

Archived version from NCDOCKS Institutional Repository <http://libres.uncg.edu/ir/asu/>

Cherry, Todd and David McEvoy. (2013). Enforcing Compliance with Environmental Agreements in the Absence of Strong Institutions: An Experimental Analysis. *Environmental and Resource Economics* 54(1): 63-77. Published by Springer [www.springer.com](http://www.springer.com) (ISSN: 0924-6460) DOI 10.1007/s10640-012-9581-3

## **Enforcing Compliance with Environmental Agreements in the Absence of Strong Institutions: An Experimental Analysis**

**Todd L. Cherry & David M. McEvoy**

### **Abstract**

This paper uses laboratory experiments to evaluate the performance of a deposit-refund mechanism used to enforce compliance with voluntary public-good commitments made in the absence of strong regulatory institutions. With this mechanism agents decide whether to join an agreement and pay a deposit prior to making their contribution decisions. If an agreement receives sufficient membership to form, members then make their contribution decisions and compliant members are refunded their deposits. If an agreement does not form, then deposits are immediately refunded and a standard voluntary contribution game is played. We find that the deposit-refund mechanism achieves nearly full efficiency when agreements require full participation, but is far less effective, and in some cases disruptive, when agreements require only partial participation. As the mechanism does not require the existence of strong sanctioning institutions, it is particularly suited for enforcing compliance with international environmental agreements.

### **Keywords**

Enforcing compliance, Environmental agreements, Experimental economics, Deposit refund, International environmental agreements

## 1 Introduction

The management of many environmental and natural resources gives rise to *social dilemma* situations in which individually rational decisions are not collectively rational. Public goods (e.g., protection of ozone layer, preservation of biological diversity) tend to be underprovided and common-pool-resources (e.g., fisheries, forest resources) tend to be overexploited. Many social dilemma situations may be easily resolved given the existence of strong regulatory institutions. For example, a well-functioning government could provide efficient levels of public goods if it can adequately collect taxes and use the revenue to produce public goods (e.g., public transportation). Likewise, common-pool resources may be managed efficiently given an effective regulatory institution and credible enforcement provisions (e.g., federal regulation of commercial fisheries). In many cases, however, these types of social dilemmas exist in the absence of strong regulatory institutions. The problem of managing greenhouse-gas emissions, or transboundary environmental problems in general (e.g., managing CFC emissions, trade in endangered species, biodiversity preservation), falls into this category. Sovereignty requires that transboundary environmental public goods are provided in a decentralized fashion without an overarching global government that can heavily regulate the management of shared resources.

In the absence of traditional regulatory approaches, transboundary resources are typically managed through voluntary institutions called international environmental agreements (IEAs). While there are many obstacles in enacting effective IEAs, there are two potentially formidable challenges. First, participation with IEAs is voluntary, meaning nations are free to decide whether to join an agreement. This presents a challenge because commitments made under an agreement, by definition, restrict a member's use of a shared environmental or natural resource, and there are often strong incentives to free ride. The second challenge is that compliance with the terms of an agreement needs to be enforced. Provisions for enforcement must be included as part of the agreement and cannot depend on the presence of a strong regulatory body. Furthermore, the enforcement provisions need to be effective in the sense that they will be severe enough to deter noncompliance and credible in the sense that they will actually be imposed. While a great deal of research has addressed the first challenge posed—the one of voluntary participation—less attention has been paid to the fact that compliance with voluntary agreements needs to be enforced. In this paper we address the issue of enforcing compliance with IEAs by examining one promising instrument that exhibits the necessary features for effective enforcement. The instrument explored here—a deposit-refund mechanism—has recently been proposed in the theoretical literature (Gerber and Wichardt 2009) and the reported results are encouraging. Building on their study, we empirically test the theoretical predictions of alternative deposit-refund schemes using experimental methods.

While some research does address the compliance issue directly (e.g., Barrett 1994, 1997, 2008; Finus and Rundshagen 1998; Barrett and Stavins 2003; Hovi et al. 2007; Finus 2008; McEvoy and Stranlund 2009; McEvoy et al. 2011) much of the economics literature avoids the problem of maintaining compliance by designing models, and

experiments, in which agreement members are required to comply with their commitments (e.g., Hoel 1992; Carraro and Siniscalco 1993, 1998; Hoel and Schneider 1997; Rubio and Ulph 2006, 2007; Kolstad 2007; McGinty 2007; Carraro et al. 2009; Kosfeld et al. 2009; Breton et al. 2010; Dannenberg et al. 2010; McEvoy 2010).<sup>2,3</sup> In other words, these studies do not give agreement members the opportunity to violate the treaty. This is not a trivial issue. Ignoring the possibility of noncompliance abstracts from a fundamental problem for the effectiveness of cooperative agreements; that is, the opportunity to violate commitments could have significant impacts on the management of shared resources. In fact, many attribute the impending failure of the Kyoto Protocol to the lack of credible enforcement of compliance. Barrett (2003, p. 360) argues, "...if the negotiators had reflected on the need for enforcement and on the difficulty of devising an effective enforcement mechanism earlier in the process, they may have negotiated a different kind of treaty—one that sustained more cooperation."

Motivated by the need to explore enforcement strategies within environmental agreements in the absence of strong institutions, we adapt the general model of Gerber and Wichardt (2009) and utilize a set of laboratory experiments to test the effectiveness of different deposit-refund schemes. While the theoretical approach is useful for predicting how the deposit-refund system can facilitate cooperation, an experimental approach allows for its empirical test in the absence of naturally occurring data.<sup>5</sup> While deposit-refund mechanisms are used in a number of simple settings (e.g., return of recyclable solidwaste, environmental performance bonds, etc.), we analyze a deposit-refund game within an agreement formation context. The mechanism is straightforward. A group of players individually decide whether or not to join an agreement, in which members to the agreement commit to contributing to the public good. Upon joining the agreement, each member is required to pay a deposit to a neutral, third-party institution. The agreement enters into force if enough players join (and pay deposits) to satisfy a minimum membership requirement.<sup>9</sup> If the membership requirement is not satisfied, then no deposits are paid and no agreement enters into force. If the membership requirement is satisfied, then the agreement enters into force and members then decide whether or not to comply with their commitments (i.e., whether to contribute to the public good). After these decisions are made, the third-party institution pays back the deposit if the member is found compliant. Otherwise, the noncompliant agreement member is not refunded their deposit. The deposit-refund scheme does not require strong institutions in the sense that the neutral third party cannot force members to contribute (e.g. through taxation) and cannot engage in ex post sanctioning (e.g., issuing fines in response to noncompliance). Rather, the neutral institution simply has to withhold the initial deposit.

In our setting the deposit-refund mechanism can be likened to performance bonds that are levied in order to ensure fulfillment of environmental obligations. An example is the Surface Mining Control and Reclamation Act (1977) which requires firms to purchase bonds before attaining the permits to begin a mining project. The bond amounts are refunded if the mined land is adequately reclaimed. Performance bonds are also used in the management of timber, oil and gas industries. While deposit-refunds/performance bonds, to our knowledge, have not been included as compliance provisions in

environmental agreements, the fact that similar mechanisms are used in other environmental policy dimensions suggests that such mechanisms may be effective.

Our experimental design consists of three treatments. The first is a simple voluntary contributions game which is used as our baseline. The baseline captures the unilateral management of shared environmental and natural resources within social dilemma situations. The second treatment is an agreement formation game with a deposit-refund mechanism that requires full participation. In this treatment all players need to join the agreement before it enters into force. If full participation is not achieved, deposits are returned and players revert to the same voluntary contributions game as in the baseline treatment. The third and final treatment is an agreement formation game with a deposit-refund mechanism that only requires partial participation before entry into force. This treatment is motivated by many international agreements that require less than full participation, most notably the Kyoto Protocol.

We find that when full participation is required under a deposit-refund mechanism, the agreement formation process is almost 100 percent efficient. Agreements enter into force in nearly every trial and compliance is full. This is a dramatic increase in efficiency compared to the baseline case (i.e., the unilateral management case). Specifically, we observe a 33 percentage point increase in efficiency under voluntary agreements with a deposit-refund scheme that require full participation compared to unilateral management. Interestingly, agreements with deposit refunds that require only partial participation do not improve upon the unilateral management situation. Agreements formed infrequently in this scenario, but when agreements did form, average contributions were higher compared to those under unilateral management. Surprisingly, when agreements failed to form, contributions were dramatically lower than under the baseline treatment (26 vs. 60% efficiency). This suggests that behavior in the first stage affects cooperative behavior in the second stage. Players contributed far less to the public good when an agreement failed compared to when there was no option for an agreement to form.

Our research is also related to the extensive literature on threshold public good experiments with money-back guarantees (Van de Kragt et al. 1983; Dawes et al. 1986; Isaac et al. 1989; Rapoport and Eshed-Levy 1989; Erev and Rapoport 1990; Bagnoli and McKee 1991; Suleiman and Rapoport 1992; Rapoport and Suleiman 1993; Marks and Croson 1998, 1999; Cadsby and Maynes 1999; McEvoy 2010; Rauchdobler et al. 2010). In these experiments, subjects make individual contribution decisions, and if the threshold level of contributions (or contributors) is satisfied, then the public good is provided. Otherwise, contributions are returned. The threshold, like the minimum participation requirement in our experiments, acts as a commitment device. Our experiment differs in an important way from these experiments by giving subjects the opportunity to violate their commitments in a second stage.

## 2 Experimental Design

Our *baseline* treatment is a familiar discrete choice public goods game. Players are in groups of five. Each subject is endowed with \$13 and she must decide whether or not to contribute \$8 from her endowment to a public account.<sup>13</sup> Every \$1 contributed to the public account returns \$0.50 to each person in the group (i.e., the marginal per-capita return is 0.5). Players make their contribution decisions simultaneously. Therefore, a player's payoff function is

$$(1) \quad \pi_i = 13 + 4(q_i + q_{-i}) - 8q_i,$$

where  $q_i$  equals one if player  $i$  contributes \$8 to the public account and zero if player  $i$  contributes zero, and  $q_{-i}$  is the sum of the binary contribution decisions by the other four players. In a non-cooperative Nash equilibrium, no contributions are made to the public account. This is confirmed by subtracting a player's payoff when contributing zero to the public good from their payoff when they contribute \$8,  $\pi_i(q_i = 0) - \pi_i(q_i = 1) = \$ - 4$ . However, achieving the social optimum requires that all five players in the group contribute \$8 to the public account. In this case all players earn \$25.

In contrast to the theoretical prediction of no contributions, years of experimental research informs us that some subjects *are* expected to contribute to public goods in social dilemma situations such as this. A common result in a repeated game setting is that contributions start in the range of 40–60 percent of the maximum and decrease as the experiment progresses (Ledyard 1995; Chaudhuri 2011).

In our second treatment, *full participation*, we give subjects the opportunity to form an agreement with the goal of increasing contributions to the public good. The agreement works as follows: In the first stage, subjects decide simultaneously whether to join an agreement. An agreement is said to *form* if and only if all five players join. When an agreement does form, its members enter a non-binding commitment to contribute \$8 to the public account. In addition, joining an agreement requires that players pay an upfront deposit of  $d$  as part of the membership process. If an agreement forms after the first-stage decisions are made, then in the second stage its members decide individually whether to comply with their commitment and contribute \$8 to the public account. If a member does contribute \$8 to the public account in the second stage (i.e., she complies) then she is refunded her deposit of  $d$ . Otherwise, if she does not contribute \$8 she is not refunded her deposit. Finally, if at least one player fails to join the agreement in the first stage then no agreement forms. In this case, all deposits are instantly refunded to those players that joined the agreement and all players revert back to the baseline game in stage two.

Therefore, if an agreement forms in the first stage of the game, individuals face the following payoff functions during stage two

$$\pi^c = 9 + 4q_{-i}$$

$$\pi^{nc} = 13 + 4q_{-i} - d,$$

(2)

where the superscripts c and nc indicate compliance and noncompliance by members to an agreement, respectively. Recall, that if at least one player does not join, then no agreement forms and all players face the same payoff function in [1] during stage two. With this agreement mechanism, members earn higher payoffs complying provided that  $d > 4$ . Indeed, as long as  $d > 4$  the efficient solution can be implemented as part of a subgame-perfect Nash equilibrium; meaning, all five players join the agreement in the first stage and then contribute \$8 to the public account in the second stage. We set the deposit amount at \$5 for our experiments. The solution is straightforward and is found using backward induction. While in the second stage, there are two possible scenarios that players may face. Consider first the scenario in which no agreement forms in the first stage. In this case the game-theoretic prediction, like the baseline game, is that no player contributes to the public account and each player earns \$13. The other possible scenario in stage two is that an agreement has formed at the end of stage one. In this case each member decides between complying with the agreement and contributing their \$8 or not complying and keeping their \$8. A comparison of the payoff functions in equation [2] illustrates that in this subgame all agreement members are expected to comply with their commitments. When compliance is full the value of the public account is \$20, and since each player is refunded their \$5 deposit they earn \$25. Given the expected earnings in stage two, joining the agreement in stage one is a weakly dominant strategy for each player.

Our third and final treatment, *partial participation*, examines the deposit-refund mechanism when full participation is not required. This treatment proceeds like the second

**Table 1 Expected Earnings for agreement members and nonmembers**

	Agreement size		
	3	4	5
Member	\$17	\$21	\$25
Nonmember	\$25	\$29	–

treatment, but now an agreement will form if at least three players join in stage one. Therefore, when agreements form, they can range from having three, four or five members. This means that in cases in which participation is less than full, members and nonmembers can coexist. Members, as before, commit to contributing \$8 to the public account and must pay an initial deposit of \$5. Nonmembers, in contrast, do not commit

to contributing \$8 and also do not pay the \$5 deposit. Rather, nonmembers simply play the baseline game.

We again solve the game using backward induction. When an agreement fails to form, as in this second treatment, players are expected to earn \$13. On the other hand, from the comparison of the payoff functions in equation [2] and a deposit amount of \$5, when agreements do form we expect members to comply with their commitments and contribute \$8 to the public account (the cost of noncompliance (\$5) is greater than the benefit (\$4)). Nonmembers, if they exist, are expected not to contribute to the public account and earn \$13 plus the value of the public account. Therefore, nonmembers will earn strictly more than members (\$8 more) and hence there is a financial incentive to not join the agreement and to free ride on the contributions to the public account. Table 1 shows the *expected* earnings for members and nonmembers given the three possible agreement sizes.

A comparison of the payoffs from Table 1 illustrates that while nonmembers always earn more than agreement members, all players are better off when an agreement forms compared to having no agreement at all (recall, players expect to earn \$13 when an agreement does not form). Hence players have two conflicting incentives; on one hand the incentive to free ride encourages nonparticipation, and on the other hand the incentive to ensure an agreement forms encourages participation. This tension allows for two subgame-perfect Nash equilibria in this game. An equilibrium exists in which no agreement forms (less than three players join in the first stage) and no contributions are made to the public account in the second stage. In addition, an agreement of size three can exist in equilibrium in which the members fully comply with their commitments and contribute to the public good.<sup>17</sup> In short, we expect this third treatment to be weakly more efficient than the baseline treatment and less efficient than the full participation treatment.

Participants were recruited from the general student body at Appalachian State University. For each of the three treatments, we conducted a session with 15 subjects. The subjects were placed in groups of five and the groups were reshuffled for each of 10 rounds. For each session, all of the rules and parameters remained constant over the ten rounds. Subjects were informed that they would only be paid for one round of play and that round would be chosen at random after all of the rounds were completed. To mitigate end-game effects, the subjects were not told the exact number of rounds. The experiments were hand run and lasted roughly one hour for the baseline treatment and 1.5 h for the deposit-refund treatments. The experiment yielded 450 individual-level observations, 150 for each treatment.

In the baseline treatment, subjects were provided with a *contribution decision sheet* at the start of each round that contained a table displaying all of their earnings possibilities as well as check boxes that subjects used to indicate their decisions. Subjects had to check whether or not to contribute \$8 to the public account. After all subjects in the room were finished making their decisions, the sheets were collected and the contributions were tallied. At that point the results for the round were recorded onto the

subjects' decisions sheets and handed back. At the end of each round subjects were informed (via their decision sheet) of how many players in their group contributed \$8 to the public account and their individual earnings. Subjects were also provided with a record sheet on which they recorded the results for each round of play.

In the other two treatments, the contribution decision was preceded by an agreement decision. An *agreement decision sheet* was used by subjects to indicate their decision whether or not to join the agreement. This sheet also contained tables that displayed all of their earnings possibilities. After all of the participants made their agreement decision, the sheets were collected, the decisions were tabulated and the results were recorded on the individual decision sheets. The sheets were then returned to the subjects. At that point each subject was informed of how many subjects joined the agreement, whether or not it formed and whether or not a deposit of \$5 was deducted. Then subjects moved to the contribution decision, as described in the baseline. Using the contribution decision sheet, subjects indicated whether or not they contribute \$8 to the public account. Once subjects were finished making their contribution decisions, the sheets were collected, the results tabulated and recorded. Subjects were then informed of how many group members contributed to the public account, whether their deposit was returned and their earnings for the round. Subjects recorded their round earnings on their record sheet and the next round began.

### 3 Results

Figure 1 illustrates the percentage of players that contributed to the public good by treatment and round. Numbers from the baseline setting initially reveal the familiar pattern of positive, though less than socially optimal, contributions that decay over time (Ledyard 1995; Chaudhuri 2011), but in later rounds, contributions exhibit a more stable pattern that is often observed in the closely related threshold public good setting (e.g., Croson and Marks 2000). Initially, 73 percent of players contribute, but the number falls to about 50 percent during the session and ends at 60 percent in the final rounds. Table 2 summarizes the contribution levels across all rounds for each treatment. The average contribution level in the baseline treatment was 60 percent.

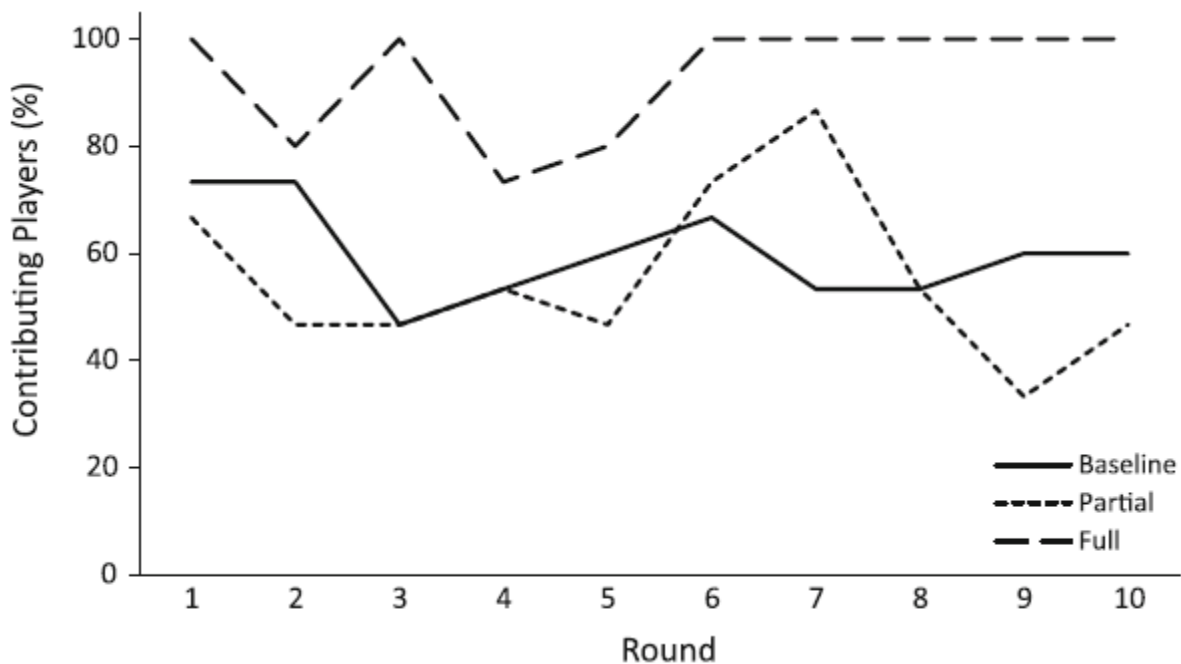
**Result 1** A deposit-refund mechanism that requires full participation leads to nearly efficient public-good provision.

Consistent with theoretical predictions, 97.3 percent of subjects (146/150) chose to join the agreement, which led to agreements forming in 90 percent of trials (27/30). As shown in



**Table 2 Percentage contributing by treatment**

	Pooled		
	Mean	SD	N
Baseline	0.600	0.491	150
Partial	0.553	0.499	150
Full	0.933	0.250	150
Total	0.696	0.461	450



**Fig. 1 Percentage contributing by treatment over rounds**

Fig. 1, the decision to join is universal (100%) in seven of the ten rounds, and such cooperation is sustained over the final five rounds. Moreover, when agreements formed subjects were perfectly compliant with their commitment to contribute to the public good. The combination of a high percentage of agreement formation and perfect compliance lead to an overall contribution level of 93.3 percent (Table 2). Therefore we find support for the theoretical prediction from section two, and from Gerber and Wichardt (2009), that a deposit-refund mechanism that requires full participation will lead to efficient public-good provision.

**Result 2** A deposit-refund mechanism that requires only partial participation does not improve upon average baseline public-good contribution levels.

When only partial participation was required, 54.7 percent of subjects (82/150) joined the agreement with agreements forming in only 53.3 percent of trials (16/30).<sup>19</sup> Overall, 55.3 percent of subjects (83/150) contributed to the public good (Table 2), which is statistically equivalent to contributions in the baseline setting (55.3 vs. 60.0%,  $p = 0.414$ ).<sup>20</sup> When agreements did form, they had an average of 3.69 members (out of five). These summary results indicate that the incentive to free ride in the partial participation setting caused low

**Table 3 Percentage contributing by treatment and membership decision**

	Partial			Full		
	Mean	SD	N	Mean	SD	N
Agreement formed	0.763	0.428	80	1.000	0.000	135
Members	0.983	0.130	59	1.000	0.000	135
Nonmembers	0.143	0.358	21	–	–	0
No agreement	0.314	0.467	70	0.333	0.488	15
Joined	0.261	0.449	23	0.454	0.522	11
Did not join	0.340	0.479	47	0.000	0.000	4
Total	0.553	0.499	150	0.933	0.250	150

levels of agreement formation, and less than full participation when agreements did form. However, the deposit-refund scenario that requires only partial participation was effective at achieving nearly full compliance from its members (98.3 percent). This implies that the deposit-refund mechanism, in both the partial and full participation cases, effectively facilitates contributions from the members when an agreement forms.

We examine the data in more detail by stratifying contribution decisions by whether the agreement formed and whether group members signed up to join the agreement. The numbers, as reported in Table 3, reveal key elements behind the overall performance measures of the deposit-refund mechanisms.

**Result 3** When agreements form, a deposit-refund mechanism that requires only partial participation causes an increase in contributions relative to the baseline.

Considering only the partial participation case, we see from Table 3 that contribution levels when an agreement formed were more than double the levels observed when an agreement does not form; 76.3 vs. 31.4 percent. A comparison of the contribution rates under the deposit-refund mechanism when agreements formed and the baseline case shows a significant increase in public-good provision (76.3 vs. 60.0%,  $p = 0.014$ ). When agreements form, most of the public-good contributions are provided by members; 98.3 percent of members contribute and 14.3 percent of nonmembers contribute. Note further that the percentage of nonmembers contributing in this scenario is significantly lower than the percentage of contributors in the baseline (14.3 vs. 60.0%,  $p = 0.000$ ).

Therefore free-riding is exacerbated among nonmembers when partial agreements form.

**Result 4** When agreements fail to form, deposit-refund mechanisms cause a decrease in public-good contributions relative to the baseline.

An interesting finding is the stark difference in behavior when agreements failed to form and the baseline. Recall, if an agreement does not form, the group and its members remain in the baseline public-good setting. Thus, at the time of the contribution decision, the only difference between the deposit-refund treatments and the baseline is a previously failed attempt to form an agreement. The data suggest that this failure matters a great deal. In the baseline treatment, 60.0 percent of members contributed to the public good, which is nearly twice the levels

**Table 4 Probit models: determinants of contribution decision**

	Model 1	Model 2	Model 3
Constant	0.637** (0.314)	0.536* (0.329)	0.762** (0.347)
D-R full	1.316*** (0.319)	1.783*** (0.314)	2.510*** (0.407)
D-R partial	-0.117 (0.325)	0.669* (0.136)	2.286*** (0.498)
D-R & agreement failed	-	-1.620*** (0.293)	-3.118*** (0.436)
D-R partial & agreement formed & did not join	-	-	-3.709*** (0.642)
Chi-square	92.93***	93.26***	143.85***
N	450	450	450

For each probit probability model, the dependent variable is the subject's binary decision whether to contribute to the public good. Estimates use robust standard errors (in parentheses) and clustering at the individual level. Models control for period-specific fixed-effects. \*, \*\* and \*\*\* indicate significance at the 10, 5 and 1 % levels

observed in the full and partial participation treatments (33.3 and 31.4 %, respectively). It appears the failure of forming an agreement can undermine cooperation. We confirm these results with a conditional analysis by estimating a series of probit models of individual contribution decisions. In each model, the dependent variable is the subject's decision to contribute to the public good (1 if contribute; 0 otherwise). Estimates control for period-specific fixed-effects and take into account that individual subjects make repeated decisions by employing robust standard errors and clustering at the individual level. The results are reported in Table 4. We begin with a basic model in which the dependent variable is estimated as a function of the two deposit-refund treatments (omitting the baseline). The results are consistent with our initial findings; that is, a deposit-refund scheme that requires full participation increases a subject's likelihood of contributing to the public good relative to the baseline (marginal effect =

0.333), while contributions are not significantly different with a partial participation scheme.

The second model adds an additional independent variable that interacts the presence of a deposit-refund mechanism (full or partial) with the failure of an agreement forming. The estimate for the interaction variable indicates how the failure to form an agreement influences the likelihood of contributing to the public good, while controlling for individual treatment effects. Results reveal a negative and highly significant relationship (marginal effect =  $-0.566$ ), indicating that subjects are less likely to contribute to the public good after a failed attempt to form an agreement, as compared to the absence of any opportunity to form an agreement. This indicates that while successful agreement formation significantly improves cooperation, an unsuccessful effort might lead to worse outcomes in comparison to not pursuing an agreement. The final model adds an interaction variable to isolate players that do not join a partial agreement to disentangle the effect of a successfully formed partial agreement on members and nonmembers. Estimates show that the new interaction term is significant and negative (marginal effect =  $-0.878$ ) while the previously insignificant coefficient for the partial participation term is now significant and positive (marginal effect =  $0.364$ ). Results show the formation of a partial agreement leads to greater contributions from members while also exacerbating free-riding from nonmembers.

#### **4 Conclusion**

Without a global governmental system that can provide direct regulation of resource use, countries are left to manage shared resources voluntarily through environmental agreements. While a decent amount of research has focused on the challenge of getting nations to voluntarily participate in these agreements, very little research has examined the problem of maintaining compliance with the commitments of the members. This paper uses lab experiments to evaluate the performance of one promising mechanism—a deposit-refund mechanism—in which players pay a deposit upon joining an agreement and are later refunded the deposit if they are not in violation of their commitments. Our results provide an empirical test of the deposit-refund mechanism recently proposed by Gerber and Wichardt 2009, but adapted here within an agreement formation game to provide a public good.

Our results show that the deposit-refund mechanism has differential impacts, depending on whether full or partial participation is required. When the participation of all players is required before an agreement enters into force, the mechanism achieves near-perfect efficiency levels. That is, agreements form in almost every trial and members are perfectly compliant when agreements do form (all deposits are refunded). The result confirms the encouraging theoretical prediction of Gerber and Wichardt (2009) that the deposit-refund mechanism will achieve significant efficiency gains relative to a voluntary contribution mechanism. However, our research explores an important variation of the deposit-refund system not analyzed by Gerber and Wichardt. That is, we show that when less than full participation is required for an agreement to enter into force—the situation that characterizes most global environmental agreements (e.g., Kyoto

Protocol)—the deposit-refund mechanism does not increase average efficiency measures. Agreements form only about half of the time. When agreements do form, the deposit-refund mechanism is successful at motivating full compliance of its members and therefore a subset of the players act more cooperatively relative to the baseline case. At the same time, however, a deposit-refund mechanism that requires only partial participation exacerbates free-riding behavior in two important ways. First, when agreements form the nonmembers are less cooperative compared to players in the absence of such a mechanism. Second, when agreements fail to form players act significantly less cooperative than when they had no opportunity to join an agreement. Therefore it appears that a previously failed attempt at forming an agreement can undermine cooperation.

Our experiments, like the model proposed by Gerber and Wichardt, necessarily abstract from many of the practical implications of implementing a deposit-refund system within an international environmental agreement. However, the basic mechanism is simple and its most critical element is that parties empower a neutral third-party institution with the power to withhold financial deposits in cases of noncompliance. The financial institution would serve as an escrow account in which monies will be disbursed given the fulfillment of contractual emissions abatement responsibilities. Since third-party financial institutions have been implemented within other IEAs (e.g., Multilateral Fund under the Montreal Protocol), this challenge appears to be achievable. In the model we test, players face predefined emissions abatement commitments, deposit amounts and a minimum number of countries required to ratify for entry into force. Apart from paying deposits, stage one in our experiments can be likened to Annex I countries deciding whether to ratify the Kyoto Protocol given defined emissions limits (relative to their 1990 levels) and a minimum participation constraint (at least 55 of the parties needed to ratify).

In our experiments, players have identical payoff functions. In reality, countries involved in IEAs typically have differential commitments as members, and most importantly, different incentives to violate those commitments. Since the up-front deposit paid by each member must weakly exceed the financial incentive to cheat, heterogeneous members should also face different deposit amounts. To provide a rough idea of the size of these deposits, a report by the Energy Information Administration (1998) compiled cost estimates for the United States to comply with the Kyoto Protocol (a seven percent reduction in 1990 level emissions) that ranged from \$52 billion to \$437 billion (0.5–4.2% of 2010 GDP). Clearly, deposit amounts for large greenhouse-gas emitting nations are non-trivial. Moreover, uncertainty in the benefit/cost functions or the realities of incomplete (or inaccurate monitoring) will likely play an important role in determining optimal deposit sizes.

Our overarching results suggest that a deposit-refund mechanism can be very effective when agreements require full participation, but is less effective, and potentially disruptive, when agreements require only partial participation. While these results help inform the design of effective institution formation, there are some obvious extensions that should be carried out to explore other relevant issues. As highlighted previously, a

useful extension would be to introduce heterogeneity into the players' payoff functions and commitment levels, which would translate into requiring differential deposit amounts. While this is likely not a problem theoretically, it is unclear how this will impact the effectiveness of the mechanism. If parties exhibit other-regarding preferences (like inequality aversion), which is sometimes demonstrated to be the case (Charness and Rabin 2002), having differential commitments and deposits could undermine the positive features of the mechanism. Other useful extensions include introducing uncertainty by making the benefits of cooperation uncertain and/or by relaxing the assumption of perfect and complete monitoring for compliance.

## References

- Bagnoli M, McKee M (1991) Voluntary Contributions Games: Efficient Private Provision of Public Goods. *Econ Inquiry* 29(2):351–366
- Barrett S (1994) Self-enforcing international environmental agreements. *Oxf Econ Pap* 46(1):878–894
- Barrett S (1997) The strategy of trade sanctions in international environmental agreements. *Resour Energy Econ* 19(1):345–361
- Barrett S (2003) *Environment and statecraft: the strategy of environmental treaty-making*. Oxford University Press, Oxford and New York
- Barrett S (2008) Climate treaties and the imperative of enforcement. *Oxf Rev Econ Policy* 24(2):239–258
- Barrett S, Stavins R (2003) Increasing participation and compliance in international climate Change agreements. *Int Environ Agreem Polit Law Econ* 3:349–376
- Breton M, Sbragia L, Zaccour G (2010) A dynamic model for international environmental agreements. *Environ Resour Econ* 45(1):25–48
- Cadsby CB, Maynes E (1999) Voluntary provision of threshold public goods with continuous contributions: experimental evidence. *J Pub Econ* 71(1):53–73
- Carraro C, Siniscalco D (1993) Strategies for the international protection of the environment. *J Pub Econ* 52(3):309–328
- Carraro C, Siniscalco D (1998) International environmental agreements: incentives and political Economy. *Eur Econ Rev* 42(3–5):561–572
- Carraro C, Marchiori C, Oreffice S (2009) Endogenous minimum participation in international environmental treaties. *Environ Resour Econ* 42(3):411–425
- Cason TN, Plott CR (1996) EPA's new emissions trading mechanism: a laboratory evaluation. *J Environ Econ Manag* 30(2):133–160
- Charness G, Rabin M (2002) Understanding social preferences with simple tests. *Q J Econ* 117(3):817–869
- Chaudhuri A (2011) Sustaining cooperation in laboratory public goods experiments: a selective Survey of the Literature. *Exp Econ* 14(1):47–83
- Croson R, Marks M (2000) Step returns in threshold public goods: a meta- and experimental analysis. *Exp Econ* 2:239–259

Dannenber A, Lange A, Sturm B (2010) On the formation of coalitions to provide public goods—experimental evidence from the lab. National bureau of economic research working paper no. 15967

Dawes RM, Orbell JM, Simmons RT, Vande Kragt AJC (1986) Organizing groups for collective action. *Am Polit Sci Rev* 80(4):1171–1185

Energy Information Administration (1998) Impacts of the Kyoto Protocol on United States energy markets and economic activity. Office of integrated analysis and forecasting, United States Department of Energy, Washington, DC

Erev I, Rapoport A (1990) Provision of step-level public goods: the sequential contribution mechanism. *J Confl Resolut* 34(3):401–425

Finus M, Rundshagen B (1998) Toward a Positive Theory of Coalition Formation and Endogenous Instrument Choice in Global Pollution Control. *Public Choice* 96(1–2):145–186

Finus M (2008) The enforcement mechanisms of the Kyoto Protocol: flawed or promising concepts. *Lett Spatial Resour Sci* 1(1):13–25

Fullerton D, Wu W (1998) Policies for green design. *J Environ Econ Manag* 36(2):131–148

Gerber A, Wichardt P (2009) Providing public goods in the absence of strong institutions. *J Pub Econ* 93 (3–4):429–439

Hoel M (1992) International environmental conventions: the case of uniform reductions of emissions. *Environ Resour Econ* 2(2):141–159

Hoel M, Schneider K (1997) Incentives to participate in an international environmental treaty. *Environ Resour Econ* 9(2):153–170

Hovi J, Froyen CB, Bang G (2007) Enforcing the Kyoto Protocol: can punitive consequences restore compliance?. *Rev Int Stud* 33(3):435–449

Isaac MR, Schmidtz D, Walker JM (1989) The assurance problem in a laboratory market. *Pub Choice* 62(3):217–236

Kolstad C (2007) Systematic uncertainty in self-enforcing international environmental agreements. *J Environ Econ Manag* 53(1):68–79

Kosfeld M, Okada A, Riedl A (2009) Institution formation in public goods games. *Am Econ Rev* 99(4): 1335–1355



Ledyard J (1995) Public good: a survey of experimental results. In: Kagel J, Roth A (eds) *The handbook of experimental economics*. Princeton University Press, Princeton pp 111–194

Marks M, Croson R (1998) Alternative rebate rules in the provision of a threshold public good: an experimental investigation. *J Pub Econ* 67(2):195–220

Marks M, Croson R (1999) The Effect of incomplete information in a threshold public goods experiment. *J Pub Econ* 67(2):195–220

McEvoy D (2010) Not it: opting out of voluntary coalitions that provide a public good. *Pub Choice* 142(1–2):9–23

McEvoy D, Murphy JJ, Spraggon JM, Stranlund JK (2011) The problem of maintaining compliance with stable coalitions: experimental evidence. *Oxf Econ Pap* 63(3):475–498

McEvoy D, Stranlund J (2009) Self-enforcing international environmental agreements with costly monitoring for compliance. *Environ Resour Econ* 42:491–508

McGinty M (2007) International environmental agreements among asymmetric nations. *Oxf Econ Pap* 59(1):45–62

Plott C (1987) Dimensions of parallelism: some policy applications of experimental methods. In: Alvin Roth (ed) *Laboratory experimentation in economics: six points of view*. Cambridge University Press, Cambridge

Rapoport A, Eshed-Levy D (1989) Provision of step-level public goods: effects of greed and fear of being gypped. *Organ Behav Hum Decis Process* 44(3):325–344

Rapoport A, Suleiman R (1993) Incremental contribution in step-level public goods games with asymmetric players. *Organ Behav Hum Decis Process* 55(2):171–194

Rauchdobler J, Sausgruber R, Tyran J (2010) Voting on thresholds for public goods: experimental evidence. *Pub Finance Anal* 66(1):34–64

Rubio SJ, Ulph A (2006) Self-enforcing international environmental agreements revisited. *Oxf Econ Pap* 58(2):233–263

Rubio SJ, Ulph A (2007) An infinite-horizon model of dynamic membership of international environmental agreements. *J Environ Econ Manag* 54(3):296–310

Shogren JF, Hurley TM (1999) Experiments in environmental economics. In: van den Bergh JCJM (ed) *Handbook of environmental and resource economics*. Edward Elgar, Cheltenham

Stranlund JK, Murphy JJ, Spraggon JM (2011) An experimental analysis of compliance in dynamic emissions markets. *J Environ Econ Manag* 62(3):414–429

Suleiman R, Rapoport A (1992) Provision of step-level public goods with continuous contribution. *J Behav Decis Mak* 5(2):133–153

Tavoni A, Dannenberg A, Kallis G, Löschel A (2011) Inequality, communication and the avoidance of disastrous climate change in a public goods game. *Proc Natl Acad Sci* 108(29):11825–11829

Tourney D, Fujiwara N (2010) National commitments, compliance and the future of the Kyoto Protocol. *Centre Eur Policy Stud* 226:1–228

Van de Kragt AJC, Orbell JM, Dawes RM (1983) The minimal contributing set as a solution to public goods problems. *Am Polit Sci Rev* 77(1):112–122