Environ Resource Econ (2012) 52:327–345
DOI 10.1007/s10640-011-9530-6

# Coalition Stability in Public Goods Provision: Testing an Optimal Allocation Rule

**Matthew McGinty · Garrett Milam · Alejandro Gelves**

Accepted: 17 November 2011 / Published online: 13 December 2011


**Abstract** We test the canonical model of international environmental agreements (IEAs) in a laboratory setting with asymmetric agents. IEA participation represents coalition formation and public good provision where there are gains to cooperation, but an incentive to free-ride. We test four competing methods of dividing the coalition's worth: a recently proposed optimal rule which accounts for subjects' payoffs as a single free-rider, the Shapley value, the Nash bargaining solution, and an equal split. Each treatment generates the theoretically predicted coalition size more often than not. The shares of the potential gains to cooperation achieved by each rule are: 51, 36, 40 and 13%, respectively. These results highlight the importance of using an optimal rule to improve IEAs, and more broadly for voluntary public good provision.

**Keywords** IEAs · Coalition formation · Experimental economics · Public goods · Shapley value · Nash bargaining solution

**JEL Classification** H41 · C92 · C72 · C78 · D63 · D64 · D74

## 1 Introduction

Coalition stability depends on the method of dividing the coalition's worth and on each member's payoff outside the coalition. Both the worth and the outside payoff vary when

M. McGinty (✉)
Department of Economics, University of Wisconsin-Milwaukee, PO Box 413, Milwaukee, WI 53201, USA
e-mail: mmcginty@uwm.edu

G. Milam
Department of Economics, University of Puget Sound, Tacoma, WA, USA

A. Gelves
Department of Economics, Midwestern State University, Wichita Falls, TX, USA

members have different characteristics, or when externalities are present. Thus, traditional sharing rules used in the international environmental agreement (IEA) literature, such as the Shapley value and the Nash bargaining solution, have different implications for stability among asymmetric players (Barrett 1997; Botteon and Carraro 2001). The Shapley value and the Nash bargaining solution do not account for the interaction between asymmetry and the outside payoff for a coalition member. Recently, optimal allocation rules have emerged that directly address the outside payoff and theoretically show that the set of stable coalitions is larger than previously thought (McGinty 2007; Weikard 2009). Yet, these rules have not been taken into the laboratory and tested. This paper tests the sharing rule proposed in McGinty (2007) to see if, indeed, IEA participation is greater than using the Shapley value, the Nash bargaining solution, and an equal split of the coalition's worth.

An important application of coalition formation and stability with asymmetry involves IEAs, such as the Kyoto treaty on greenhouse gas emissions. The theoretical literature on IEAs shows participation by individually rational nations is trivial when there are significant gains to cooperation. Barrett (1994) shows only three out of a hundred nations will participate in an IEA when the marginal benefit and marginal cost of abatement have slopes of equal magnitude. In response, there is a growing body of applied IEA literature which examines coalition stability under different rules of coalition formation and division of worth[1] (Finus 2003; Weikard and Dellink 2010; Weikard et al. 2010). This paper presents an experimental test of four methods of dividing a coalition's worth using payoffs generated by the canonical model of IEAs (Carraro and Siniscalco 1993; Barrett 1994). We show that subjects form larger coalitions under an optimal sharing rule.

The Shapley value and the Nash bargaining solution are appropriate for convex games where the grand coalition (consisting of all players) is stable and there are no externalities. However, abatement of pollutants under IEAs are public goods, and as such they typically suffer from the free-rider problem. The other $n - 1$ nations are the primary beneficiaries of a nation's costly abatement. Given the positive externalities associated with abatement, traditional burden sharing rules fare poorly in the context of IEA stability. Typically, in these types of games there are gains to cooperation and the grand coalition is Pareto optimal. However, as Barrett (1997) and Botteon and Carraro (2001) show, it is not internally stable when there are significant gains to cooperation since at least one member earns a greater payoff as a free-rider. As a result there is an incentive for nations to deviate, and coalitions break down. Barrett (1997) obtains a maximum stable coalition of only three members (out of seven) using the Shapley value to determine a zero-sum system of transfers among members.

McGinty (2007) obtains more encouraging results by addressing IEA stability via an optimal rule. This identifies an IEA as internally stable when there is a positive surplus, meaning the worth of a coalition is at least as great as the sum of payoffs that each member would earn as a single defector (Weikard 2009; Weikard et al. 2010).[2] When there is a positive surplus a zero-sum system of transfers, implemented in the form of abatement requirements and tradable pollution permits, exists such that no member has an incentive to free-ride.[3] Each

[1] Methods of dividing a coalition's worth are often referred to as burden sharing rules in public goods games. These burden sharing rules implicitly define a zero-sum system of transfers among coalition members.

[2] Similar allocations with at least the outside payoff have recently emerged. These have different names in the literature. Weikard (2009) has a "claims rights condition", Carraro et al. (2006) refer to a "potentially internally stable coalition," and McGinty (2011) shows that an equal share of the surplus "risk-dominates" all other allocations.

[3] For a given coalition, actual abatement is the same for every allocation rule since the worth is a result of minimizing the coalition's aggregate abatement cost. This requires equating the marginal abatement cost of the last unit for all coalition members. However, the rules differ in how they allocate the worth, even though

member earns her payoff outside a given coalition, plus a share of the surplus determined by a positive parameter $\theta$. This $\theta$ rule explicitly addresses the relevant alternative payoff in the presence of externalities, which the Shapley value and Nash bargaining solution (NBS) do not. McGinty (2007) shows stable coalitions defined in this fashion can obtain 72% of the aggregate payoff difference between the grand and singleton coalitions when there are twenty fully asymmetric nations. By contrast, with symmetric nations only 9% of this difference is obtained by the largest stable coalition.

This paper presents a controlled laboratory test of coalition stability comparing the $\theta$ rule to the Shapley value, the Nash bargaining solution, and an equal split of the coalition's worth. The coalition's worth is not arbitrarily chosen, but rather given by the benchmark model of IEAs (Carraro and Siniscalco 1993; Barrett 1994) in the presence of asymmetry (Barrett 1997; McGinty 2007). Previous coalition experiments focus on subjects' division of a coalition's worth and then compare actual behavior with competing theoretical predictions (Croson et al. 2004, Bolton et al. 2003, Leopold-Wildburger 1992). Instead, we focus on coalition stability by providing subjects with the division of the coalition's worth and their payoff outside a given coalition. Our approach eliminates uncertainty with respect to the bargaining process and reduces subjects' choice to either being in or out of a given coalition. The typical coalition experiment essentially assumes the grand coalition will form. By contrast, we look at coalition stability in the presence of a free-rider incentive since the grand coalition is not stable in our experiment. Rather than measure the distance between the actual outcomes and traditional solution concepts as Croson et al. (2004) we impose the solution concept and address coalition formation and stability in isolation. In this respect our paper answers Bolton (1998) call for subsequent laboratory research linking bargaining and dilemma games.

We find coalition membership is broadly consistent with the theoretical predictions across treatments. In our four player game, the $\theta$ rule results in the predicted three member coalitions 56% of the time, whereas Shapley and NBS result in the predicted two member coalitions 57 and 59%, respectively. No coalition formed 64% of the time in the equal split treatment. Our performance metric is the proportion of the payoff gap that is closed by coalition formation. Efficiency is defined as the proportion of the potential aggregate payoff difference between the grand and singleton coalitions. The $\theta$ rule results in the largest average efficiency of 51%, compared to 36% for Shapley and 40% for NBS. Efficiency for the equal split treatment was only 13%. These results highlight the importance of using the appropriate abatement requirements, implicitly defining a burden sharing rule, for the Kyoto treaty.

Our results are consistent with recent work on IEAs and coalitions. Dannenberg et al. (2010) conduct an IEA experiment with symmetric subjects. Under three different institutional arrangements subjects choose both coalition membership and a level of public good provision. As predicted by theory, they find that only small coalitions results when members are forced to choose the coalition's optimal abatement level. They find larger coalitions emerge when abatement requirements only partially internalize the positive externalities from abatement. Carraro et al. (2006), McGinty (2007) and Weikard (2009) show that coalition membership is largest when players' receive an allocation that exceeds the payoff that would be earned from individually leaving the coalition. McGinty (2011) shows that a risk-dominant

Footnote 3 continued
each rule implies a zero-sum system of transfers. One transfer method is a system of tradable pollution permits, where the permit price $p$ is the common marginal abatement cost of the last unit. Under a permit system, coalition member $s$ is required to abate an amount $q_s^r$. Actual abatement is $q_s$, thus the transfer $\tau$ for member $s$ in coalition $k$, is: $\tau_s(k) = p(q_s - q_s^r)$. McGinty (2007) and Sect. 5.2 of McGinty (2011) provide additional details.

optimal allocation rule may result in larger stable coalitions than the Shapley value or the Nash bargaining solution. An example applies this risk-dominant rule to the fully asymmetric IEA model in McGinty (2007), and details the associated abatement requirements with transfers implemented via a system of tradable pollution permits.

The remainder of the paper is organized as follows. The theoretical model is presented in Sect. 2, detailing the underlying model, definitions of stable coalitions and the four treatments. Section 3 presents the experimental environment describing the data set, experimental procedures and theoretical predictions. Descriptive and regression results from fixed effects models are presented in Sect. 4. Section 5 concludes and suggests further research.

## 2 Theoretical Model and Treatments

Following Barrett (1994, 1997) and Carraro and Siniscalco (1993) we implement the standard public goods model from the IEA literature, but allow for fully asymmetric players as in McGinty (2007). The set of players is denoted $N = \{1, \ldots, n\}$. Social benefit is a concave function $B(Q) = b\left(aQ - \frac{Q^2}{2}\right)$ of the total quantity of the public good, $Q = \sum_{i \in N} q_i$. Parameters $a$ and $b$ are assumed to be strictly positive resulting in declining marginal benefit. Benefit shares are given by $\alpha_i$, where $\alpha_i > 0\, \forall i \in N$ and $\sum_{i \in N} \alpha_i = 1$. The benefit for player $i$ is: $B_i(Q, \alpha_i) = b\alpha_i \left(aQ - \frac{Q^2}{2}\right)$. Players have convex cost functions $C_i(q_i, c_i) = \frac{c_i q_i^2}{2}$, generating marginal cost curves that are rays from the origin with asymmetric slopes $c_i > 0$. In the absence of transfers the payoff for player $i$ is: $\pi_i = B_i(Q, \alpha_i) - C_i(q_i, c_i)$.

Variations of this model appear in the experimental literature using the voluntary contribution mechanism (VCM) where subjects contribute a portion of their endowment to a public good (see Ledyard 1995; Laury and Holt 2005 for surveys). Several VCM experiments test either declining marginal benefit, increasing marginal cost (Isaac et al. 1985; Keser 1996) or asymmetric public and private good values (Burlando and Guala 2005; Chan et al. 1999). Two stylized facts emerge from this literature. First, subjects systematically over-contribute to public goods beyond the individually rational level. Second, this over-contribution decays as the experiment progresses. Our experimental environment examines an alternative model of public goods provision with less uncertainty regarding others' actions and a binary decision space. However, our design features an incentive structure analogous to VCM games which allows examination of the issues of over-contribution and time-decay.

We begin by identifying the singleton and grand coalition outcomes. The singleton coalition, denoted $\{i\}$, is obtained by each player acting in his own self-interest choosing $q_i$ to maximize $\pi_i$. The singleton individual and group levels of public good provision are:

$$
\begin{aligned}
q_i^{\{i\}} &= \frac{ab\theta_i}{1 + b\sum_{j=1}^{n}\theta_j} \quad \forall i \in N \\
Q^{\{i\}} &= \frac{ab\sum_{j=1}^{n}\theta_j}{1 + b\sum_{j=1}^{n}\theta_j}
\end{aligned} \tag{1}
$$

where $\theta_i \equiv \frac{\alpha_i}{c_i}$ is the ratio of benefit share to marginal cost slope. Each player provides the public good in proportion to their $\theta$ in the singleton outcome. The grand coalition (social optimum) maximizes the sum of payoffs, $\Pi \equiv \sum_{i=1}^{n} \pi_i$, where each player internalizes the positive externality that accrues to others. The grand coalition, denoted $\{N\}$, results in the social optimum, $q_i^{\{N\}}$ and $Q^{\{N\}}$:

$$q_i^{\{N\}} = \frac{ab}{c_i\left[1 + b\sum_{j=1}^{n}\frac{1}{c_j}\right]} \quad \forall i \in N$$
$$Q^{\{N\}} = \frac{ab\sum_{j=1}^{n}\frac{1}{c_j}}{1 + b\sum_{j=1}^{n}\frac{1}{c_j}} \tag{2}$$

All coalitions are assumed to allocate public good provision across members to maximize the coalition's worth. This occurs where the marginal cost of the last units are equalized across coalition members. The division of the coalition's worth then entails a zero-sum system of transfers among members. In the context of IEAs these transfers may take the form of tradable pollution permits, equating the marginal cost of the last unit of provision with the permit price. Abatement requirements under the IEA relative to $q_i^{\{N\}}$ then determine the sign and magnitude of the transfer for each player. The solution to all other potential coalition structures is presented in McGinty (2007).

## 2.1 Coalition Stability

The central issue surrounds the conditions under which a given coalition achieves stability. The literature on IEAs uses internal and external stability from the cartel literature (d'Aspremont et al. 1983) to describe individual incentives. We model an open membership single coalition game (Finus 2003). Thus, existing members may not prevent new members from joining the coalition and each non-member acts as an individual. Let $k$ be an element of the set of all possible coalitions $K$. Also, let $k \diagdown \{i\}$ denote the remaining coalition when member $i$ leaves the coalition, and $k \cup \{j\}$ denote the resulting coalition when non-member $j$ joins. The subscripts $s$ and $t$ denote signatories and non-signatories (free-riders) choosing provision levels to maximize coalition worth and individual payoff, respectively. The following conditions from McGinty (2007) define individual and coalition stability.

$$\pi_s(k) - \pi_t(k \diagdown \{t\}) + \tau_s(k) \geq 0, \forall s = t \in k \tag{3}$$

$$\pi_t(k) - \pi_s(k \cup \{s\}) - \tau_s(k \cup \{s\}) > 0, \forall t = s \notin k \tag{4}$$

$$\Phi = \left\{k \in K : \sum_{s \in k}\pi_s(k) - \sum_{t \in k}\pi_t(k \diagdown \{t\}) \geq 0\right\} \tag{5}$$

The conditions for internal and external stability are given by Eqs. (3) and (4). Internal stability ensures each member at least as high a payoff choosing $q_s(k)$ with transfer $\tau_s(k)$ than as a single defector from the coalition choosing $q_t(k \diagdown \{t\})$. External stability indicates that no non-member earns a higher payoff joining coalition $k$ and choosing $q_s(k \cup \{s\})$ with transfer $\tau_s(k \cup \{s\})$. Equation (5) defines the set of all stable coalitions: $\Phi \subset K$, given the appropriate (zero-sum) transfers. This set consists of members such that the coalition's worth, $v(k) \equiv \sum_{s \in k}\pi_s(k)$, exceeds the sum of payoffs that each member would earn individually leaving the coalition. For all elements $\phi \in \Phi$ there exists a zero-sum system of transfers that simultaneously satisfies the internal stability condition for all members $s \in k$.

To illustrate the central issues of superadditivity and coalition stability consider the following three player characteristic function game with positive externalities from Croson et al. (2004). While the payoffs are not generated by an underlying model, the structure is similar to the public goods game tested in the present paper.

$$\begin{aligned} v\{i\} &= 50, i = A, B, C \\ v\{AB\} &= 240, \pi_C(\{AB\}) = 140 \\ v\{AC\} &= 210, \pi_B(\{AC\}) = 150 \\ v\{BC\} &= 180, \pi_A(\{BC\}) = 180 \\ v\{ABC\} &= 400 \end{aligned} \tag{6}$$

Each row represents one of the five possible coalition structures. The first row is the singleton payoffs, $v\{i\} = 50$, $i = A, B, C$. The worth of the two member coalitions are in rows two through four, with the payoff of the player outside the coalition denoted $\pi_t$, $t = A, B, C$. The worth of the grand coalition in row five. Note, $\pi_t$ is the payoff that a player would earn if they were to leave the grand coalition. This game is superadditive since the worth of each coalition exceeds the sum of members worth in all sub-coalitions. Formally, strict superadditivity is defined as: $v\{k \cup t\} > v\{k\} + v\{t\}$. For the two member coalitions $v\{ij\} > v\{i\} + v\{j\}$, $i, j = A, B, C$ all distinct. The grand coalition: $v\{ABC\} > v\{ij\} + \pi_h(\{ij\})$ $h, i, j = A, B, C$ all distinct. The two member coalitions show the positive externalities since the player outside the coalition benefits from the coalitions formation: $\pi_h(\{ij\}) > v\{h\}$, $h, i, j = A, B, C$ all distinct.

The grand coalition is socially optimal, however it is not internally stable since the game is concave. The worth of the grand coalition does not exceed the sum of the free-rider payoffs $v\{ABC\} < \pi_A + \pi_B + \pi_C$ indicating (5) is not satisfied. There exists no division of 400 such that all three players simultaneously earn a payoff greater than they could earn as a single free-rider. However, the two member coalitions are (potentially) internally stable since there exists a division of the coalition's worth where no member has an incentive to leave. The singletons are not externally stable since internally stable two member coalitions can form.

Surprisingly, Croson et al. (2004) find the grand coalition forms 91% of the time, even though it is not internally stable. Several possible explanations come to mind. First, subjects may be systematically over-optimistic. Each player may anticipate their own earnings to be greater than their outside payoffs, even if they recognize that this is not simultaneously possible for all coalition members. Second, subjects may have been conditioned to form the grand coalition since it was stable in all other treatments. Third, subjects may have exhibited some form of altruism.

## 2.2 Treatments

The system of transfers defines the division of the coalition's worth. We test four methods of distribution: the $\theta$ rule, the Shapley value, the NBS, and an equal split. Internal and external stability depends on the system of transfers. With an inappropriate system of transfers a potentially stable coalition may violate internal stability for at least one member. The worth of each coalition, each members payoff outside that coalition, $\pi_t(k \diagdown \{t\}$ for $t \in k$, and each non-members payoff, $\pi_t(k)$ for $t \notin k$, are invariant across treatments.

Satisfying condition (5) indicates that the worth of the coalition is sufficient to make membership incentive compatible for all members simultaneously. The issue is whether the transfer scheme is sufficient to satisfy the condition for internal stability in (3 ). Under the $\theta$ rule each member receives the payoff they would earn outside the coalition and any remaining surplus is distributed by $\theta$ share. Each coalition member receives:

$$\pi_s^\theta(k) = \pi_t(k \diagdown \{t\}) + \frac{\theta_s}{\sum_{s \in k} \theta_s} \left[ v(k) - \sum_{t \in k} \pi_t(k \diagdown \{t\}) \right] \quad \forall s = t \in k \tag{7}$$

Botteon and Carraro (2001) and Barrett (1997) have used rules from cooperative game theory to distribute a coalition's worth in the context of IEAs. The Shapley value and Nash bargaining solution are unique payoff allocations, but they do not ensure internal stability (3), even if (5) holds. The Shapley value (SV) is:

$$\pi_s^{SV}(k) = \sum_{j \subseteq k} \frac{(|j|-1)!(|k|-|j|)!}{|k|!} [v(k) - v(k \backslash \{s\}] \quad \forall s \in k \tag{8}$$

where $|i|, i = j, k$ denotes the number of members (cardinality) in coalition $i$.

The Shapley value is the weighted sum of players' marginal contributions across all possible sub-coalitions $j$. The weights are binomial coefficients indicating the probability of each coalition formation sequence occurring, with the assumption that all sequences are equally likely. The Shapley value is unique and always exists, but there is no reason to expect that it, or the Nash bargaining solution, exceeds the free-rider payoff $\pi_t(k \diagdown \{t\})$ in a concave game. The Shapley value incorporates how much a player's inclusion adds to the worth of the coalition, but not how much a player earns outside of the coalition.

The Nash bargaining solution maximizes the product of differences from the singleton payoff. Assuming equal weights, the Nash bargaining solution (NBS) for each member $\pi_s^{NBS}$ is the solution to the maximization problem:

$$\pi_s^{NBS}(k) = \max \left\{ \prod_{s \in k} [v(k) - \pi_s(\{s\})] \right\} \tag{9}$$

The NBS does not account for neither the worth of the sub-coalitions of $k$, nor the payoff a player would earn outside the coalition in games with externalities.

Finally, the equal split simply divides the worth of the coalition by the number of coalition members $|k|$.

$$\pi_s^E(k) = \frac{v(k)}{|k|} \tag{10}$$

Inclusion of the equal split treatment serves two purposes. The primary motivation stems from the need for a clear benchmark. This allows us to verify that agents are responding across treatments to the individual incentives generated by the payoff structure. The secondary rationale concerns previous literature which investigates equal division of surplus in bargaining games. Bolton (1997) shows that splitting equally can be an evolutionarily stable equilibrium in a bilateral bargaining game. The idea of a "fair share" comes up frequently in follow-up questionnaires administered in bargaining experiments (Roth 1987). If these are important issues in our environment, the equal split treatment should help identify their influence.

## 3 Experimental Environment

### 3.1 Information and Data Set

Six laboratory sessions were conducted, three each at Ryerson University in Toronto and the University of Wisconsin-Milwaukee. Subjects were recruited from first-year undergraduate Economics courses. Subjects were not allowed to communicate with one another, nor did they know the identity of the other group members. Coalitions were referred to as subgroups to avoid any demand effects (Friedman and Sunder 1994). Thus, subjects did not know that the context of the experiment was IEAs, or even more broadly, public goods provision. Five sessions had twelve subjects, while one session was run only with eight due to no-shows. Subjects were randomly assigned a type ($A$, $B$ , $C$, or $D$) and a group consisting of one of each type. The Earnings Table in the "Appendix" shows the payoff for each type and allocation rule. Sessions consisted of four treatments each lasting ten periods, with three practice periods for the first treatment. Each subject participated in all four treatments, and the treatment order was randomly determined. Subjects were randomly assigned a new type and group at the end of each treatment. [4] The data set consists of 40 final observations for 68 subjects, resulting in 2, 660 observations.[5] All sessions were completed in under two hours. Subjects earned an average of $28, including the $5 show-up payment.

The experiment was programmed and conducted using the z-Tree software (Fischbacher 2007). Subjects were given the earnings table, showing payoffs for each group member for all permissible coalition structures, and the written instructions in the "Appendix". As in the earnings table, we adopt the shorthand notation $[A, B]$ or $AB$ to denote coalition $\{A, B\}$ for the remainder of the paper. Subjects interacted through a series of three different computer screens. The first screen indicated their type ($A$, $B$, $C$, or $D$) and asked if they wished to join or remain out of the subgroup. The second screen revealed all group member's initial decisions showing both the subject's payoff with that subgroup and the payoff if they reversed their decision, ceteris paribus. Subjects then either confirmed or reversed their decision. If any of the subjects reversed their decision the entire group was returned to the first screen (with subjects who previously chose to join the subgroup starting in, rather than out). This process repeated until all group members confirmed a particular subgroup, or time ran out. Period length was sixty seconds, plus zero to sixty seconds randomly drawn from a uniform distribution.[6] The third screen showed the final subgroup and the subject's earnings for that and all previous periods.

Given our primary emphasis is the formation and stability of coalitions we allowed subjects to reverse or confirm after observing other's decisions. To test for both internal and external stability we allowed those outside to join, or those inside to leave, an already established coalition. This necessitated the time constraint and uncertain period length given the structure of the game. For example, with few exceptions, each subject's most preferred outcome is to be outside a three member coalition.[7] This means subjects have an incentive to try promote the grand coalition and then leave at the last second. The uncertain period length made this type of strategy impossible to implement. However, to avoid repeated game strategies subjects knew that each run would consist of exactly ten periods. The Folk Theorem (Friedman 1971) states that the one-shot Nash equilibrium is obtained via backwards induction in repeated games with a known final period. Thus, the repeated interaction in the experiment has the same predictions as the one-shot game in the theoretical section.

[4] The matching procedure results in a "partial strangers" design.

[5] One of the $\theta$ treatment sessions ended after five periods due to a computer crash resulting in a loss of 60 observations.

[6] Monitors announced when the initial sixty seconds was up, indicating that the period could end at any time.

[7] The exceptions are: (i) subject A's most preferred coalition structure is $[A, B, D]$ for Shapley, (ii) subject A's most preferred coalition structure is $[A, B, C, D]$ for Nash bargaining and equal split, and (iii) subject B's most preferred coalition structure is $[B, C, D]$ for equal split.

**Table 1** Payoffs

| Coalition $k$ | $v(k)$ | Efficiency (%) | $\pi_t(k\diagdown\{t\})$ |
|---|---|---|---|
| $[A]$, $[B]$, $[C]$, $[D]$ | [30.7, 46.0, 61.4, 76.7] | 0 | na |
| $[A, B]$ | 77.7 | 28.5 | [30.7, 46.0] |
| $[A, C]$ | 94.3 | 24.3 | [30.7, 61.4] |
| $[A, D]$ | 111.3 | 37.6 | [30.7, 76.7] |
| $[B, C]$ | 108.5 | 21.6 | [46.0, 61.4] |
| $[B, D]$ | 124.7 | 23.5 | [46.0, 76.7] |
| $[C, D]$ | 139.3 | 16.2 | [61.4, 76.7] |
| $[A, B, C]$ | 150.7 | 83.0 | [32.9, 50.4, 66.1] |
| $[A, B, D]$ | 169.4 | 45.0 | [33.2, 51.4, 82.6] |
| $[A, C, D]$ | 186.7 | 46.9 | [32.9, 68.5, 84.1] |
| $[B, C, D]$ | 199.4 | 57.7 | [49.3, 66.5, 82.3] |
| $[A, B, C, D]$ | 255.6 | 100 | [39.0, 59.5, 79.1, 98.0] |

### 3.2 Predictions: Stability and Efficiency

In order to best examine the gains from coalition formation our design assigns to the highest (lowest) cost player the highest (lowest) benefit share.[8] This type of asymmetry increases the potential gains to coalition formation, by creating larger gains from trade in pollution permits. By contrast, the Nash equilibrium is closer to the social optimum when the high benefit nations are also low cost. This type of asymmetry makes the sum $\sum_{j=1}^{n} \theta_j$ larger than if the high benefit nations are high cost. The Nash equilibrium in Eq. (1) shows that aggregate abatement is increasing in $\sum_{j=1}^{n} \theta_j$. Since the social optimum in Eq. (2) is independent of the benefit shares, the difference between Nash and optimal abatement is greatest when the high benefit share countries are also high cost.[9] Table 1 shows the worth of each coalition $v(k)$, efficiency as the proportion of the payoff gap that is obtained and each members payoff should they individually leave the coalition $\pi_t(k\diagdown\{t\})$. Table 2 presents each member's payoff for all four treatments and indicates which coalitions are internally and externally stable. Subjects were only given the instructions and earnings Table 7 in the "Appendix".

The parameterization generates different theoretical predictions across the four treatments. The game is superadditive and the grand coalition is socially optimal. However, the grand coalition is unstable for all treatments since condition (5) is not satisfied: $v(N) = 255.6 < \sum_{t \in N} \pi_t(N\diagdown\{t\}) = 275.5$. Condition (5) is satisfied for all two and three member coalitions. Thus, all three member coalitions are both internally and externally stable for the $\theta$ treatment. However, none of the three member coalitions are internally stable in the other three treatments. The Shapley value and Nash bargaining solutions have identical stability predictions, even though the magnitude of the decisions at the margin are different. Under Shapley and NBS only coalition $AB$ is both internally and externally stable. For these two treatments all other two member coalitions are internally stable, but not externally stable. Thus, given two member coalitions that are not externally stable, we expect three member coalitions to form and then break down since the resulting three member coalitions are not internally stable. By contrast, under $\theta$ all three member coalitions are both internally and externally stable. No larger coalition is both internally and externally stable under equal split, thus the singleton outcome is the unique prediction. Some of the individual payoff differences between being

[8] The parameterization is: $n = 4$, $a = 50$, $b = 0.25$, $c = [\frac{2}{14}, \frac{3}{14}, \frac{4}{14}, \frac{5}{14}]$, $\alpha = [\frac{2}{14}, \frac{3}{14}, \frac{4}{14}, \frac{5}{14}]$, $\theta = [1, 1, 1, 1]$, for the ordering $[A, B, C, D]$.

[9] See McGinty (2007) for more on this point.

**Table 2** Internal and external stability

| Coalition $k$ | $\pi_s^{\theta}(k)$ | $\pi_s^{SV}(k)$ | $\pi_s^{NBS}(k)$ | $\pi_s^{E}(k)$ |
|---|---|---|---|---|
| $[A], [B], [C], [D]$ | [30.7,46.0,61.4,76.7]* | [30.7,46.0,61.4,76.7]* | [30.7,46.0,61.4,76.7]* | **[30.7,46.0,61.4,76.7]***** |
| $[A, B]$ | [31.6, 46.1]* | **[31.2, 46.5]***** | **[31.2, 46.5]***** | [38.9, 38.9]** |
| $[A, C]$ | [32.6, 61.7]* | [31.8, 62.5]* | [31.8, 62.5]* | [47.1, 47.1]** |
| $[A, D]$ | [33.8, 77.4]* | [32.6, 78.7]* | [32.6, 78.7]* | [55.6, 55.6] |
| $[B, C]$ | [47.0, 61.5]* | [46.6, 61.9]* | [46.6, 61.9]* | [54.2, 54.2] |
| $[B, D]$ | [47.8, 76.9]* | [47.0, 77.7]* | [47.0, 77.7]* | [62.4, 62.4] |
| $[C, D]$ | [62.4, 76.9]* | [62.0, 77.3]* | [62.0, 77.3]* | [69.7, 69.7] |
| $[A, B, C]$ | **[33.3, 50.8, 66.5]***** | [35.1, 49.8, 65.8]** | [34.9, 50.2, 65.6]** | [50.2, 50.2, 50.2]** |
| $[A, B, D]$ | **[33.9, 52.1, 83.4]***** | [41.3, 56.1, 72.0]** | [36.0, 51.3, 82.0]** | [56.5, 56.5, 56.5]** |
| $[A, C, D]$ | **[33.3, 68.9, 84.5]***** | [37.3, 66.6, 82.8]** | [36.7, 67.4, 82.7]** | [62.2, 62.2, 62.2] |
| $[B, C, D]$ | **[49.7, 66.9, 82.7]***** | [51.2, 66.2, 82.0]** | [51.1, 66.5, 81.8]** | [66.5, 66.5, 66.5] |
| $[A, B, C, D]$ | [34,54.5,74.2,93.0]** | [41.2,55.1,71.2 88.1]** | [40.9 56.2,71.6,86.9]** | [63.9,63.9 63.9,63.9]** |

* Denotes internally stable
** Denotes externally stable
*** Denotes both internally and externally stable (in bold)

inside or outside of a given coalition are small, however this is the nature of the underlying theoretical model. The parameters were chosen with the expectation that on average subjects would earn approximately \$25 if actual behavior matched the theoretical predictions. Subjects earned a \$5 fee for showing up and earnings each period depended on which coalition formed. The theoretical prediction for the four runs (ten periods each) resulted in an expected value of \$20 for each subject.

As a performance metric we define efficiency ($E$) as the percentage of the potential aggregate payoff difference between the singleton and grand coalitions obtained by $k$ forming: $E(k) = \frac{\sum_{i\in N}\pi_i(k)-\sum_{i\in N}\pi_i(\{i\})}{\sum_{i\in N}\pi_i(N)-\sum_{i\in N}\pi_i(\{i\})}$. The potential gains vary considerably across treatments. For the equal split treatment 0% efficiency gain is predicted since there are no stable coalitions larger than the singleton. Efficiency slightly improves under the Shapley value and NBS. Given that only coalition $[A, B]$, with the two lowest share agents, is both internally and externally stable we expect a gain of 28.5% under these two treatments. Predicted performance from stable coalitions is highest with the $\theta$ rule. The four three-member coalitions achieve between 45 and 83% of the efficiency gains. No clear equilibrium selection criteria allows us to make a prediction within the range, even if subjects form only stable coalitions. The inherent free-riding incentives generate a four-way "Battle of the Sexes" type structure where each subject would prefer to be outside a three member coalition.

The analysis below examines two related sets of hypotheses. The first concerns the prediction that agents will form coalitions which are stable under each of the four burden sharing rules. The second considers the relative efficiency of coalition formation under each rule. We are also interested in the actual efficiency performance of the $\theta$ rule given the wide range of predictions within the set of stable coalitions.

## 4 Results

Figure 1 presents the frequency distribution of the data in Table 3 across each of the twelve possible coalitions by treatment. For each treatment, those coalitions which are stable both

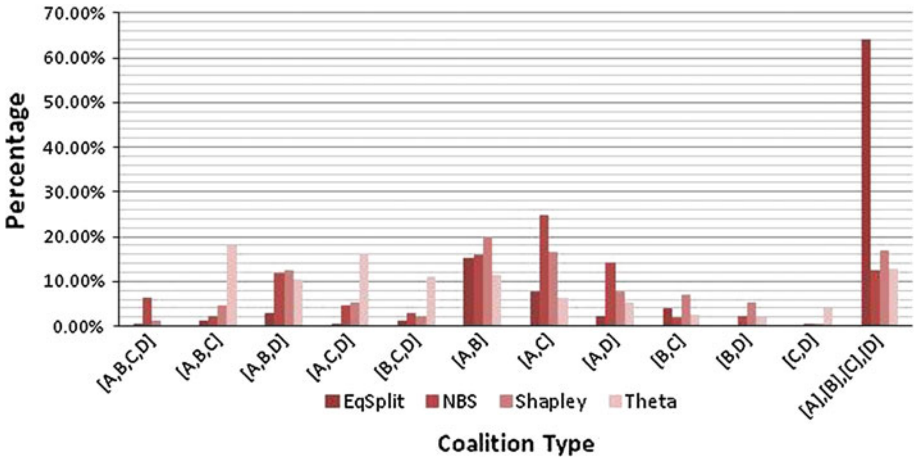


**Fig. 1** Coalition frequency

**Table 3** Coalition frequency (%)

| | $\theta$ | Shapley | NBS | Equal |
|---|---|---|---|---|
| $[A]$,$[B]$,$[C]$,$[D]$ | 12.90 | 17.06 | 12.35 | 64.12 |
| $[A, B]$ | 11.61 | 20.00 | 15.88 | 15.29 |
| $[A, C]$ | 6.45 | 16.47 | 24.71 | 7.65 |
| $[A, D]$ | 5.16 | 7.65 | 14.12 | 2.35 |
| $[B, C]$ | 2.58 | 7.06 | 1.76 | 4.12 |
| $[B, D]$ | 1.94 | 5.29 | 2.35 | 0 |
| $[C, D]$ | 3.87 | 0.59 | 0.59 | 0 |
| $[A, B, C]$ | 18.06 | 4.71 | 2.35 | 1.18 |
| $[A, B, D]$ | 10.32 | 12.35 | 11.76 | 2.94 |
| $[A, C, D]$ | 16.13 | 5.29 | 4.71 | 0.59 |
| $[B, C, D]$ | 10.97 | 2.35 | 2.94 | 1.18 |
| $[A, B, C, D]$ | 0 | 1.18 | 6.47 | 0.59 |

internally and externally are the most common outcomes. The observed departures from stability varies across treatments, the most notable being a non-trivial fraction of singletons. In addition, while the grand coalition rarely forms, coalitions are clearly largest under the $\theta$ rule.

## 4.1 Descriptive Statistics

The theoretical predictions under Shapley value and NBS are identical. Our experimental results vary only slightly across these two treatments. Coalition formation is most evenly distributed here even though the only internally and externally stable coalition is $[A, B]$. This is the modal coalition in the Shapley value treatment, and second most frequent to $[A, C]$ for NBS. The distribution of coalitions is clearly dominated by the various two member coalitions, occurring 59 and 56% of the time for Shapley and NBS respectively. These represent the internally stable set.

The singleton outcome represents 17% of Shapley and 12% of NBS results. This raises an important issue relating the stability of a final coalition to the dynamic coalition formation

**Table 4** Coalition membership frequency (%)

| | $\theta$ | Shapley | NBS | Equal |
|---|---|---|---|---|
| Singleton | 12.90 | 17.06 | 12.35 | 64.12 |
| Two member coalition | 31.61 | 57.06 | 59.41 | 29.41 |
| Three member coalition | 55.48 | 24.71 | 21.76 | 5.88 |
| Grand coalition | 0 | 1.18 | 6.47 | 0.59 |
| $A$ join = 1 | 72.26 | 75.29 | 87.06 | 72.35 |
| $B$ join = 1 | 58.06 | 55.29 | 45.88 | 34.12 |
| $C$ join = 1 | 61.94 | 42.94 | 45.29 | 15.29 |
| $D$ join = 1 | 50.32 | 35.29 | 43.53 | 7.65 |

process. To illustrate, consider the decision process facing agents, independent of their type. The first two agents to join the coalition increase their payoff relative to the singleton payoff under all treatments except the equal split. This stems from the internal stability of two member coalitions. When these two member coalitions are not externally stable an outside player receives a larger payoff from joining the coalition. The resulting three member coalitions are not internally stable under Shapley and NBS, thus more than one member may simultaneously leave. If two or more players leave a three-member coalition and time runs out, then the final outcome is a singleton.

The composition of coalitions we observe clearly relates to the relative strength of incentives facing different subject types. Subject $A$ has a dominant strategy of always joining a coalition in treatments other than $\theta$, and observed participation shown in Table 4 was high: 87.06% in NBS and 75.29% in Shapley. The stronger players (in terms of threat points) have weaker incentives which are reflected in their participation rates. For example, subject $D$ was a member of a coalition 35.29% of the time for Shapley and 43.53% time for NBS.

Under the $\theta$ rule, coalition membership in Figure 1 is clearly the highest. The three member coalitions formed 55% of the time. Thus, a majority of the outcomes fit the theoretic stability prediction. Further, the modal coalition under this treatment $[A, B, C]$ generates the highest possible efficiency (83%) short of the grand coalition. These increases in coalition membership come in spite of the fact that subjects face the same incentives to delay their decision to join stable coalitions as under Shapley and NBS.

The $\theta$ rule allocates the coalition's worth as a function of individual $\theta$ shares and the free-rider payoff. The realigned incentives are reflected in increased willingness to join by stronger members. Table 4 shows that type $A$ subjects chose to be members of a coalition around 72% of the time, a reduction from Shapley and NBS. However increases are seen in the percentages of $B$, $C$, and $D$ types decisions as they rose to 58.06, 61.94, and 50.32% of the time, respectively.

From a social or policy perspective our primary interest concerns not the composition but the efficiency of coalitions under our burden sharing rules. Recall that the expected efficiency is 28.5% under Shapley and NBS. Observed efficiency was 36 and 40%, respectively. The efficiency observed under $\theta$ was 51%, within the range predicted by the theoretical model. This represents a substantial improvement over the other rules in spite of the fact that they outperform their predicted efficiency. The equal split treatment obtained only 13% of the potential gain. The $\theta$ rule generates four stable three agent coalitions; with each agent type having a unique preferred coalition. We have no prior beliefs regarding which of these might dominate but the modal coalition of $[A, B, C]$ represents the highest efficiency in the set of internally and externally stable coalitions.

The results from the equal split treatment, in which the singleton is predicted to be the only stable coalition, match the predictions. No coalition formed 64.12% of the time under equal split. We take this as an indication that individuals respond primarily to their own payoff incentives and fairness considerations do not appear to be a primary factor of concern. While the results above present a clear correspondence between the data and our theoretical expectations, we now turn to a more formal analysis.

## 4.2 Regression Results

To isolate the factors influencing coalition formation we estimate two regression models. The first investigates if the $\theta$ rule generates higher coalition membership than the alternative rules. The second tests if the treatments alter the decision by subjects to join a given coalition.

In order to test if the $\theta$ rule generates larger coalitions we fit an ordered logit model. The dependent variable takes one of the four ordered categories reflecting the number of coalition members. Equation (11) describes the probability of the response variable equaling a given coalition size conditional on the values of the explanatory variables, assuming a logit distribution:

$$P(y = j|x) = \exp(x\beta_j)/\left[1 + \sum_{h=2}^{4}\exp(x\beta_h)\right], \qquad j = 2, 3, 4 \tag{11}$$

where $j$'s are the three possible outcomes when excluding the zero members outcome, $x$ is a vector of dummy variables, and $\beta_j$ is a $k \times 1$ vector of unknown parameters. The vector $x$ consists of three treatment dummies, three treatment order dummies, and nine period dummies. The excluded comparison groups are the $\theta$ treatment, the fourth treatment order, and the final period. Thus, we interpret the value of the coefficients on the treatment dummies (Shapley, NBS and equal split) as the effects on coalition size of these treatments relative to $\theta$. Our theory predicts that these will be negative. The three treatment order dummies are compared to the fourth treatment order. Our theory predicts that none of these dummies will be significantly different from zero. The nine time period dummies are included to identify significant differences in coalition size for earlier interactions relative to the final period. We predict these will not be significantly different from zero, however VCM experiments typically find these to be negative, indicative of time decay. Finally, we included a geographical dummy variable taking the value one for observations from Ryerson University to capture any difference from location. We predict that the coefficient on the Ryerson dummy will not be significantly different from zero. Our estimated equation is:

$$g(p_{ij}) = u_j + x_i b, \qquad j = 2, 3, 4 \tag{12}$$

where $g$ is a logit cumulative probability function, $u_j$ is the intercept that depends on category $j$, and $i$ represents a certain group. Results are presented in Table 5.

The coefficients on all three treatment dummies are negative and significant at the 5% level. The magnitude of the results can be explained by calculating the odds ratios, exponentiating the estimated coefficients. These coefficients show the odds of a given treatment, relative to the odds of the $\theta$ treatment, in forming larger coalitions. For example, the equal split treatment has an exponentiated coefficient of 0.0483, meaning there is less than a 5% chance that there will be a higher number of coalition members in equal split than $\theta$. All estimated odds ratios are less than one, indicating a smaller chance of the other three treatments generating a larger coalition size than $\theta$.

**Table 5** Estimates for coalition members as dependent variable

| Parameter | Estimate | Standard error | Odds ratios |
|---|---|---|---|
| Intercept 4 | −2.89 * ** | 0.42 | |
| Intercept 3 | 0.24 | 0.35 | |
| Intercept 2 | 2.73 * ** | 0.37 | |
| Shapley | −0.91 * ** | 0.24 | 0.40 |
| NBS | −0.61 * * | 0.24 | 0.55 |
| Equal split | −3.03 * ** | 0.26 | 0.05 |
| First treatment | 0.48 * * | 0.24 | 1.62 |
| Second treatment | 0.29 | 0.24 | 1.33 |
| Third treatment | 0.10 | 0.23 | 1.11 |
| Period 1 | −0.09 | 0.34 | 0.91 |
| Period 2 | −0.10 | 0.34 | 0.91 |
| Period 3 | −0.26 | 0.34 | 0.77 |
| Period 4 | −0.31 | 0.34 | 0.73 |
| Period 5 | −0.07 | 0.34 | 0.93 |
| Period 6 | −0.37 | 0.34 | 0.69 |
| Period 7 | −0.11 | 0.33 | 0.90 |
| Period 8 | −0.12 | 0.33 | 0.90 |
| Period 9 | −0.40 | 0.34 | 0.67 |
| Ryerson | −0.70 * ** | 0.15 | 0.50 |

*, **, *** indicate significant at the 10%, 5%, and 1% levels, respectively

In our second estimation we are interested in the individual decisions of subjects to join the coalition. We estimate a simple binary logistic model, where join or not join was the dependent variable. By letting $j = 1$ in Eq. (12) we obtain the probabilities for our response variable. For $j = 1$ Eq. (12) becomes:

$$g(p_{i1}) = u + x_i b \tag{13}$$

Note that in this equation the probability modeled is $join = 1$. To the vector $x$ of treatment, treatment order and time period dummies we add the individual characteristics of the profit differential and subject type. Profit differential is defined as the incentive to free-ride: $\pi_t(k \diagdown \{t\}) - \pi_s^i(k), \forall s = t \in k, i = \theta, SV, NBS, Equal$. We expect a lower (higher) probability of a subject choosing to join a coalition the greater (smaller) the differential. There are three subject type dummies with type $A$ as the excluded comparison group. Results are given in Table 6.

The estimated coefficients for the subject type dummies are all negative, meaning that relative to subject type $A$, subjects $B$, $C$, and $D$ have lower incentives to join a coalition. These estimates are all significant, and their odds ratios are all less than one. They monotonically decrease from subject $B$ to $D$ meaning that more powerful players have a lower probability of joining coalitions. The coefficient on the profit differential is also negative, as expected. The odds ratio shows that for every dollar of positive differential, the odds of not joining are 0.9 times the odds of joining. The treatment dummies estimates are all negative and significant. The odds ratios are all less than one, meaning that relative to the $\theta$ treatment the other treatments offer less incentive to join coalitions. The interpretation of the odds ratios is the marginal effects on the probability of a certain category occurring relative to the excluded category.

**Table 6** Estimates for join (not join) as dependent variable

| Parameter | Estimate | Standard error | Odds ratios |
|---|---|---|---|
| Intercept | 1.43 *** | 0.21 | |
| Type B | −1.07 *** | 0.13 | 0.34 |
| Type C | −1.20 *** | 0.13 | 0.30 |
| Type D | −1.42 *** | 0.14 | 0.24 |
| Profit differential | −0.08 *** | 0.01 | 0.92 |
| Shapley | −0.31 ** | 0.13 | 0.73 |
| NBS | −0.12 | 0.13 | 0.89 |
| Equal split | −0.85 *** | 0.15 | 0.43 |
| First treatment | 0.23* | 0.13 | 1.26 |
| Second treatment | 0.25* | 0.14 | 1.29 |
| Third treatment | 0.12 | 0.13 | 1.13 |
| Period 1 | −0.06 | 0.19 | 0.94 |
| Period 2 | 0.03 | 0.19 | 1.03 |
| Period 3 | −0.10 | 0.19 | 0.91 |
| Period 4 | −0.18 | 0.19 | 0.84 |
| Period 5 | 0.09 | 0.19 | 1.10 |
| Period 6 | −0.07 | 0.19 | 0.93 |
| Period 7 | −0.04 | 0.19 | 0.97 |
| Period 8 | −0.01 | 0.19 | 0.99 |
| Period 9 | −0.12 | 0.19 | 0.89 |
| Ryerson | −0.32 *** | 0.09 | 0.72 |

*, **, *** indicate significant at the 10%, 5%, and 1% levels, respectively

There are several conclusions that are consistent across both regressions. First, period dummies are not significant, indicating there is no pattern of time decay typical of VCM public goods experiments (Ledyard 1995). Second, only the first treatment order is significant, indicating more cooperation relative to the fourth treatment. The binary decision to join or not and certainty over payoffs are potential explanations for the difference from previous VCM results.

## 5 Conclusion

This experiment shows that traditional cooperative allocation rules result in lower efficiency and coalition size than an optimal rule which explicitly addresses subjects' payoffs outside the coalition. The Shapley value and the Nash bargaining solution are used in the IEA literature as a zero-sum system of transfers, attempting to expand coalition membership by self-interested nations. Uniform rules fare even worse. When agents are inherently asymmetric they may not view "fair" and equal as the same thing.

Despite the strength of our results some aspects of the experimental setting might be improved. Given the sequence of actions, subjects did not have immediate information about other subjects leaving or joining a coalition. Thus, some may have been caught inside a three member coalition when they would have left if they had the choice. Our random period length attempted to mitigate this issue, but was imperfect. Since being outside a three member coalition is almost always the most preferred coalition structure, subjects may have tried to form the grand coalition and then leave. With Shapley and NBS all two member coalitions other than $[A, B]$ are internally, but not externally stable. This means that we would expect three member coalitions to form and then break down since the resulting three member coalitions are not internally stable. This "cycling" is one explanation why more three member coalitions formed for Shapley and NBS than we expected. There is no cycling effect for $\theta$ and equal split. Another possible design is to remove the time limit on periods, making subjects unanimously agree to end the period. Within period coalitions would provide additional information, but at the expense of fewer periods, particularly for the Shapley and NBS treatments where cycling should be more prevalent. Another possible design limitation is that some of the payoff differences between being in or out of a given coalition were very small. This resulted because we chose to test the standard IEA model and payoffs were scaled for appropriate subject earnings. Finally, we chose the Shapley value because of its use in the IEA literature. Croson et al. (2004) among others have noted that the Myerson-Shapley value (Myerson 1977) may be more appropriate for games with externalities, and when more than one coalition may form.

The practical implications of this experiment are clear. Efficiency in public goods games can be significantly improved by using a rule which directly addresses the free-rider payoff rather than an equal split or traditional cooperative game theory rules. The payoffs were generated using the benchmark model of IEAs, however the results apply to any pure public good. Compared to the VCM environment we have changed the subject's choice variable from a contribution, with an unknown contribution by other group members, to a binary decision to be in or out of a coalition with public information on coalition members. Our results refute two VCM stylized facts. First, we observe very little over-contribution relative to the theoretical prediction. Under $\theta$, over-contribution (relative to the individual optimum) would imply the grand coalition. The grand coalition did not occur in any of the 170 observations. Second, we observe none of the time-decay present in VCM experiments.

One direction for future research would be to test the different optimal allocations. The optimal rules in Carraro et al. (2006), Weikard (2009) and McGinty (2007, 2011) propose different shares of the coalition surplus than the one tested in this paper. Each rule will result in the same set of stable coalitions, yet they differ in the payoff advantage to remaining a coalition member. McGinty (2011) shows theoretically that an equal share of the surplus maximizes the stability of a stable coalition, however no experimental comparison of these rules has determined which one performs better in the lab.

**Acknowledgments** This paper has benefited from comments by Rachel Croson, Dan Friedman, John Heywood and seminar participants at the ESA meetings. The authors wish to thank the UWM Center for Research on International Economics, the UWM Foundation and the Ryerson University Faculty of Arts for generous financial support. The usual caveat applies.

## Appendix

### Instructions

This experiment will consist of four different 'runs' each lasting 10 periods. In each run you will be randomly assigned a type [$A$, $B$, $C$, or $D$] and to a group of four people (you and three others). You will have the same type and the same group members for all 10 periods. At the end of each run you will randomly be assigned a new type and a new group. There is a 1/4 chance that you will be assigned a certain type each run. There are three four-person groups in this room.

In each period group members will decide to join or not join a subgroup. Your earnings for each period depend on your type ($A$, $B$, $C$, or $D$), if you are in or out of the subgroup, and the other members of the subgroup. The Earnings Table shows the earnings for each group member for all possible subgroups. You will have the information about your type and who is in the subgroup at all times on your computer screen.

In each period you will asked whether or not you wish to join the subgroup. After you and the other group members decide, you will see your earnings. You will also see who the other subgroup members are. At this point, you can click on the 'Confirm' button to confirm your decision or the 'Reverse' button to change your decision.

If anyone wishes to change their decision before confirming, everyone will be taken back to the initial decision screen. Once the person who wants to change makes their new decision, everyone returns to the confirmation screen. Everyone's earnings will be updated with the new information and you will again have the opportunity to confirm before the period ends. The period ends after all members of the group confirm their choices or when time runs out. Each period will last at least 60 s, after which it could end at any time.

At the end of the period, you will see your earnings and the subgroup members for that period and a history of subgroups and earnings for all prior periods from that run. Do not discuss your decisions with anyone in the room. Only the monitors will learn of your decisions. The other players do not know your type, only their own, and only the monitors know the identities of the group members.

Example: During run I, you are type B. The subgroup consists of $[A, D]$. Your earnings for the period are 51.4 if you stay out of the subgroup. If you decide to join, the subgroup will become $[A, B, D]$ and if that is the final subgroup your earnings will become 52.1.

Example: During run II, you are type C. The subgroup consists of $[A, C, D]$. Your earnings for the period are 66.6 if you remain in the subgroup. If you decide to leave the subgroup, it will become $[A, D]$ and your earnings will become 68.5. However, say the type A subgroup member also decides to leave. Then there is no subgroup (only D is left, so in the table it is $[A]$, $[B]$, $[C]$, $[D]$) and you will earn 61.4.

Example: During run IV, you are type D. The subgroup consists of $[A, B, C,$ and $D]$. Your earnings for the period are 63.9 if this is the final subgroup. However, if both the type A and type B members leave the subgroup, your earnings will become 69.7.

See Table 7.

**Table 7** Earnings table

| Subgroup | Earnings for subgroup members | Earnings for non-subgroup members |
|---|---|---|
| *Run I* | | |
| $[A]$, $[B]$, $[C]$, $[D]$ | NA | $[A = 30.7]$, $[B = 46.0]$, $[C = 61.4]$, $[D = 76.7]$ |
| $[A, B]$ | $[A = 31.6, B = 46.1]$ | $[C = 66.1]$, $[D = 82.6]$ |
| $[A, C]$ | $[A = 32.6, C = 61.7]$ | $[B = 50.4]$, $[D = 84.1]$ |
| $[A, D]$ | $[A = 33.8, D = 77.4]$ | $[B = 51.4]$, $[C = 68.5]$ |
| $[B, C]$ | $[B = 47.0, C = 61.5]$ | $[A = 32.9]$, $[D = 82.3]$ |
| $[B, D]$ | $[B = 47.8, D = 76.9]$ | $[A = 33.3]$, $[C = 66.5]$ |
| $[C, D]$ | $[C = 62.4, D = 76.9]$ | $[A = 32.8]$, $[B = 49.3]$ |
| $[A, B, C]$ | $[A = 33.3, B = 50.8, C = 66.5]$ | $[D = 98.0]$ |
| $[A, B, D]$ | $[A = 33.9, B = 52.1, D = 83.4]$ | $[C = 79.1]$ |
| $[A, C, D]$ | $[A = 33.3, C = 68.9, D = 84.5]$ | $[B = 59.5]$ |
| $[B, C, D]$ | $[B = 49.7, C = 66.9, D = 82.7]$ | $[A = 38.9]$ |
| $[A, B, C, D]$ | $[A = 34.0, B = 54.5, C = 74.2, D = 93.0]$ | NA |

**Table 7** continued

| Subgroup | Earnings for subgroup members | Earnings for non-subgroup members |
|---|---|---|
| *Run II* | | |
| $[A], [B], [C], [D]$ | NA | [A = 30.7], [B = 46.0], [C = 61.4], [D=76.7] |
| $[A, B]$ | $[A = 31.2, B = 46.5]$ | $[C = 66.1], [D = 82.6]$ |
| $[A, C]$ | $[A = 31.8, C = 62.5]$ | $[B = 50.4], [D = 84.1]$ |
| $[A, D]$ | $[A = 32.6, D = 78.7]$ | $[B = 51.4], [C = 68.5]$ |
| $[B, C]$ | $[B = 46.6, C = 61.9]$ | $[A = 32.9], [D = 82.3]$ |
| $[B, D]$ | $[B = 47.0, D = 77.7]$ | $[A = 33, 3], [C = 66.5]$ |
| $[C, D]$ | $[C = 62.0, D = 77.3]$ | $[A = 32.8], [B = 49.3]$ |
| $[A, B, C]$ | $[A = 35.1, B = 49.8, C = 65.8]$ | $[D = 98.0]$ |
| $[A, B, D]$ | $[A = 41.3, B = 56.1, D = 72.0]$ | $[C = 79.1]$ |
| $[A, C, D]$ | $[A = 37.3, C = 66.6, D = 82.8]$ | $[B = 59.5]$ |
| $[B, C, D]$ | $[B = 51.2, C = 66.2, D = 82.0]$ | $[A = 38.9]$ |
| $[A, B, C, D]$ | $[A = 41.2, B = 55.1, C = 71.2, D = 88.1]$ | NA |
| *Run III* | | |
| $[A], [B], [C], [D]$ | NA | $[A = 30.7], [B = 46.0], [C = 61.4], [D = 76.7]$ |
| $[A, B]$ | $[A = 31.2, B = 46.5]$ | $[C = 66.1], [D = 82.6]$ |
| $[A, C]$ | $[A = 31.8, C = 62.5]$ | $[B = 50.4], [D = 84.1]$ |
| $[A, D]$ | $[A = 32.6, D = 78.7]$ | $[B = 51.4], [C = 68.5]$ |
| $[B, C]$ | $[B = 46.6, C = 61.9]$ | $[A = 32.9], [D = 82.3]$ |
| $[B, D]$ | $[B = 47.0, D = 77.7]$ | $[A = 33.3], [C = 66.5]$ |
| $[C, D]$ | $[C = 62.0, D = 77.3]$ | $[A = 32.8], [B = 49.3]$ |
| $[A, B, C]$ | $[A = 34.9, B = 50.2, C = 65.6]$ | $[D = 98.0]$ |
| $[A, B, D]$ | $[A = 36.0, B = 51.3, D = 82.0]$ | $[C = 79.1]$ |
| $[A, C, D]$ | $[A = 36.7, C = 67.4, D = 82.7]$ | $[B = 59.5]$ |
| $[B, C, D]$ | $[B = 51.1, C = 66.5, D = 81.8]$ | $[A = 38.9]$ |
| $[A, B, C, D]$ | $[A = 40.9, B = 56.2, C = 71.6, D = 86.9]$ | NA |
| *Run IV* | | |
| $[A], [B], [C], [D]$ | NA | $[A = 30.7], [B = 46.0], [C = 61.4]. [D = 76.7]$ |
| $[A, B]$ | $[A = 38.9, B = 38.9]$ | $[C = 66.1], [D = 82.6]$ |
| $[A, C]$ | $[A = 47.1, C = 47.1]$ | $[B = 50.4], [D = 84.1]$ |
| $[A, D]$ | $[A = 55.6, D = 55.6]$ | $[B = 51.4], [C = 68.5]$ |
| $[B, C]$ | $[B = 54.2, C = 54.2]$ | $[A = 32.9], [D = 82.3]$ |
| $[B, D]$ | $[B = 62.4, D = 62.4]$ | $[A = 33.3], [C = 66.5]$ |
| $[C, D]$ | $[C = 69.7, D = 69.7]$ | $[A = 32.8], [B = 49.3]$ |
| $[A, B, C]$ | $[A = 50.2, B = 50.2, C = 50.2]$ | $[D = 98.0]$ |
| $[A, B, D]$ | $[A = 56.5, B = 56.5, D = 56, 5]$ | $[C = 79.1]$ |
| $[A, C, D]$ | $[A = 62.2, C = 62.2, D = 62.2]$ | $[B = 59.5]$ |
| $[B, C, D]$ | $[B = 66.5, C = 66.5, D = 66.5]$ | $[A = 38.9]$ |
| $[A, B, C, D]$ | $[A = 63.9, B = 63.9, C = 63.9, D = 63.9]$ | NA |

Runs I–IV are the theta proportion rule, Shapley value, Nash bargaining solution, and equal split, respectively. Treatment order was randomly determined

## References

Barrett S (1994) Self-enforcing international environmental agreements. Oxf Econ Pap 46:878–894
Barrett S (1997) Heterogeneous international environmental agreements. In: Carraro CInternational environmental negotiations. Strategic policy issues, Edward Elgar
Bolton G (1997) The rationality of splitting equally. J Econ Beha Org 32:365–381
Bolton G (1998) Bargaining and dilemma games: from laboratory data towards theoretical synthesis. Exp Econ 1:257–281
Bolton G, Chatterjee K, McGinn K (2003) How communication links influence coalition bargaining: a laboratory investigation. Manag Sci 49(5):583–598
Botteon M, Carraro C (2001) Environmental coalitions with heterogenous countries: burden sharing and carbon leakage. In: Ulph A (ed) Environmental policy, international agreements, and international trade. Oxford University Press, New York
Burlando R, Guala F (2005) Heterogeneous agents in public goods experiments. Exp Econ 8:35–54
Carraro C, Eyckmans J, Finus M (2006) Optimal transfers and participation decisions in international environmental agreements. Rev Int Org 1(4):379–396
Carraro C, Siniscalco D (1993) Strategies for the international protection of the environment. J Public Econ 52:309–328
Chan K, Mestelman S, Moir R, Muller RA (1999) Heterogeneity and the voluntary provision of public goods. Expl Econ 2:5–30
Croson R, Gomes A, McGinn K, Nöth M (2004) Mergers and acquisitions: an experimental analysis of synergies. externalities and dynamics. Rev Financ 8:481–514
d'Aspremont C, Jacquemin A, Gabszewicz JJ, Weymark J (1983) On the stability of collusive price leadership. Can J Econ 16:17–25
Dannenberg A, Lange A, Sturm B (2010) On the formation of coalitions to provide public goods-experimental evidence from the lab. NBER working paper #15967
Finus M (2003) Stability and design of international environmental agreements: the case of transboundary pollution. In: Folmer H, Tietenberg T (eds) International yearbook of environmental and resource economics, 2003/2004. Edward Elgar, Cheltenham pp 82–158
Fischbacher U (2007) z-Tree: zurich toolbox for ready-made economic experiments. Exp Econ 10:171–178
Friedman D, Sunder S (1994) Experimental methods: a primer for economists. Cambridge University Press, Cambridge
Friedman J (1971) A non-cooperative equilibrium for supergames. Rev Econ Stud 38(113):1–12
Isaac RM, McCue K, Plott C (1985) Public goods provision in an experimental environment. J Public Econ 26:51–74
Keser C (1996) Voluntary contributions to a public good when partial contribution is a dominant strategy. Econ Lett 50:359–366
Laury S, Holt C (2005) Voluntary provision of public goods: Experimental results with interior Nash equilibria. In: Plott C, Smith V (eds) Handbook of experimental economics results, vol. 1
Ledyard J (1995) Public goods: a survey of experimental research. In: Kagel J, Roth A (eds) Handbook of experimental economics. Princeton University Press, Princeton pp 111–194
Leopold-Wildburger U (1992) Payoff divisions on coalition formation in a three-person characteristic function experiment. J Econ Behav Org 17:183–193
McGinty M (2007) International environmental agreements among asymmetric nations. Oxf Econ Pap 59(1):45–62
McGinty M (2011) A risk-dominant allocation: maximizing coalition stability. J Public Econ Theory 13(2):311–325
Myerson R (1977) Values of games in partition function form. Int J Game Theory 6(1):23–31
Roth A (1987) Bargaining phenomena and bargaining theory. In: Laboratory experimentation in economics: six points of view, Cambridge, UK, pp. 14–41
Weikard H-P (2009) Cartel stability under an optimal sharing rule. Manch Sch 77(5):575–593
Weikard H-P, Dellink R (2010) Sticks and carrots for the design of international climate agreements with renegotiations. Annal Oper Res (forthcoming)
Weikard H-P, Dellink R, Ecco van I (2010) Renegotiations in the greenhouse. Environ Resour Econ 45:573–596