. *Environmental and Resource Economics*, 62(4):711–727.

Appendix

Table A1: Stable Coalitions under Varying levels of α and β

	RV	C_GDP	C_EM	I_GDP	I_EM	EQ_SH	EQ_pcSH	No Transfer
2025, $\alpha = 0.833, \beta = 0.576$	*Fully Stable Coalitions*							
	JPN, CHN, IND, AFR EU, CHN, IND, AFR RUS, CHN, IND, AFR LAT, CHN, IND, AFR CHN, EUA, IND, AFR CHN, OTH, IND, AFR EUA, OTH, IND, AFR	CHN, IND	CHN, IND	–	–	–	CHN, IND	CHN, IND IND, AFR
	Number of Internally Stable Coalitions							
	120	22	21	32	26	28	23	8
	Number of Externally Stable Coalitions							
	923	665	665	668	663	663	665	672
2025, $\alpha = 0.4165, \beta = 0.233$	*Fully Stable Coalitions*							
	EU, CHN, IND US, CHN, IND, AFR OHI, CHN, IND, AFR JPN, CHN, IND, AFR RUS, CHN, IND, AFR LAT, CHN, IND, AFR MEA, CHN, IND, AFR MEA, OTH, IND, AFR CHN, OTH, IND, AFR CHN, EUA, IND, AFR	–	–	CHN, OTH, IND CHN, IND, AFR	CHN, OTH, IND CHN, IND, AFR	CHN, OTH, IND	CHN, OTH, IND	CHN, IND
	Number of Internally Stable Coalitions							
	189	66	70	98	76	82	77	10
	Number of Externally Stable Coalitions							
	480	439	439	439	439	439	439	440
2025, $\alpha = 0.95, \beta = 0.144$	*Fully Stable Coalitions*							
	RUS, EUA US, CHN, IND, AFR OHI, CHN, IND, AFR JPN, CHN, IND, AFR RUS, CHN, IND, AFR LAT, CHN, IND, AFR MEA, CHN, IND, AFR EU, CHN, IND, AFR CHN, OTH, IND, AFR	–	–	CHN, EUA, IND MEA, CHN, IND	–	CHN, OTH, IND	CHN, OTH, IND	CHN, IND
	Number of Internally Stable Coalitions							
	173	58	66	155	74	123	80	12
	Number of Externally Stable Coalitions							
	335	191	191	192	191	191	191	192