Department of Economics Working Paper

Number 11-01 | April 2011

The Endogenous Formation of Coalitions to Provide Public Goods: Theory and Experimental Evidence

David M. McEvoy
Appalachian State University

Todd L. Cherry
Appalachian State University

John K. Stranlund
University of Massachusetts at Amherst

Department of Economics
Appalachian State University
Boone, NC 28608
Phone: (828) 262-6123
Fax: (828) 262-6105
www.business.appstate.edu/economics

The Endogenous Formation of Coalitions to Provide Public Goods: Theory and Experimental Evidence

David M. McEvoy (corresponding author)
Department of Economics
Appalachian State University
3108 Raley Hall, 416 Howard Street
Boone, NC 28608
Phone: 828-262-6126
Email: mcevoydm@appstate.edu

Todd L. Cherry
Department of Economics
Appalachian State University

John K. Stranlund
Department of Resource Economics
University of Massachusetts at Amherst

Abstract: This paper examines the endogenous formation of coalitions that provide public goods in which players implement a minimum participation requirement before deciding whether to join. We demonstrate theoretically that payoff-maximizing players will vote to implement efficient participation requirements and these coalitions will form. However, we also demonstrate that if some players are averse to inequality they can cause inefficient outcomes. Inequality-averse players can limit free riding by implementing larger than efficient coalitions or by blocking efficient coalitions from forming. We test the theory with experimental methods and observe individual behavior and coalition formation consistent with a model of inequality-averse players.

1. Introduction

The provision of public goods constitutes a social dilemma in which individually rational choices are not collectively rational. Absent appropriate institutions, voluntary contributions to public goods will be underprovided relative to the social optimum. The analysis and design of institutions to coordinate voluntary provision of public goods is important for understanding how groups may overcome this social dilemma. Most economic studies on institutions in public goods games analyze the performance of exogenous institutions; however, in many contexts there is no governing body that can impose a set of rules on a group of agents. Important examples are found in the context of international public goods. The management of many international resources constitute social dilemmas in which countries' unilateral management of resources result in suboptimal international management (e.g., green-house-gas emissions, nuclear non-proliferation). Because nations are sovereign, any institutional arrangement created to increase provision of an international public good has to be developed endogenously.

In this paper we explore, theoretically and experimentally, a form of endogenous institution formation in which agents voluntarily form coalitions to provide a public good. Prior to making the decision whether to join the coalition all players vote on the minimum number of members required for a coalition to form. A coalition forms if enough players voluntarily join so that the minimum membership requirement is met or surpassed, and only then are members required to make contributions to the public good. Many international treaties and almost all international environmental treaties contain minimum participation requirements like the one explored in this paper. For example, the Kyoto Protocol required ratification by at least 55 parties accounting for at least 55 percent of the total 1990 greenhouse gas emissions. The Montreal Protocol on Substances that Deplete the Ozone Layer entered into force when at least 11 countries ratified and members needed to represent at least two-thirds of the total 1986 consumption of ozone-depleting substances. The Treaty on the Non-Proliferation of Nuclear Weapons required ratification by the then five nuclear nations plus 40 additional nations for entry into force. Some treaties require all affected parties to join; for example, the Convention

for the Protection of the Marine Environment of the North-East Atlantic required ratification by all 13 parties.[1]

This study adds to a growing literature on endogenous institution formation to confront social dilemmas (see, for examples, Walker et al. 2000; Gurerk et al. 2006; Tyran and Feld 2006; Kroll et al. 2007; Burger and Kolstad 2009; Kosfeld et al. 2009; Sutter et al. 2010). Our work is perhaps closest to that of Kosfeld et al. (2009). They also analyze the formation of a voluntary coalition designed to increase contributions to a public good. In their game players first decide whether to join a coalition, and then later the members vote on whether to contribute to the public good. Hence, in their analysis members of a coalition decide what the coalition should accomplish *after* they make their participation decision. In contrast, the players in our study understand *ex ante* what they are required to do in a coalition. This is closer to the actual process of coalition formation to provide international public goods, where countries typically decide the commitments of coalition members and what triggers a coalition to form *before* they decide whether to join.

We examine this institution both theoretically and with laboratory experiments. Our theoretical model is a two-stage public goods game in which individuals with identical financial payoffs decide whether to contribute a single unit of the public good. In the first stage of the game all members of a group vote on the adoption of a minimum membership requirement that must be met if a coalition is to form. In the second stage each individual decides whether they will join the coalition. If enough individuals join to meet the membership requirement, a coalition forms, the members of the coalition each provide their unit of the public good, while nonmembers keep theirs. A unique feature of the theoretical model and our experiments is that efficient provision of the public good may require that all individuals contribute (i.e., the grand coalition is efficient), or efficiency may require that only a subset of individuals contribute.

To our knowledge, Carraro et al. (2009) provide the only theoretical account of endogenous membership requirements for coalitions that form to provide public goods (they focus on the provision of international environmental quality, but their model can be applied more generally). They show that this feature increases participation in these coalitions. Our initial theoretical model is a special case of the Carraro et al. model. We demonstrate that, if all

---

[1] See Barrett (2003) for a more thorough examination of participation requirements in international environmental agreements.

players only care about maximizing their expected financial payoffs, they will vote to implement an efficient coalition size as the membership requirement and this coalition will form. This is true regardless of whether the efficient coalition is the grand coalition or some smaller coalition.

We then extend the coalition formation model to allow for the possibility that some individuals are inequality averse. Individual preferences for equity or fairness have been well documented in the economics literature (Fehr and Schmidt 1999; Fehr and Gachter 2000a; Fehr and Gachter 2000b; Charness and Rabin 2002; Camerer 2003; Engelmann and Strobel 2004; Fehr et al. 2006). We suspect that these preferences may be particularly important for situations in which efficiency requires only partial coalition participation, because in these cases it is optimal that some individuals free ride, and hence, earn higher payoffs than coalition members. We demonstrate that if some players are averse to inequality they can move groups to inefficient outcomes, particularly if an efficient coalition is smaller than the grand coalition. Inequality-averse individuals can cause larger-than-efficient coalitions to form, and they may incur financial losses to block efficient coalitions from forming.

We conducted laboratory experiments to test the theory. In treatments for which the efficient coalition was the grand coalition, coalitions formed most of the time and most of these were efficient. However, in treatments for which the efficient coalition required only half the members of a group, coalitions formed in just over half the trials, but only half of these were efficient. In fact, the grand coalition, despite being inefficient, formed at close to the same rate (18%) as the efficient coalition (23%). Moreover, the efficient coalition formed in only about half the trials in which it was adopted as the membership requirement. These results are not consistent with a theoretical model of players who are only concerned with material payoffs, but they are consistent with a model that includes inequality-averse players. Our experiments also demonstrate the value of the endogenous implementation of a minimum membership requirement in promoting coalition formation and public good provision, at least when the efficient coalition is the grand coalition. In these cases we observe significant efficiency gains compared to previous studies that use similar incentives but without endogenous minimum membership requirements (Kosfeld et al., 2009; Dannenberg et al., 2010; McEvoy et al., 2010).

2. Theory

In this section we lay out the theoretical underpinnings of our experiments. We first present a model of identical, risk-neutral players who vote on the number of members a coalition must contain for it to form, and then decide whether to join such a coalition. Players in this model seek to maximize their expected financial payoffs. Like Kosfeld et al. (2009), we say that such players have *standard* preferences. We then extend the model to examine the consequences of the presence of some players with an aversion to inequality in the spirit of Fehr and Schmidt (1999).

2.1 A model of coalition formation with standard preferences

Consider a game in which n-players decide whether to contribute a single unit to a public good. Let the number of individuals who contribute to the public good (and the total supply of the public good) be s. Let $b > 0$ denote the shared benefit players receive from contributions to the public good up to $\bar{s} \leq n$ and let $c > 0$ denote the individual cost of contributing. The basic financial payoff function of a player that contributed to the public good is

$$\pi_i = A + bs - c, \text{ for } s \leq \bar{s} , \qquad [1]$$

where A is a positive constant. Assume that $c > b$ so that no player would contribute to the public good in a non-cooperative Nash equilibrium. However, suppose that collective welfare is maximized when $\bar{s}$ individuals contribute their units to the public good. We allow for the possibility that it is efficient that the public good be provided by a coalition that is smaller than the grand coalition. We assume a simple case of this in which individual contributions yield a return of b up to an aggregate level and additional contributions yield a return of zero. Examples of this include any project for which contributions in excess of what is required to provide the good do not yield public benefits. Our specification also covers cases in which collective welfare is maximized when everyone contributes their unit to the public good; that is, when $\bar{s} = n$. Note that if $\bar{s}$ is efficient then $n(A + b\bar{s}) - c\bar{s} > nA$, which requires $nb > c$.

Given the motivation for collective action to provide the public good, suppose that the players are able to form a cooperative coalition in a two-stage game. In the first stage (*the voting stage*), all players vote on the minimum number of members required for a coalition to form. We call the outcome of this vote the *minimum membership requirement*. In our experiments we implement a modal-response voting rule, so that the number that receives the most votes becomes the membership requirement, but other rules are possible. In stage two of the coalition

formation game (*the coalition stage*), the players decide independently whether or not to join the coalition. If enough players join so that the membership requirement is met, the coalition forms and its members provide their units of the public good. Throughout we call such a coalition an *effective coalition*. Those that do not join an effective coalition do not provide their units of the public good (because $c > b$), but still benefit from its provision. If the minimum membership requirement is not met, an effective coalition does not form, and no player contributes to the public good.

The equilibrium of this game is found by backward induction, so we start by describing the coalition stage. At this point in the game a minimum membership requirement has been chosen. Denote the membership requirement as s_p. Using [1], those who decide to join a coalition with s members earn:

$$\pi^m(s) = \begin{cases} A + bs - c & \text{for } s \leq \bar{s} \text{ if } s \geq s_p; \\ A + b\bar{s} - c & \text{for } s > \bar{s} \text{ if } s \geq s_p; \\ A & \text{if } s < s_p. \end{cases}$$

[2]

Throughout the superscript *m* indicates the player is a member of a coalition. Nonmembers, identified by the superscript *nm,* earn:

$$\pi^{nm}(s) = \begin{cases} A + bs & \text{for } s \leq \bar{s} \text{ if } s \geq s_p; \\ A + b\bar{s} & \text{for } s > \bar{s} \text{ if } s \geq s_p; \\ A & \text{if } s < s_p. \end{cases}$$

[3]

From [2] and [3] it is clear that when an effective coalition forms, nonmembers earn strictly higher profits than members because they enjoy the benefits of public good contributions without incurring the cost of contributing. However, if the minimum membership requirement is not met, both members and nonmembers earn their noncooperative payoffs.

Players will join an effective coalition only if it is profitable for them to do so in the sense that they earn at least as much in the coalition as they would if no coalition formed. Thus, a coalition is profitable for its members if and only if $\pi^m(s \mid s \geq s_p) = A + bs - c \geq A$. Assume throughout that a coalition with $\bar{s}$ members is profitable for individuals with standard preferences so that $\pi^m(\bar{s} \mid \bar{s} \geq s_p) \geq A$, or rather, $b\bar{s} - c \geq 0$. Note from [2] that $\pi^m(s \mid s \geq s_p)$ is increasing in s up to $\bar{s}$, and then is constant. For analytic convenience, we assume that s is a

6

continuous variable throughout. Therefore, the smallest possible profitable coalition is the solution to $\pi^m(s \mid s \geq s_p) = A + sb - c = A$, yielding

$$s_{\min} = c / b. \tag{4}$$

Since a coalition with $\bar{s}$ members is profitable, $\bar{s} \geq s_{\min}$. For an effective coalition to form in the second stage of the game both the membership requirement and the profitability requirement must be satisfied; that is, a coalition can form if $s_{\min} \leq s_p$. However, it is not guaranteed that a coalition forms even when $s_{\min} \leq s_p$. Other equilibria are possible, including one in which an effective coalition does not form. In addition, when $s_{\min} \leq s_p < n$, coalitions of size $s \in \{s_p, s_p + 1, ..., n\}$ can exist in equilibrium. However, if we require that equilibrium coalitions satisfy a *strictness* refinement then we have a clear prediction from the coalition stage. A Nash equilibrium is considered strict if every player is strictly worse off by deviating from equilibrium play. Under the strictness refinement, if $s_{\min} \leq s_p$ then a coalition of size s_p will form in the second stage of the game.[2]

　　Now let us examine the determination of s_p in the voting (first) stage of the game. In this stage players simultaneously submit votes for the minimum membership requirement to be implemented in the coalition formation stage. It is straightforward to show that a player with standard preferences would never vote to implement a membership requirement for a coalition that is unprofitable, so we do not have to consider membership requirements below $s_{\min}$. For larger coalitions, we assume that players vote to implement membership requirements that yield their highest expected payoff, where the uncertainty is about whether they will eventually join an effective coalition. Since the players are identical it is reasonable to assume that each of them believes that they have the same chance of joining an effective coalition as any other member. Therefore, a player's evaluation of the probability of being part of an effective coalition is s/n, and his or her subjective expected payoff is:

---

[2] Kosfeld et al. (2009) also appeal to the strictness refinement to arrive at a unique equilibrium prediction in the stage game. The equilibrium in their game is the minimum size profitable coalition (under standard preferences). In contrast, the equilibrium in our game is the minimum membership requirement which will be a weakly larger than the minimum size profitable coalition.

$$v(s) = \pi^m(s)(s/n) + \pi^{nm}(s)((n-s)/n).$$

$$[5]$$

Substitute [2] and [3] for $s \geq s_p$ into [5] and differentiate to obtain

$$v'(s) = \begin{cases} (nb-c)/n > 0, & \text{for } s \leq \bar{s}; \\ -c/n < 0, & \text{for } s > \bar{s}, \end{cases}$$

$$[6]$$

which indicates that the expected payoff of every player is increasing in the size of a cooperative coalition between $s_{\min}$ and $\bar{s}$. This, in turn, implies that under the modal response voting rule, as well as more restrictive voting rules (e.g., unanimity, super majority, or majority), all players will vote to implement the efficient coalition membership requirement in the first stage of the game. In the second stage, the efficient coalition size will form. Under these circumstances, therefore, players are expected to form the efficient cooperative coalition endogenously, and this is independent of whether efficiency requires full participation (i.e., the grand coalition) or only partial participation.

2.2 A model of coalition formation with inequality-averse players

In social dilemma games such as this one, it is possible that players have preferences beyond financial self-interest. In coalition formation games, inequality aversion, triggered by the difference between the payoffs of coalition members and free-riding nonmembers, may affect players' voting and participation decisions. To investigate this possibility, we amend our coalition formation model to allow individuals to be inequality averse as in Fehr and Schmidt (1999).[3]

To incorporate inequality aversion, first suppose that $\bar{s} = n$ and $s_p = 0$. Given the financial payoffs [2] and [3] with these restrictions, define the utility of a member of an effective coalition with s members as

$$u_i^m(s) = \pi_i^m(s) - \frac{\alpha_i}{n-1} \sum_{j \neq i} \max\left(\pi_j - \pi_i^m(s), 0\right) - \frac{\beta_i}{n-1} \sum_{j \neq i} \max\left(\pi_i^m(s) - \pi_j, 0\right),$$

$$[7]$$

---

[3] We adapt the model of inequality aversion developed by Fehr and Schmidt (1999) because it translates nicely into a model of coalition formation. However, other models of inequality aversion and social preferences exist (e.g., Rabin 1993; Bolton and Ockenfels 2000; Charness and Rabin 2002; Cox et al. 2007).

where $\alpha_i > 0$ captures the player's loss from disadvantageous inequality and $\beta_i > 0$ captures her loss from advantageous inequality. Since $\pi_j^{nm}(s) - \pi_i^m(s) = c$ and $\pi_j^m(s) - \pi_i^m(s) = 0$ from [2] and [3], [7] can be written as

$$u_i^m(s) = A + bs - c - \frac{\alpha_i c(n-s)}{n-1}. \qquad [8]$$

Similarly, the utility of a nonmember of an effective coalition with s members is:

$$u_i^{nm}(s) = \pi_i^{nm}(s) - \frac{\alpha_i}{n-1}\sum_{j \neq i}\max\left(\pi_j - \pi_i^{nm}(s), 0\right) - \frac{\beta_i}{n-1}\sum_{j \neq i}\max\left(\pi_i^{nm}(s) - \pi_j, 0\right), \qquad [9]$$

which can be written as

$$u_i^{nm}(s) = A + bs - \frac{\beta_i sc}{n-1}. \qquad [10]$$

The free-riding incentive is preserved in this model if $(n-1)(\beta_i - 1) < \alpha_i$; that is, as long as the aversion to advantageous inequality is not too strong relative to the aversion to disadvantageous inequality.[4]

    Incorporating the efficient coalition size $\bar{s} \leq n$ yields:

$$u_i^m(s) = \begin{cases} A + bs - c - \dfrac{\alpha_i(n-s)c}{n-1} & \text{for } s \leq \bar{s} \text{ if } s \geq s_p; \\[4mm] A + b\bar{s} - c - \dfrac{\alpha_i(n-s)c}{n-1} & \text{for } s > \bar{s} \text{ if } s \geq s_p; \end{cases} \qquad [11]$$

$$u_i^{nm}(s) = \begin{cases} A + bs - \dfrac{\beta_i sc}{n-1} & \text{for } s \leq \bar{s} \text{ if } s \geq s_p; \\[4mm] A + b\bar{s} - \dfrac{\beta_i sc}{n-1} & \text{for } s > \bar{s} \text{ if } s \geq s_p. \end{cases} \qquad [12]$$

All players earn A for coalition sizes that do not meet a minimum membership requirement.

    Using [11], an individual's minimum profitable coalition size can be characterized as:

---

[4] The freeriding incentive is preserved if $u_i^m(s) - u_i^{nm}(s) = c\left(s(\alpha_i + \beta_i) - \alpha_i n - (n-1)\right)/(n-1) < 0$ for each possible s. Note that $u_i^m(s) - \pi_i^{nm}(s)$ is increasing in s. Therefore, a player prefers to freeride on any effective coalition if $u_i^m(s) - \pi_i^{nm}(s)$ is negative when $s = n-1$. Calculate $u_i^m(n-1) - u_i^{nm}(n-1) = c\left((n-1)(\beta_i - 1) - \alpha_i\right)/(n-1)$, which is negative if $(n-1)(\beta_i - 1) < \alpha_i$.

$$\tilde{s}_i \begin{cases} = \hat{s}_i \text{ if } \hat{s}_i \le \overline{s} \\ > \hat{s}_i \text{ if } \hat{s}_i > \overline{s}, \end{cases} \quad \text{where} \quad \hat{s}_i = \frac{c(n-1) + \alpha_i cn}{b(n-1) + \alpha_i c}. \tag{13}$$

To demonstrate $\tilde{s}_i$, we first derive $\hat{s}_i$ as the solution to

$$A + bs - c - \frac{\alpha_i(n-s)c}{n-1} = A. \tag{14}$$

Since $u_i^m(s)$ in [11] is increasing in s, if $\hat{s}_i \le \overline{s}$ then $\hat{s}_i$ is i's minimum profitable coalition size.

However, if $\hat{s}_i > \overline{s}$, then $\tilde{s}_i$ must be the solution to

$$A + b\overline{s} - c - \frac{\alpha_i(n-s)c}{n-1} = A. \tag{15}$$

Now plug $\hat{s}_i$ into [14] and $\tilde{s}_i$ into [15], set the resulting equations equal to each other and collect

terms to obtain $b(\hat{s}_i - \overline{s}) = \alpha_i c(\tilde{s}_i - \hat{s}_i)/(n-1)$ which implies that $\tilde{s}_i > \hat{s}_i$ if $\hat{s}_i > \overline{s}$.

The specification of $\tilde{s}_i$ may be the most important result of this model because it reveals why inequality aversion can produce inefficient outcomes. Disadvantageous inequality aversion (i.e., $\alpha_i > 0$) implies that $\tilde{s}_i > s_{\min}$.[5] Thus, there are coalitions that are profitable in terms of financial payoffs, but an inequality-averse individual would not find profitable in terms of utility. This limits the range of coalitions that an individual finds acceptable: a larger-than-efficient coalition might be required to be profitable. We will see shortly how this can lead to the formation of inefficiently large coalitions. Moreover, such individuals might refuse to join an efficient coalition when it is smaller than the grand coalition, which can prevent efficient coalitions from forming. To understand how, define a *critical* player as one whose choice to not join a coalition prevents the coalition from forming. A critical individual *blocks* a coalition if he or she decides to not join. (This does not imply that a coalition forms if a critical player chooses to join—there are likely to be multiple critical players). For example, suppose as in our experiments that we have six players who vote to implement a membership requirement of three and assume that a three-player coalition is efficient. Suppose further that the first three players to make their second-stage choice decide to not to join the coalition. Then, each of the remaining three players is critical. Now, suppose that one of the remaining critical players has an aversion to payoff inequality. Although she is critical for the coalition to form, she may choose to opt out

---

[5] This follows from $\tilde{s}_i \ge \hat{s}_i$, $\hat{s}_i$ is increasing in α_i, and $\hat{s}_i = s_{\min}$ for $\alpha_i = 0$.

if she finds a three-player coalition unprofitable because of her aversion to disadvantageous inequality. In this way, an inequality-averse player can prevent an efficient coalition from forming.

Now let us turn to the players' voting preferences. Since they are heterogeneous because of differences in their aversion to inequality, it seems too restrictive to assume that each of them would believe that they have the same chance of being part of an effective coalition as any other player. Let $p_i(s)$ be i's subjective probability that it will join a coalition with s members in the coalition stage. This probability is zero for coalition sizes that i finds unprofitable and is equal to one for the grand coalition. That is, $p_i(s) = 0$ for $s < \tilde{s}_i$ and $p_i(n) = 1$. Suppose further that $p_i'(s) > 0$ in the range of $(\tilde{s}_i, n)$. A player's subjective expected payoff in this range is

$$v_i(s) = p_i(s)u_i^m(s) + (1 - p_i(s))u_i^{nm}(s) . \qquad [16]$$

Remove the minimum membership requirement from [11] and [12], substitute the result into [16], differentiate with respect to s in the range of $(\tilde{s}_i, n)$, and simplify terms to obtain the following. An individual's marginal expected utility when $\tilde{s}_i < \bar{s}$ is:

$$v_i'(s) = \begin{cases} p_i'(s)\left(-c - \dfrac{\alpha_i(n-s)c}{n-1} + \dfrac{\beta_i sc}{n-1}\right) + \dfrac{p_i(s)c(\alpha_i + \beta_i)}{n-1} + b - \dfrac{\beta_i c}{n-1}, \text{for } s \in [\tilde{s}_i, \bar{s}), \\[4mm] p_i'(s)\left(-c - \dfrac{\alpha_i(n-s)c}{n-1} + \dfrac{\beta_i sc}{n-1}\right) + \dfrac{p_i(s)c(\alpha_i + \beta_i)}{n-1} - \dfrac{\beta_i c}{n-1}, \text{for } s \geq \bar{s}. \end{cases} \qquad [17]$$

The individual's marginal expected utility when $\tilde{s}_i \geq \bar{s}$ is:

$$v_i'(s) = p_i'(s)\left(-c - \frac{\alpha_i(n-s)c}{n-1} + \frac{\beta_i sc}{n-1}\right) + \frac{p_i(s)c(\alpha_i + \beta_i)}{n-1} - \frac{\beta_i c}{n-1}. \qquad [18]$$

The first term in [17] and [18] is negative because $-c - \dfrac{\alpha_i(n-s)c}{n-1} + \dfrac{\beta_i sc}{n-1} = u_i^m(s) - u_i^{nm}(s)$ is negative for $s < n$ given that free-riding nonmembers are better off than coalition members. The second term in the sum is positive. $b - \beta_i c / (n-1)$ is positive as along as freeriders are better off with larger coalitions: above $\bar{s}$ this term reduces to $-\beta_i c / (n-1) < 0$. Thus, in all cases the sign of $v_i'(s)$ is indeterminate at this level of generality. This indeterminacy is due to inequality aversion and our inability to pin down an individual's subjective probability that she will join a coalition when others in her group have heterogeneous preferences.

11

We do know, however, that an individual with standard preferences in a population that includes inequality-averse individuals will likely vote to implement the efficient coalition as the minimum membership requirement and would never vote for a larger membership requirement. To see this, we can use [13] to show that the minimum profitable coalition for an individual with standard preferences is $\tilde{s}_i = s_{\min} \leq \bar{s}$. This implies that [17] is the relevant marginal expected payoff function with $\alpha_i = \beta_i = 0$. Thus, $v_i'(s) = b - p_i'(s)c$ for $s \in [s_{\min}, \bar{s})$, and $v_i'(s) = -p_i'(s)c$ for $s \geq \bar{s}$. If $p_i'(s)$ is not too large (e.g., not too much larger than $1/n$), then $v_i'(s) = b - p_i'(s)c > 0$, indicating that the individual would vote to implement $\bar{s}$ as the membership requirement. However, $v_i'(s) < 0$ for larger coalitions. Thus, an individual with standard preferences would never vote for an inefficiently large coalition as the group's membership requirement.

In contrast, inequality-averse individuals might vote to implement membership requirements that exceed the efficient coalition size. Depending on the voting rule and the distribution of preferences in a group, inequality aversion can lead to the implementation of inefficiently large coalitions. Suppose, for example, that all individuals are inequality averse and that their expected payoffs are higher for coalitions above $\bar{s}$. We would expect this group to implement a minimum participation rule above $\bar{s}$ and that such a coalition will form in the second stage of the game.

It is also possible that inequality aversion can prevent coalitions from forming, because it creates tension in the desired size of coalitions between those with standard preferences who prefer efficient coalitions and those who are inequality averse who prefer larger coalitions. There is a special case of this result that turns out to be important in our experiments. We find a small but significant number of situations in which a player voted to implement the efficient coalition size as the membership requirement in the first stage, this requirement was implemented, but the individual then blocked the coalition in the coalition formation stage. These seemingly contradictory actions are actually consistent with our model. Certainly an individual with standard preferences would never vote to implement the efficient coalition as the membership requirement and then block it, because efficient coalitions are profitable for individuals with standard preferences. However, these actions might be rational for an inequality-averse individual. Consider such an individual with a minimum profitable coalition above $\bar{s}$. This

individual will never join a coalition with $\bar{s}$ members, but she might vote to implement $\bar{s}$ as the membership requirement. To see why, suppose that her assessment of the likelihood of joining a profitable coalition, along with her preferences for equity and the fundamental parameters of the game result in [18] being negative. This person would not vote for a membership requirement that exceeds her minimum profitable coalition size. She will instead vote for a smaller membership requirement, but will refuse to join a coalition of this size because it is not profitable for her. Which coalition will she vote for? In this case note that her expected payoff [16] reduces to $v_i(s) = u_i^{nm}(s)$. From [12], this is increasing in s up to $\bar{s}$ and then is decreasing for $s > \bar{s}$. Therefore, the individual will vote to implement $\bar{s}$ as the membership requirement, but of course, she would block this coalition if she ever became critical for its formation. This occurrence turns out to be significant in our experiments.

In summary, our coalition formation model with individual with standard preferences predicts that players will implement efficient coalitions, regardless of whether efficiency requires full or partial participation. The presence of inequality-averse individuals makes additional equilibria possible. In particular, inequality-averse individuals can cause inefficiently large coalitions to form and can cause efficient coalitions to fail. With theory as a guide, the experiments discussed next explore the endogenous formation of coalitions to provide public goods.

3. Experiments
Following our theoretical models, our experiments have two stages, a voting stage and a coalition formation stage. The context of the experiment is fairly generic: subjects decide whether to join a coalition (called an agreement) with its members making a discrete choice to contribute to the public good (called a public account). We implement four experimental treatments that differ according to the returns to individual contributions to the public good. In two treatments, the marginal return to contributions is constant so that the efficient coalition is the grand coalition. In the other two treatments, the marginal return to public good contributions is constant up until an aggregate level and then zero so that the efficient coalition is smaller than the grand coalition. Each pair of treatments includes one in which players have homogenous payoffs and another in which payoffs are heterogeneous. We parameterized the treatments so that the equilibrium predictions given standard preferences were the same under both homogeneous

and heterogeneous payoffs. Although our theoretical treatment of the coalition formation problem did not include heterogeneous payoffs, we included these treatment treatments to test whether this feature affected coalition formation behavior.

We chose parameter values of $n = 6$, $A = 10$, $c = 10$ for all four treatments. In one of the treatments in which all subjects had identical payoff functions, the marginal return to all individual contributions was set at $b = 4.5$. Since $nb - c > 0$ for all contributions, the efficient coalition in this case is six individuals ($\bar{s} = 6$). In the other treatment with identical payoffs, $b = 4.5$ for contributions up to three units and then zero after three units. The efficient coalition size in this case is three individuals ($\bar{s} = 3$). In one of the treatments in which players had heterogeneous payoffs, three individuals had $b_h = 5$ (high earners) while the other three had $b_l = 4$ (low earners) for all contributions. The grand coalition is efficient in this treatment. In the other heterogeneous payoff treatment, individual marginal returns were zero after three units were contributed. The efficient coalition in this case contains three members of any combination of high and low earners. The payoff parameters are such that all three-player coalitions in all treatments are profitable for individuals with standard preferences.

All sessions were run at the experimental economics laboratory at [insert institution here] using software specifically designed for this experiment. In each session, three groups of six subjects were in the lab. These groups of six remained constant throughout the experiment (i.e., partners design) which lasted 20 periods. We ran one session for each of the four treatments and therefore we have 72 subjects who generated 1,440 individual-level observations. These observations include their votes for the membership requirement in the first stage and their decision to join a coalition in the second stage.

In the first stage of the experiments, subjects voted on the minimum membership requirement for a coalition to form in the second stage. They were given 60 seconds to vote by selecting a number, in the range of one to six, from a listbox. Subjects could change their votes during the voting period, but the number selected in the subject's listbox at the end of the 60 seconds, was recorded as their vote. A unique feature of our voting protocol was that a subject's selection from the listbox was viewed by the other five group members. Although only the subject's final selection was recorded as their vote, this feature allowed for a type of structured group communication in which subjects exchanged information about their intended votes. This communication feature better represents the process of determining minimum participation

14

requirements for international agreements compared to simultaneous voting without the ability to communicate. A modal response voting rule was used. The membership requirement that received the most votes was implemented in the second stage of the game. Ties were settled by a random draw.

In the coalition formation stage, each subject decided whether or not to join the coalition. Again, players are given 60 seconds to make their decision. Once a subject made their choice they could not change it. During the stage, all subjects were given real-time information about the decisions made by the other five members in the group. Specifically, they were informed about the number of other subjects that had joined the coalition, the number of others that decided not to join, and the number of others that had not yet made a decision. Therefore, each player knew whether their participation decision was critical for the coalition to form.[6] If someone failed to make a decision within the allotted time, she was made not to join the coalition by default.[7]

If enough players joined the coalition to satisfy the minimum membership requirement, then the coalition 'formed' and those that joined contributed to the public good. Those who did not join did not contribute. If too few subjects joined the coalition to satisfy the membership requirement, then no coalition formed and no one contributed to the public good. This stage of the experiment is very similar to traditional threshold public goods experiments that utilize

---

[6] The experiments deviate slightly from the theory in the coalition stage in that subjects make their participation decisions sequentially and with complete information. This deviation from theory, however, does not change the size of the coalition in a sub-game perfect Nash equilibrium in comparison to the simultaneous move game. The equilibrium coalition is s_p provided that this coalition size is profitable for enough members. The sequential decision making feature under complete information is useful in our experiments because it limits the extent of coordination problems since each player knows whether they are critical for a coalition to form.

[7] The software was designed with a number of features to ensure that subjects had complete information regarding the choices made by the other group members when making their choice. For example, if more than one subject made a decision within the same second, only the first decision was recorded. In those situations the subjects whose decision did not register received a message informing them that their action was not recorded and instructed them that the group's information had changed. Those subjects could then reevaluate their position and were given the opportunity to make another decision. In addition, if a group member made a decision within the last five seconds of the round when some subjects remained undecided, an additional five seconds were added to the time remaining. This feature provided undecided subjects enough time to assimilate changes before making their decision.

money-back guarantees (Dawes et al. 1986; Erev and Rapoport 1990; Bagnoli and McKee 1991; Cadsby and Maynes 1999).

Our theoretical development suggests two sets of predictions, one for situations in which all players in a group have standard preferences and another when some individuals in a group are inequality-averse. If all individuals have standard preferences, then the theory suggests that subjects will vote to make the efficient size coalition the minimum membership requirement in the first stage, and this coalition will form in the second. This is true whether the efficient coalition is the grand coalition or the smaller coalition of size three. Given our chosen parameters, this prediction holds under homogeneous and heterogeneous payoffs.

However, if some subjects are inequality averse, many outcomes are possible, particularly when the efficient coalition is smaller than the grand coalition; that is, when inequality itself is efficient. In the experiments in which the grand coalition is efficient, we expect that this coalition will form even if some subjects are inequality averse, since there is no inequality when the grand coalition forms. If inequality aversion is important, we expect its effects to show up in the experiments for which 3-player coalitions are efficient. In these cases, we've shown that inefficiently large coalitions may form and that efficient coalitions may fail to form. Thus, under the treatments in which a 3-person coalition is efficient, we may observe a significant number of larger coalitions forming. Moreover, if the 3-person membership requirement is ever implemented, we may observe that it will be blocked a significant number of times.

4. Results

Our experimental data suggest the following broad conclusions. In the treatments for which the grand coalition was efficient (the $\bar{s} = 6$ treatments), individuals voted overwhelmingly to implement the 6-player membership requirement. The grand coalition formed in a significant majority of trials, leading to high efficiency as measured by the ratio of group earnings to maximum group earnings. Performance was significantly worse in the treatments for which a 3-player coalition was efficient (the $\bar{s} = 3$ treatments). Coalitions formed in only about half of the trials and only half of these were efficient. Consequently, average efficiency was significantly lower in these treatments than in the $\bar{s} = 6$ treatments. These results, and the individual voting decisions and decisions to join coalitions that we now look at in detail, are not consistent with a

16

theory of coalition formation that includes only individuals with standard preferences. They are, however, consistent with a theory that allows for the presence of inequality-averse players.

We first examine the data on voting for the membership requirement in the first stage of the experiments. Table 1 provides votes and referenda outcomes by membership requirement and treatment. The first row in each cell contains the number of votes and percentage of total votes (out of 360 for each treatment) for that minimum membership requirement. The second row in each cell contains the number of times and percentage of trails in the treatment (out of 60) that membership requirement was implemented. Kolmogorov-Smirnov (K-S) tests of the null hypothesis that the samples of votes under homogenous and heterogeneous payoff treatments are drawn from the same distribution are rejected but only at the 10 percent level of significance ($p = 0.097$ for both tests under $\bar{s} = 3$ and $\bar{s} = 6$). K-S tests of the null hypotheses that the samples of adopted membership requirements under homogenous and heterogeneous payoff treatments are drawn from the same distribution cannot be rejected ($p = 0.809$ under $\bar{s} = 3$; 0.925 under $\bar{s} = 6$). Since the patterns of votes and membership requirements for the homogeneous and heterogeneous treatments are so similar, we discuss only the combined results for the $\bar{s} = 3$ and $\bar{s} = 6$ treatments.

Table 1: Individual votes and referenda outcomes by minimum membership requirement and treatment

	Minimum membership requirement					
	1	2	3	4	5	6
$\bar{s} = 6$						
Homogeneous	6 (1.7%)	7 (1.9%)	17 (4.7%)	26 (7.2%)	72 (20.0%)	232 (64.4%)
	1 (1.7%)	1 (1.7%)	2 (3.3%)	4 (6.7%)	6 (10.0%)	46 (76.7%)
Heterogeneous	19 (5.3%)	18 (5.0%)	21 (5.8%)	31 (8.6%)	63 (17.5%)	208 (57.8%)
	1 (1.7%)	3 (5.0%)	4 (6.7%)	4 (6.7%)	8 (13.3%)	40 (66.7%)
Combined	25 (3.5%)	25 (3.5%)	38 (5.3%)	57 (7.9%)	135 (18.8%)	440 (61.1%)
	2 (1.7%)	4 (3.3%)	6 (5.0%)	8 (6.7%)	14 (11.7%)	86 (71.7%)
$\bar{s} = 3$						
Homogeneous	25 (6.9%)	39 (10.8%)	160 (44.4%)	22 (6.1%)	33 (9.2%)	81 (22.5%)
	2 (3.3%)	6 (10.0%)	33 (55.0%)	4 (6.7%)	3 (5.0%)	12 (20.0%)
Heterogeneous	16 (4.4%)	40 (11.1%)	150 (41.7%)	21 (5.8%)	19 (5.3%)	114 (31.7%)
	2 (3.3%)	7 (11.7%)	26 (43.3%)	3 (5.0%)	3 (5.0%)	19 (31.7%)
Combined	41 (5.7%)	79 (11.0%)	310 (43.1%)	43 (6.0%)	52 (7.2%)	195 (27.1%)
	4 (3.3%)	13 (10.9%)	59 (49.2%)	7 (5.8%)	6 (5.0%)	31 (25.8%)

Top of each cell: Number of votes for each minimum membership requirement (percent of total votes by treatment). Note there are 360 individual votes per treatment. **Bottom of each cell**: Number of times each minimum membership requirement was implemented (percent of total trials by treatment). Note there are 60 group-level observations per treatment.

Under the $\bar{s} = 6$ treatments, the 6-player membership requirement received 61.1 percent (440 of 720) of total votes, which is considerably more than the 18.8 percent received by the second most preferred option of a 5-player membership requirement. The remaining four options received even fewer votes. This voting behavior resulted in the selection of the efficient grand coalition in 71.7 percent (86 of 120) of referenda. Voting behavior is generally consistent with our predictions for the $\bar{s} = 6$ treatments with and without the presence of inequality-averse players.

Voting under the $\bar{s} = 3$ treatments was more complicated. Even though the 3-player membership requirement received the greatest number of votes (43.1%), this is significantly less than the percentage of votes to implement the efficient coalition as the membership requirement under $\bar{s} = 6$ treatments (61.1%, $p = 0.000$).[8] Moreover, the subjects under the $\bar{s} = 3$ treatments

[8] We report unconditional summary statistics in our tables. However, we recognize that our observations are not entirely independent as the same subject makes repeated decisions as part of

showed a considerable tendency to vote to implement higher membership requirements, in particular the 6-player membership requirement (27.1% of the votes). Membership requirements mirror the votes, with the 3-player requirement being implemented in 49.2 percent of the trials and the 6-player requirement implemented in 25.8 percent of the trials. Therefore groups under the $\bar{s} = 3$ treatments implemented the efficient coalition size as the membership requirement considerably more often than the 6-player requirement. However, we will see shortly that this difference did not result in the efficient coalition forming significantly more often than the 6-player coalition.

Our theoretical model with standard preferences predicts that groups will always implement the efficient coalition size as the minimum membership requirement. This hypothesis is clearly violated in our $\bar{s} = 3$ treatments. However, implementing larger membership requirements is consistent with the presence of inequality-averse subjects. While the 3-player coalition is efficient, this coalition size requires a subset of players to not join the coalition and free ride off contributions of the members. By requiring that everyone join a coalition for it to form, groups can eliminate the freeriding. The support for 3-player and 6-player membership requirements suggests a fundamental tension between efficiency and eliminating inequality in our experiments.

Table 2 contains results concerning coalition formation by minimum membership requirement and treatment. For each membership requirement/treatment combination we provide the number of times a coalition formed under the membership requirement, this number as a percentage of total trials, and coalition formations as a percentage of times the membership requirement was adopted. In the final column in Table 2 we present the number and percentage of trials a coalition of any size formed (i.e., all effective coalitions). A z-test comparing the proportion of coalition formation between homogeneous and heterogeneous treatments suggests the two are equivalent ($p = 0.583$ under $\bar{s} = 3$; $p = 0.307$ under $\bar{s} = 6$) and therefore we examine the combined data.

---

a stable group. To address this issue our hypotheses tests use linear regression techniques to control for period, subject (for individual-level data) and group (for group-level data) fixed effects. We report these p-values throughout.

Table 2: Coalition formation by minimum membership requirement and treatment

	Minimum membership requirement						
	1	2	3	4	5	6	Total
$\bar{s} = 6$							
Homogeneous	1	0	2	1	4	45	53
	1.7%	0.0%	3.3%	1.7%	6.7%	75.0%	88.3%
	100.0%	0.0%	100.0%	25.0%	66.7%	97.8%	
Heterogeneous	1	0	2	2	7	37	49
	1.7%	0.0%	3.3%	3.3%	11.7%	61.7%	81.7%
	100.0%	0.0%	50.0%	50.0%	87.5%	92.5%	
Combined	2	0	4	3	11	82	102
	1.7%	0.0%	3.3%	2.5%	9.2%	68.3%	85.0%
	100.0%	0.0%	66.7%	37.5%	78.6%	95.4%	
$\bar{s} = 3$							
Homogeneous	1	1	15	1	0	12	60
	1.7%	1.7%	25.0%	1.7%	0.0%	20.0%	50.0%
	50.0%	16.7%	45.5%	25.0%	0.0%	100.0%	
Heterogeneous	0	3	13	0	1	10	27
	0.0%	5.0%	21.7%	0.0%	1.7%	16.7%	45.5%
	0.0%	42.9%	50.0%	0.0%	33.3%	52.6%	
Combined	1	4	28	1	1	22	57
	0.8%	3.3%	23.3%	0.8%	0.8%	18.3%	47.5%
	25.0%	30.8%	47.5%	14.3%	16.7%	71.0%	

Top of each cell: Number of times coalitions formed. **Middle of each cell**: Percentage coalition formation by number of trials per treatment. **Bottom of each cell**: Percentage coalition formation by adopted membership requirement.

Under the $\bar{s} = 6$ treatments, note that coalitions of any size formed in 102 of 120 (85%) trials. Eighty-two of these coalitions (80.4%) were efficient coalitions. Thus, coalitions formed quite frequently in the $\bar{s} = 6$ treatments and the greatest majority of these were efficient. Overall, the efficient coalition formed in 82 out of 120 trials (68.3%). Other coalitions formed far less frequently. When groups implemented the 6-player membership requirement, the coalition formed 95.4 percent of the time. This suggests that the main reason that smaller than efficient

coalitions formed in a minority of the trials is because groups sometimes failed to implement the 6-player membership requirement.

In contrast, coalitions formed far less frequently under the $\bar{s} = 3$ treatments and only about half of these were efficient. Coalitions formed in 47.5 percent of trials under these treatments, which is significantly lower than the 85.0 percent coalition formation rate under the $\bar{s} = 6$ treatments ($p = 0.000$). Of the 57 coalitions that formed only 28 of these were efficient. The rate at which efficient coalitions formed under the $\bar{s} = 3$ treatments (23.3%) is far lower than the rate at which the efficient coalition formed under the $\bar{s} = 6$ treatment (68.3%, $p = 0.000$). One reason for the low rate of efficient coalition formation is the low rate at which the 3-player membership requirement was implemented (49.2% of trials from Table 1). Recall that the 6-player membership requirement was implemented in a significant number of trials. In fact, the grand coalition formed in 18.3 percent of all trials in the $\bar{s} = 3$ treatments, while the efficient coalition formed in 23.3 percent of trials.

Another reason the efficient coalition failed to form in the $\bar{s} = 3$ treatments is that it formed only 47.5 percent of the time when the 3-player membership requirement was adopted. Let us explore this phenomenon more closely, since it implies that some individuals willfully blocked efficient coalitions. Recall that an individual blocks a coalition if her willingness to join is critical for its formation, but she refuses to join so the coalition does not form. The 3-player membership requirement was implemented in 59 out of 120 trials (49.2% from Table 1) in the combined homogenous and heterogeneous payoff treatments under $\bar{s} = 3$. The efficient coalition failed to form in 31 of these trials. Of these 31 failures, 29 were blocked by individuals who voted for a membership requirement of three or more players. Of these 29, 11 were due to blocks by individuals who voted for membership requirements with more than three players, while 18 were due to blocks by individuals who actually voted to implement the 3-player membership requirement. All of these blocks would be inconsistent with a coalition formation model with only individuals with standard preferences: these individuals would never block an efficient coalition because they find them profitable. These blocks are, however, consistent with the presence of inequality-averse subjects. It is easy to understand why an inequality-averse subject whose minimum profitable coalition is larger than the efficient coalition would vote for a higher membership requirement and also refuse to join an efficient coalition. It is more surprising that some individuals who voted for the 3-player membership requirement blocked

this coalition from forming. However, we showed in section 2 why this could occur. Such an individual envisions little chance that a coalition they find profitable would form; they are better off if the efficient coalition forms than if no coalition forms so they vote to implement this membership requirement, but they are better off if no coalition forms than joining the efficient coalition so they refuse to join if given the opportunity. We find it interesting that this behavior plays a significant role in preventing efficient coalitions from forming.

We complete our data analysis with results on average public good provision and average efficiency in Table 3. Efficiency for each group in each period is calculated as the ratio of aggregate payoffs to maximum payoffs. As expected, public good provision was lower in the $\bar{s} = 3$ treatments than in the $\bar{s} = 6$ treatments. More importantly, the inability of subjects under the $\bar{s} = 3$ treatments to form efficient coalitions consistently produced significantly lower efficiency as compared to the $\bar{s} = 6$ treatments. For the homogenous and heterogeneous payoff treatments combined, subjects earned 87.5 percent of maximum earnings under the $\bar{s} = 6$ treatments, while subjects earned significantly less, 68.7 percent, under the $\bar{s} = 3$ treatments ($p = 0.000$). On average, public good provision and efficiency were less under the heterogeneous treatment than under the homogenous treatments, but these differences are not significant (for public good provision, $p = 0.213$ and $p = 0.585$; for efficiency, $p = 0.194$ and $p = 0.428$ for $\bar{s} = 6$ and $\bar{s} = 3$ respectively).

Table 3. Public good provision and efficiency

Treatment	Average Public Good Provision	Efficiency
$\bar{s} = 6$		
Homogeneous	5.05	90.25%
	(0.26)	(2.74)
Heterogeneous	4.55	84.78%
	(0.30)	(3.15)
Combined	4.8	87.5%
	(0.20)	(2.09)
$\bar{s} = 3$		
Homogeneous	2.07	70.12%
	(0.31)	(2.57)
Heterogeneous	1.83	67.34%
	(0.30)	(2.37)
Combined	1.95	68.7%
	(0.21)	(1.74)

Standard errors are in parentheses. Each treatment consists of 60 group-level observations.

5. Conclusion

We have analyzed, both theoretically and experimentally, endogenous coalition formation in which players determine a minimum participation requirement before deciding whether they will join the coalition. We demonstrate theoretically that if all players only care about maximizing their expected financial payoffs, they will vote to implement an efficient coalition size as the membership requirement and this coalition will form. However, we also demonstrate that if some players are averse to inequality they can move groups to inefficient outcomes, particularly if an efficient coalition is smaller than the grand coalition. Inequality-averse individuals can cause larger-than-efficient coalitions to form, and they may block the formation of efficient coalitions.

In general our experimental results are consistent with a model of coalition formation in which players have an aversion to inequality. In treatments for which the efficient coalition was the grand coalition, coalitions formed 85 percent of the time and most of these were efficient. While this result is predicted by a model of players having standard preferences, adopting the grand coalition is also consistent with inequality aversion since this coalition size minimizes inequality. On the other hand, in treatments for which the efficient coalition required only a

23

subset of the group, coalitions formed in just over half the trials, but only half of these were efficient. In fact, the grand coalition formed at about the same rate as the efficient coalition. We also find that efficient coalitions that allowed for free riding were blocked about half the time. Although individuals with standard payoff-maximizing preferences would never adopt larger-than-efficient coalitions or block efficient coalitions from forming, these actions are consistent with inequality-averse individuals.

Our results also demonstrate the value of the endogenous implementation of a minimum membership requirement in promoting coalition formation and public good provision when the efficient coalition is the grand coalition. When efficiency required full participation in our experiments, coalitions were 87.5 percent efficient on average. This efficiency level is high compared to other experiments on coalition formation with similar incentives as ours but without minimum membership requirements. For example, Kosfeld et al. (2009), Dannenberg et al. (2010) and McEvoy et al. (2010) report average efficiency measures of 60.5, 22 and 57.9 percent respectively.

The research also provides insight into the role preferences for equity play in the design of effective institutions that govern international public goods. One of the striking results from this study is the high frequency of trials in which groups adopted the grand coalition even though efficiency required some degree of free riding. This finding may help explain the choice of participation requirements in many existing voluntary institutions. Specifically, the fact that many international environmental agreements require full, or very high levels of participation (Barrett 2003). In light of our results it is possible that some of these existing agreements have participation thresholds that are inefficiently high in order to limit the extent of free-rider payoffs.

Many economists now appreciate the role that equity and fairness play in the design of effective institutions to resolve social dilemma situations. The typical result in the literature is that inequality aversion can help foster cooperation between group members in public goods and common-pool resource games and lead to more-efficient outcomes. Our study contributes to this growing theoretical and experimental literature on inequality. With the endogenous institution we consider, we show that inequality aversion can actually move groups away from efficient outcomes. The complete picture shows that inequality aversion can either foster or frustrate

cooperation among group members, and which prevails likely depends on whether resolving the social dilemma mitigates or exacerbates inequality measures.

References

Barrett, Scott. 2003. *Environment and Statecraft: The Strategy of Environmental Treaty Making*. Oxford, U.K.: Oxford University Press.

Bagnoli, Mark, and Michael McKee. 1991. "Voluntary Contribution Games: Efficient Private Provision of Public Goods." *Economic Inquiry*, 29(2): 351-66.

Bolton, Gary E., and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90(1): 166–93.

Burger, Nicholas E., and Charles D. Kolstad. 2010. "International Environmental Agreements: Theory Meets Experimental Evidence."
http://fiesta.bren.ucsb.edu/~kolstad/HmPg/papers/IEA%20Experiments.9.2010.pdf

Cadsby, Charles B., and Elizabeth Maynes. 1999. "Voluntary Provision of Threshold Public Goods with Continuous Contributions: Experimental Evidence." *Journal of Public Economics*, 71(1): 53-73.

Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.

Carraro, Carlo, Carmen Marchiori, and Sonia Oreffice. 2009. "Endogenous Minimum Participation in International Environmental Treaties." *Environmental and Resource Economics*, 42(3): 411-25.

Charness, Gary and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117(3): 817-69.

Cox, James C., Daniel Friedman, and Steven Gjerstad. 2007. "A Tractable Model of Reciprocity and Fairness." *Games and Economic Behavior*, 59(1): 17–45.

Dannenberg, Astrid and Andreas Lange. 2010. "On the Formation of Coalitions to Provide Public Goods – Experimental Evidence from the Lab." *ZEW – Centre for European Economic Research Discussion* Paper No. 10-037.

Dawes, Robyn M., John M. Orbell, Randy T. Simmons, and Alphons J. C. Van De Kragt. 1986. "Organizing Groups for Collective Action." *The American Political Science Review*, 80(4): 1171-85.

Engelmann, Dirk, and Martin Strobel. 2004. "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments." *American Economic Review*, 94(4): 857-69.

Erev, Ido, and Amnon Rapoport. 1990. "Provision of Step-Level Public Goods: The Sequential Contribution Mechanism." *Journal of Conflict Resolution*, 34(3): 401-25.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3): 817-68.

Fehr, Ernst, and Simon Gachter. 2000a. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90(4): 980-94.

Fehr, Ernst, and Simon Gachter. 2000b. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic* Perspectives, 14(3): 159–81.

Fehr, Ernst, Michael Naef, and Klaus M. Schmidt. 2006. "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment." *American Economic Review*, 96(5): 1912-17.

Gurerk, Ozgur, Bernd Irlenbusch, and Bettina Rockenbach. 2006. "The Competitive Advantage of Sanctioning Institutions." *Science*, 312(5770): 108-11.

Kosfeld, Michael, Akira Okada, and Arno Riedl. 2009. "Institution Formation in Public Goods Games." *American Economic Review*, 99(4):1335-55.

Kroll, Stephan, Todd L. Cherry, and Jason F. Shogren. "Voting, Punishment, and Public Goods." *Economic Inquiry*, 45(3): 557-70.

McEvoy, David, James Murphy, John Spraggon, and John Stranlund. 2010. "The Problem of Maintaining Compliance with Stable Coalitions: Experimental Evidence." *Oxford Economic Papers*, published online October 1, 2010 doi:10.1093/oep/gpq023

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83(5): 1281–1302.

Sutter, Matthias, Stefan Haigner and Martin G. Kocher. 2010. "Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations." *Review of Economic Studies*, 77(4): 1540-66.

Tyran, Jean-Robert, and Lars P. Feld. 2006. "Achieving Compliance when Legal Sanctions are Non-deterrent." *Scandinavian Journal of Economics*, 108(1): 135-56.

Walker, J.M., Roy Gardner, Andrew Herr and Elinor Ostrom. 2000. "Collective Choice in the Commons: Experimental Results on Proposed Allocation Rules and Votes." *The Economic Journal*, 110(460): 212-34.