Public Goods Agreements with Other-Regarding Preferences

Charles D. Kolstad*

DRAFT: March 25, 2011

Abstract

Stimulation of cooperation when noncooperation appears to be individually rational has been an issue in economics for at least a half century. In the 1960's and 1970's the context was cooperation in the prisoner's dilemma game; in the 1980's voluntary provision of public goods; in the 1990's, the literature on coalition formation for public goods provision emerged, in the context of coalitions to provide transboundary pollution abatement. The problem is that theory suggests fairly low (even zero) levels of contributions to the public good. Experiments and empirical evidence suggests higher levels of cooperation. This is a major reason for the emergence in the 1990's and more recently of the literature on other-regarding preferences (also known as social preferences). Such preferences tend to expand cooperation (though not always). This paper contributes to the literature on coalitions, public good provision and other-regarding preferences. For standard preferences, the marginal per capita return (MPCR) to investing in the public good must be greater than one for contributing to be individually rational. We find that Charness-Rabin preferences tend to reduce this threshold for individual contributions. We also find that Charness-Rabin preferences reduce the equilibrium size of a coalition of agents formed to provide the public good. In addition to theoretical results, some experimental implications of the theoretical model are provided. In contrast to much of the literature, we treat the wealth of agents as heterogeneous.

---

*Department of Economics and Bren School, University of California, Santa Barbara, Resources for the Future and NBER. Research assistance from Trevor O'Grady has been greatly appreciated, as have been conversations with Gary Charness.

I. INTRODUCTION

Stimulation of cooperation when noncooperation appears to be individually rational has been an issue in economic theory for at least a half century. In the 1960's and 1970's the context was cooperation in the prisoner's dilemma game (eg, Schelling, 1973). In the 1980's, voluntary provision of public goods became the focus (eg, Bergstrom et al, 1986). In the 1990's, the literature on international environmental agreements (IEAs) began to develop (eg, Barrett, 1994 and Carraro and Siniscalco, 1993). Within economics, IEAs are typically viewed simply as coalitions of agents providing public goods (abatement) and thus properly belong in this literature.

A parallel empirical literature (primarily experimental) began to emerge in the early 1980's calling into question the theoretical results, primarily the dismal theoretical findings that free riding was common and cooperation was difficult to sustain. Kim and Walker (1984) were one of the first to offer evidence from laboratory experiments suggesting people were far less likely to free ride and more likely to cooperate than theory suggests. Other games of cooperation and competition (other than the public goods game) were also fraught with disparities between laboratory behavior and experiments.

In response to this disparity between empirics and theory, the notion of other-regarding or social preferences began to emerge in the early 1990's. Andreoni (1990) suggested that agents receive utility from public goods through an additional route. He suggested that agents receive utility from the act of contributing to the public good – something he terms "warm glow." This perspective that simple utility was not enough to explain cooperation led to a burgeoning literature on social preferences (eg, Fehr and Schmidt, 1999; Charness and Rabin, 2002). A primary thrust of this literature is that agents care about three things: their own private payoff, fairness in payoffs (within the population), and overall efficiency (the aggregate economic surplus accruing to all agents). "Standard" preferences would only involve private payoffs.

Several important questions remain unanswered. How should the theory of voluntary provision of public goods be modified when agents have a specific form of social preferences? How should the theory of coalitions to provide public goods (as in the IEA literature) be modified when agents have social preferences?

This paper is one of the first papers to address these unanswered questions. We posit a particularly common form of social preferences, due to Charness and Rabin (2002). We then develop (1) a theory of voluntary contributions to public goods for the linear public goods game and (2) a theory of voluntary coalitions to provide public goods. Unlike many other papers, we allow income to vary over the population, which yields significantly richer results. Using experimental results from Kosfeld et al (2009), we estimate the parameters of a social preference function and show the implications for cooperation.

A number of results emerge from this paper. In terms of theory, we find that social preferences do expand the set of MPCRs for which contributing to the public good. This is confirmed in our experimental results. In the experimental results, an MPCR of 0.65 is sufficient for the large majority of subjects to voluntarily contribute to the public good. Even when the MPCR is as low as 0.4, 10% of agents should find it individually rational to contribute.

In terms of coalitions to provide public goods, we find that the introduction of social preferences tends to reduce the equilibrium size of stable coalitions. Furthermore, we find that the distribution of income has a significant impact on the theoretical stability of coalitions. Coalitions in which the endowment of agents is quite similar are more likely to be stable than coalitions in which there are wide disparities in endowment. This result is particularly significant for international environmental agreements.

II. BACKGROUND

A. Private Provision of Public Goods

Inducing individual contributions to a public good in a noncooperative setting is a classic problem in public economics. Bergstrom et al (1986) provide the standard treatment of this problem, developing a simple model involving individual provision of a private good, x_i, the public good, g_i, and aggregate provision of the public good, G (=Σg_i). Each identical agent (i) has simple preferences and an endowment of wealth, w_i, to be divided between x_i and g_i. The individual chooses x and go to maximize utility, subject to a budget constraint:

$$u(x_i,G) \quad \text{s.t. } x_i + g_i = w_i \tag{1a}$$

where the first argument of u embodies the opportunity cost to the individual of providing the public good and the second term reflects the benefit of the aggregate provision. The authors show that in most cases there is a nonzero equilibrium provision of the public good. A second interesting result involves identical preferences but different wealth levels. In this case, there is a cutoff level of wealth. People who are poorer than the cutoff provide none of the public good whereas people above the cutoff provide a nonzero amount.

Andreoni (1988) uses this model to determine how contributions increase as the size of the economy (N —the number of individuals) increases. He shows that as N increases towards infinity, average individual contributions approach zero, the size of the contributing group approaches zero and the aggregate contributions approach a limit which is finite and nonzero. He points out that this is at variance with casual empiricism that individuals do contribute to public goods, despite the economy being very large. For instance, according to Andreoni, half of all US households claim charitable donations on their tax returns (in the US, charitable donations are generally deductible from taxable income).

Early experimental work on public good provision established that subjects tend to provide public goods at higher rates than predicted by the theory described above. Kim and Walker (1984) set up a laboratory experiment to test the "free rider hypothesis," which had been the subject of a number of papers in the 1970s (in the context of the prisoner's dilemma). The hypothesis simply is that individuals will prefer to free-ride rather than make contributions to the public good. The authors distinguish between the "strong" free riders and other free riders. Strong free riders are closer to the theoretical behavior of contributing little to the public good. The authors show that although free-riding exists, they are not able to conclude that the free-riding is as strong as theory suggests. Isaac and Walker (1988) provide additional experimental evidence, exploring the role of an important variable, the *marginal per capita return* (MPCR). The MPCR is defined as the ratio of the marginal benefit to an individual of privately providing a public good to the marginal cost of that provision by the individual. Put differently, for every dollar a person spends on privately providing the public good, the MPCR measures how much the individual gets back. Clearly the MPCR is less than one (otherwise there is no issue). Higher MPCRs mean that the private gain from the public good is higher. A lower MPCR means that the individual is getting less private reward from providing the public good. Isaac and Walker (1988) demonstrate experimentally that MCPR is the primary determinant of contribution levels—there is no separate pure group size effect.[1] Furthermore, the authors demonstrate that the strong free rider effect is more pronounced for lower values of MPCR.

In an interesting review of this literature, Chaudhuri (2010) characterizes five main findings of the pre-1995 literature (attributing the last three to Ledyard, 1995): (1) in one shot versions of the noncooperative public goods game (described above) there is much less free-riding (more contribution) than predicted by theory; (2) if players repeat the one-shot game, free-riding increases with repeated interaction; (3) communication facilitates cooperation; (4) thresholds facilitate cooperation; and (5) higher MPCRs lead to increased cooperation and decreased free-riding.

Over the past decade or two, researchers have been moving beyond simple characterizations of payoffs to include a variety of "other-regarding preferences" or "social preferences" on the part of participants. One of the first extensions of this nature is the model of "impure altruism" by Andreoni (1989,1990), building in part on suggestions decades earlier by Mancur Olson and Gary Becker on charitable giving. Impure altruism holds that there are two avenues for personal utility gain from making a voluntary contribution to a public good: via the aggregate level of the public good and via a "warm glow" associated with making a donation. Thus the individual may appear altruistic but that is because the individual obtains utility from giving. Thus the utility function in Eqn. (1a) becomes $u(x_i,G,g_i)$. It is easy to see that including a private good dimension to public good contributions can remedy the apparent anomalies between the experimental results on free-riding and the theoretical results on contributions to public goods.

[1] Isaac et al (1994) provide support for these findings using significantly larger groups.

Other authors provide alternative models of contributing to public goods, always with the issue of free-riding as a motivator. Drawing on the fairness literature in psychology and economics (eg, Kahneman et al, 1986), Fehr and Schmidt (1999) posit that inequality aversion drives cooperation. They propose the importance of inequality aversion as a dimension of utility that promotes cooperation and support the thesis with experimental evidence. Charness and Rabin (2002) present evidence in direct contradiction to this result, suggesting that efficiency also plays a role in outcomes in prisoner's dilemma games. In the Prisoner's Dilemma game shown in Fig. 1, theory would suggest defection repeatedly occurs. However, in an experimental setting, Charness et al (2008) find cooperation rates of 15%, 45% and 70% for values of x of 4, 5, and 6, respectively. This suggests more nuanced objectives. In particular, agents seem to be concerned with the total size of the "pie" as well as their own private payoffs.

Figure 1: Prisoner's Dilemma Payoffs from Charness et al (2008).

	B Cooperates	B Defects
A Cooperates	(x,x)	(1,7)
A Defects	(7,1)	(2,2)

Note: With payoff (a,b), a is payoff to player A and b is payoff to player B; $2<x<7$.

Bliss and Nalebuff (1984), in a paper with a superb title, examine the case where individuals have different "abilities" or costs to supply the public good. They show that even in a noncooperative, repeated setting, the lowest cost individuals will eventually take it upon themselves to supply the public good.

B. Coalitions for Public Good Provision

One way of enhancing the overall provision of public goods is through the formation of groups or coalitions of players to coordinate the provision of public goods. Over the past two decades, most of the work on this problem has been done in the context of the international environmental agreement (IEA). Only recently has the general public goods literature begun to address coalition formation (Kosfeld et al, 2009; Charness and Yang, 2011).

The literature on IEAs starts with a framework nearly identical to Eqn (1) for voluntary provision of public goods. The interesting twist added by the IEA literature (drawn from the cartel stability literature[2]) is that the noncooperative behavior is represented as a two stage game. In the first stage (the "membership game") countries decide whether they wish to be in a coalition (an IEA). Specifically,

[2] See d'Aspremont et al, 1983; Donsimoni et al, 1986

each country announces "in" or "out;" the first stage game generates a coalition as the Nash equilibrium in these announcements. In the second stage (the "emissions game"), the coalition acts as one and emissions choices of the coalition and fringe emerge as a Nash equilibrium in emissions conditional on the coalition formed in the first stage.

Barrett (1994) provides the first analysis of this problem in the literature. Unfortunately, he is unable to come up with analytic results without simplifying; he uses simulations to suggest that welfare gains from an IEA (relative to the noncooperative outcome) are modest. An IEA may have many members (relative to N), but in such cases , welfare gains are slight compared to the noncooperative equilibrium; conversely, when cooperation would increase net benefits significantly, the equilibrium size of an IEA is small. In other words, generally (but not always) there is an inverse relationship between the equilibrium number of coalition members and the gains from cooperation (ie, the welfare difference between a full cooperative outcome and a noncooperative outcome).[3]

One simplification of the model (Barrett, 1999; Ulph, 2004) is for payoffs to be linear with identical preferences and identical endowments and a common MPCR. In this case, it is easy to show that the equilibrium consists of fringe countries not abating ($g_i = 0$) and coalition members abating providing the coalition is large enough. Let n* be the smallest size of a coalition which chooses to abate. A basic result is that n* = 1/MPCR. One can also show that the benefits from cooperation increase in MPCR whereas the size of a coalition decreases in MPCR. The assumption of identical endowments is almost universal in this literature, despite the fact that the wealth of a country seems to be an important factor in driving participation in IEAs. Few in the IEA literature have explicitly treated social preferences. An exception is Lange (2006).

A number of authors focus on incentives to hold coalitions together, whether these be punishments or transfers among countries. Barrett (2002, 2003) examines credible punishments that can hold a coalition together, in the context of a repeated game. Allowing repeated interaction opens up the possibilities for a variety of outcomes, primarily because punishment strategies for defectors can be built into an agreement and then applied should a country defect.

One of the first papers to explore coalitions for providing public goods outside the context of international environmental agreements is Kosfeld et al (2009). In that paper, the authors suggest a stage game structure very similar to the standard static IEA problem, though with one additional stage. The first stage is a membership game, the second stage is an implementation stage and the third stage is a contribution stage. The membership and contribution stages are identical to the IEA problem. The participation stage involves members of the coalition deciding whether to implement the coalition. An implemented coalition involves payment of a fixed fee and punishment for not contributing enough

[3] This inverse relationship between MPCR and the equilibrium size of an IEA holds generally in Barrett (1994), as articulated in his Prop. 1. He does offer specific functional forms where it does not necessarily hold. For instance, he shows that with constant marginal damage and quadratic costs that the maximum size of an IEA is 3 members. In this special case the relationship between MPCR and IEA size will not of course hold. See also Finus (2003).

(applied to coalition members). Their results for standard preferences are a repetition of IEA results. An interesting extension is their introduction of fairness, using Fehr-Schmidt social preferences. They show that for a subset of Fehr-Schmidt preferences, the grand coalition is an organizing equilibrium.

Researchers have only recently begun to use experiments to validate theory on the formation of coalitions for public goods provision (Kosfeld et al, 2009; Burger and Kolstad, 2009; Dannenberg et al, 2010). Results are ambiguous though consistent with the private provision literature – experimental evidence suggests more cooperation and less free-riding.

III. THEORY: PRIVATE PROVISION

A. Basic Conditions.

Rather than postulate a general utility function which depends on individual and aggregate contributions (as in Bergstrom et al, 1986), we specify a linear (monetary) payoff to each agent and separately, the utility an individual gains from the distribution of monetary payoffs. We start with a standard homogeneous preferences linear public good game with respect to monetary payoffs. Specifically, there are i=1,…,N agents each with monetary payoffs. Eqn. (1a) can then be re-written as a problem of choosing g_i to yield payoff

$$\pi_i = w_i - g_i + aG \quad \text{where } G = \Sigma_i\, g_i \quad \text{and } 0 \le g_i \le w_i \qquad (1b)$$

Here w is wealth, g is the individual contribution to the public good and G is the aggregate contribution. This is equivalent to a linear payoff from a private good x_i and a public good G, with the agent making a contribution to a public good, g_i, subject to a budget constrain $x_i+g_i=w_i$. The variable a is the marginal per capita return (MPCR), indicating how much of an investment in the public good is returned privately. To keep the problem interesting, we restrict a to be in the open interval (1/N,1). Clearly a could vary from one agent to another, though that would complicate our analysis. We will assume wealth may be different from one individual agent to another in our group.

Assuming linear payoffs as in Eqn. (1b) inevitably leads to knife-edge outcomes wherein the agent either contributes all of his wealth or nothing to the public good. Nonlinear payoffs would make for more subtle behavior but it would be difficult to provide an analytic representation of the Nash equilibrium, particularly when other-regarding preferences are treated.

Agents act on the basis of their utility from payoffs, reflecting a possible concern for the distribution of payoffs within the game. There are a number of ways of representing these "other-regarding" preferences, including Andreoni's (1989, 1990) model of "impure altruism," discussed earlier. In this model agents receive utility (as opposed to monetary payoff) from their own monetary payoff (Eqn. 1c) as well as from their individual contribution g_i.

Fehr and Schmidt (1999) stipulate that utility depends on relative payoffs (for agent i):

$$u_i = \pi_i - \alpha_i/(N\text{-}1)\ \Sigma_{j\neq i}\max(\pi_j - \pi_i,0) - \beta_i/(N\text{-}1)\ \Sigma_{j\neq i}\max(\pi_i - \pi_j,0) \qquad (2)$$

where the α and β reflect aversion to personally disadvantageous (people doing better than I) and advantageous (people doing worse than i) inequality. The authors specifically state that even though agents may be homogenous in terms of the payoff function, some may have different attitudes towards inequality than others. Different mixes of "selfish" and "fair minded" people can result in very different levels of cooperation. We refer to Eqn. (2) as F-S social preferences.

Charness and Rabin (2002) suggest that efficiency is also important (see the discussion in the context of Figure 1). Although they are careful not to reject the Fehr and Schmidt representation, they suggest that it is incomplete. In fact, they suggest that utility depends on three dimensions of payoff: personal payoff, an equity term and an efficiency term. Their approach is to posit utility for agent i as a linear combination of own monetary payoff, the minimum monetary payoff over the rest of the population (a Rawlsian-like criterion reflecting a concern for equity) and total monetary payoffs over the population (reflecting concerns for social efficiency):

$$u_i = (1-\lambda_i)\ \pi_i + \lambda_i[\delta_i \min_{j\neq i} \pi_j + (1-\delta_i)\Sigma_j\ \pi_j] \qquad (3)$$

where $\lambda_i \varepsilon$ [0,1] and $\delta_i \varepsilon$ [0,1]. A λ_i=0 implies that own monetary payoffs are all that matter. A δ_i=0 with λ_i>0 implies that own payoffs and aggregate efficiency both matter (with λ_i as a weight) and that equity does not. We refer to Eqn. (3) as C-R preferences, though the equity term in Eqn. (3) is slightly different from the original representation in Charness and Rabin (2002) in that the minimum excludes own payoffs. Although this is in part for tractability, the fact is that a concern for equity is usually thought of as a concern for the well being of others, particularly those less well-off.

The Andreoni warm-glow model could be viewed as a variant of C-R when the warm glow arises from providing social benefits.

Because the C-R preferences appear to represent a broader perspective on social preferences (by including equity and efficiency, not just equity as in F-S), we adopt that representation here.

B. Efficient and Noncooperative Outcomes

Assume C-R preferences as characterized in Eqn. (1b) and (3), though individual λ and δ may vary from one person to another. By assumption, a>1/N; thus the aggregate monetary payoff is maximized when everyone is contributing their entire wealth to the public good. Similarly, the minimum monetary payoff will be highest when all are contributing. Clearly, a Pareto optimum will occur when $g_i = w_i$ for all i and $\pi_i = u_i = aN$, for all i.

If individuals are interacting non-cooperatively, we seek a Nash equilibrium. It would be helpful to apply the results of Bergstrom et al (1986) to characterize the equilibrium. However in Bergstrom et al (1986), utility of an individual agent is a function of g and G; in our case, the vector of g's enters each utility function due to the equity criterion (see Eqn. 5b below). If $\delta=0$ then in fact only g and G would be arguments of the utility function.

Let $G_{-i}\equiv\Sigma_{j\neq i}\, g_j$. From Eqn. (1), a person's monetary payoff will be

$$\pi_i = w_i - g_i + a\,(G_{-i} + g_i) = w_i - (1\text{-}a)g_i + aG_{-i} \qquad (4)$$

Assume person $m\neq i$ consumes the least amount of private goods (ie, $w_m - g_m$ is lowest); person m will have the lowest payoff. Thus person i chooses g to maximize u(g), defined as:

$$u_i(g_i)= (1\text{-}\lambda_i)(\, w_i - (1\text{-}a)g_i + aG_{-i}) + \lambda_i\{\delta_i[w_m - g_m + a(\Sigma_j g_j)]+(1\text{-}\delta_i)\Sigma_k[w_i - g_k + a(\Sigma_m g_m)]\} \qquad (5a)$$

$$= (1\text{-}\lambda_i)(\, w_i - (1\text{-}a)g_i + aG_{-i}) + \lambda_i\{\delta_i[w_m - g_m + a(G_{-i} + g_i)]+(1\text{-}\delta_i\,)[W+(aN\text{-}1)(G_{-i}+g_i)]\} \qquad (5b)$$

where $W=\Sigma_k w_k$ (5c)

To simplify, let $\Delta_i(g) = u_i(g) - u_i(g=0)$, which, after some simplifying, yields:

$$\Delta_i(g_i)= g_i\{a[1+(1\text{-}\delta_i)(N\text{-}1)\lambda_i]\text{-}1+\lambda_i\delta_i\} \qquad (6)$$

Clearly utility in Eqn. (6) is maximized at either $g_i=0$ or $g_i=w_i$, depending on the sign of the term in braces in Eqn (6), which leads to the following proposition:

<u>Prop. 1.</u> Assuming the N homogeneous player public goods game with C-R social preferences (Eqn. 3), then

(1) Efficient (Pareto Optimal) outcomes involve all agents contributing their entire resources to the public good; and

(2) The Non-cooperative Nash equilibrium involves all agents either contributing nothing (g=0) or everything (g=w) to the public good according to

$$g_i = 0 \text{ if } a < \bar{a}_i \qquad (7a)$$

$$g_i = w_i \text{ if } a > \bar{a}_i \qquad (7b)$$

where $\bar{a}_i = (1\text{-}\lambda_i\delta_i)/[1+\lambda_i(N\text{-}1)\,(1\text{-}\delta_i)]$ (7c)

In the case where $\bar{a}_i = a$, then any affordable contribution level for agent i is a Nash equilibrium.

Note in Prop. 1 that if $\lambda=0$ (standard preferences), then $\bar{a}=1$ and the Nash equilibrium is for all agents to contribute nothing to the public good, since by assumption $a < 1$. The $\bar{a}_i$ in Prop. 1 can be interpreted as the cutoff MPCR (varying from agent to agent) between cooperation and noncooperation. The effect of other-regarding social preferences ($\lambda > 0$) is to lower $\bar{a}$, effectively expanding the levels of MPCR wherein cooperation takes place. Note further that when efficiency is of some concern $\delta < 1$), then increasing N has the effect of lowering $\bar{a}$. The logic is simply that from an efficiency point of view, the payoff from contributing to the public good increases as the number of agents increase. When $\delta=1$ (utility depends on equity, not efficiency), N drops out of $\bar{a}$ and the effect of N on $\bar{a}$ disappears:

Corollary 1. Assuming the N homogeneous player public goods game with C-R social preferences as represented in Eqn. 3, then the cutoff level between noncooperation and cooperation for an individual agent ($\bar{a}$) as defined in Prop 1, exhibits the following comparative statics:

a. If $\lambda > 0$ and $\delta < 1$, then increasing the number of players (N) has the effect of lowering the cutoff MPCR level between cooperation and noncooperation ($\bar{a}$), effectively shrinking the range of values of MPCR associated with noncooperation.

b. If $\lambda = 0$ (conventional preferences) or $\delta = 1$ (equity but no efficiency), then increasing the number of players (N) has no effect on the cutoff MPCR level between cooperation and noncooperation ($\bar{a}$).

C. Aggregate Provision of the Public Good.

One of the interesting findings from Bergstrom et al (1986) is that there is a crucial value of wealth (w^*) such that agents who are poorer than w^* contribute nothing and agents who are richer contribute $w_i - w^*$ (thus private consumption is w^*). Does this also hold in case of social preferences and the linear model?

Let $P=\{i|\bar{a}_i>a\}$. The set P consists of all the agents who will unambiguously contribute to the public good in a Nash equilibrium (of course P may be empty). Without loss of generality, assume that the agents are ordered in ascending order of $\bar{a}_i$ (in other words, agent 1 has the lowest $\bar{a}$ and agent N has the highest $\bar{a}$). And let $\hat{\imath}$ be the value of i such that $\bar{a}_{\hat{\imath}} < a$ and $\bar{a}_{\hat{\imath}+1} \geq a$. The population is divided into noncontributors and contributors who contribute their entire wealth. Clearly the total provision of public goods is given by $G = \Sigma_{i\varepsilon P}\, w_i$. Furthermore, the Bergstrom et al (1986) neutrality result (their Theorem 5) applies in that the redistributions of income among contributors and among non-contributors will not change the aggregate amount of public goods provided (this follows directly from our Prop. 1). Further, a redistribution of wealth from contributors to noncontributors will decrease the aggregate provision of public goods.

Andreoni (1988) shows for the standard public goods game that as the size of the economy (N) grows the set of contributors shrinks, though in the limit the aggregate amount of public goods provided is nonzero, though finite. Thus even when there is a continuum of agents and the set of contributors is of measure zero, there is a finite nonzero aggregate contribution to the public good.

This question can be addressed by extending the analysis in Corollary 1: as N grows without bound, $\bar{a}$ approaches a limit of 0, provided $\lambda > 0$ and $\delta < 1$. In other words, provided social preferences have some concern for efficiency, as an economy grows, contributions will grow, until everyone is a contributor. This result is somewhat counterintuitive. Most likely this is an anomalous consequence of the linear model. One would expect the marginal utility for an individual of either private consumption or public welfare to diminish with increased levels of either variable. This does not of course happen in the linear model.

IV. THEORY: COALITIONS FACILITATING PROVISION

We now consider a slightly more complicated institution. We allow a subset of agents to form a coalition for the express purpose of coordinating contributions to the public good. Agents voluntarily join the coalition and may leave the coalition. Furthermore, any public goods provided by the coalition benefit both coalition members and non-members (thus it is not a club good in the standard sense). This leads to the obvious question: why would anyone join the coalition when the fringe enjoys all of the benefits and none of the costs of the coalition? The answer to this legitimate question lies in the nature of a Nash equilibrium. A Nash equilibrium is a state-of-the-world wherein it is not in any agent's individual interest to unilaterially change behavior. It is not relevant what path an agent took to find him or herself in the coalition or in the fringe.

In contrast to the previous section, we simplify the model somewhat by restricting all agents to have the same preferences (ie, λ and δ do not vary over i). Relaxing this assumption on preferences, would be natural extensions. Furthermore, we consider the problem from a static point of view, not in terms of repeated (through time) interactions. As we know from the theory of repeated games, the equilibrium that may be supported with repetition may be considerably richer than for a simple on shot static game.

As is standard in the literature on international environmental agreements (which, as was earlier argued, is essentially a problem of coalition formation to provide public goods), we view the problem as a two stage game. In the first stage, agents decide whether or not to join the coalition. In the second stage, agents decide how much to contribute to the public good, with the coalition acting as one – as a joint payoff maximizer. We solve the problem using backwards induction.

Before moving to these two stages, some notation is in order. Define the members of the coalition by C = {i|agent i is a member of the coalition}, the size of which we denote by n. Let W_C be the aggregate wealth of the coalition and G_C be the total contributions from the coalition members and G_F the total contributions from the fringe.

A. Contributions Stage

In the two-stage process, the second stage is the contributions stage, when the fringe and the coalition determine how much to contribute to the public good, conditional on the size and composition of the coalition. Which leads to our first result regarding the actions of the fringe.

Lemma 1. With homogeneous C-R preferences and agents divided into members of the coalition and members of the fringe, it is a dominant strategy for each member of the fringe to contribute everything or nothing to the public good, depending on the value of ā relative to a (where ā is defined by Eqn 7c):

Contribute w_i if $a > \bar{a}$ (8a)

Contribute 0 if $a < \bar{a}$ (8b)

Pf: Consider agent i in the fringe. As defined earlier, W and G_{-i}, are, respectively, total wealth and the contributions to the public good of all agents other than i. Let $\underline{\pi}$ represent the minimum payoff of all the agents other than i. Therefore the utility for agent i, should he contribute nothing to the public good (u_N) is

$$u_N = (1-\lambda)\ (w_i + a\ G_{-i}) + \lambda\{\delta\ \underline{\pi} + (1-\delta)[W - G_{-i}\ + aN\ G_{-i}\ ]\} \quad (9a)$$

and should he choose to contribute xi to the public good:

$$u_C = (1-\lambda)\ [w_i - x_i + a\ (G_{-i} + x_i)] + \lambda\{\delta\ (\underline{\pi} + ax_i) + (1-\delta)[W - G_{-i}\ - x_i + aN\ (x_i + G_{-i}\ )]\} \quad (9b)$$

The difference between these is easily determined and simplified:

$$u_C - u_N = x_i\{\ (1-\lambda)(a-1) + \lambda[\delta a + (1-\delta)(aN-1)]\} \quad (9c)$$

It is easy to see from Eqn (9c) that it is optimal for agent i to either contribute all or nothing, depending on the sign of the term in braces—positive implies all. With some simple manipulation, it is easy to verify that this condition is exactly the same as Eqn. 7. ■

We can similarly examine the incentives of the coalition. Clearly we can limit the analysis to the case wherein a < ā since otherwise (as shown in Lemma 1 above), it is a dominant strategy to contribute everything to the public good. It is straightforward to demonstrate the following analog to Lemma 1:

Lemma 2. Assume homogeneous C-R preferences and agents divided into members of the coalition and members of the fringe. Furthermore, in the C-R preferences, assume equity concerns for coalition members are with respect to the minimum payoff of agents outside the coalition. Then, conditional on the size of the coalition being n, it is a dominant strategy for the coalition to either contribute as much as possible ($G_C=W_C$) or as little as possible ($G_C=0$):

Contribute $G_C=W_C$ if $n > \tilde{n}$ (10a)

Contribute $G_C=0$ if $n < \tilde{n}$ (10b)

Where

$$\tilde{n} = (1-\lambda)/\{a+\lambda(1-\delta)[a(N-1)-1]\} \quad (10c)$$

Pf: Let m represent the identity of the agent with the smallest wealth outside the coalition. The aggregate utility for the members of the coalition when individual contributions are g_k for $k\varepsilon C$ is

$$\Pi_C(\mathbf{g}) = \Sigma_{k\varepsilon C}\{(1-\lambda)(w_k-g_k+a(G_C+G_F)) + \lambda[\delta(w_m+a(G_C+G_F)) + (1-\delta)\,\Sigma_i(w_i-g_i+a(G_C+G_F))\}$$

$$= (1-\lambda)[W_C + (an-1)(G_C+G_F)] + \lambda\delta n(w_m+a(G_C+G_F)) + \lambda(1-\delta)n[W+(aN-1)(G_C+G_F)] \quad (11)$$

and thus the payoff for the coalition contributing nothing is

$$\Pi_C(\mathbf{0}) = (1-\lambda)[W_C + (an-1)G_F] + \lambda\delta n(w_m+aG_F) + \lambda(1-\delta)n[W+(aN-1)G_F] \quad (12)$$

Which implies that the difference in payoff between contributing and not, $\Delta(\mathbf{g}) \equiv \Pi_C(\mathbf{g}) - \Pi_C(\mathbf{0})$ is

$$\Delta(\mathbf{g}) = G_C\{(1-\lambda)(an-1)+\lambda\delta na+\lambda(1-\delta)n(aN-1)\} \quad (13)$$

which clearly is maximized at G equals 0 or W, depending on the sign of the term in braces. It is easily shown that the term in braces is positive (which means G_C=W) iff

$$n > (1-\lambda)/\{a+\lambda(1-\delta)[a(N-1)-1]\} \quad (14)$$

This completes the proof. ■

Note in Lemma 2 that the result is conditional on the size of the coalition being n. Clearly if n < ñ, there is no gain from contributing. However, if n > ñ, any G>0 improves the aggregate utility of the coalition, and the more the better.

Note also in the above that if standard preferences apply (λ=0), then the right hand side of Eqn (14) simplifies to 1/a, which is the standard result in the coalition literature.

One obvious question is how moving from pure self-interested standard preferences (λ=0) to other regarding preferences changes the cutoff size of a coalition in Lemma 2. This is easily answered with comparative statics:

Lemma 3. Assuming homogeneous C-R preferences as in Lemma 2, with 0 < λ < 1, then

(1) Increases in λ result in decreases in ñ, as defined in Eqn. (10c);

(2) Increases in δ have ambiguous effects on ñ, depending on the value of a(N-1). For a > 1/(N-1), increases in δ increase ñ and for a < 1/(N-1), increases in δ decrease ñ.

(3) For δ<1, increases in N result in decreases in ñ.

Pf: Simply differentiating the right-hand-side of Eqn. (14) and then signing the result proves this lemma. ■

B. Membership Stage

We now turn to the first stage of the two-stage game. Each agent knows what will happen in the second stage, conditional on how large the coalition is, which is determined in the first stage. Each agent announces "in" or "out" in this stage and the well-known equilibrium conditions for the coalition to result from a Nash equilibrium in announcements are:

(a) Internal stability: no agent in the coalition can do better by unilaterally defecting

and

(b) External stability: no member of the fringe can do better by unilaterally joining the coalition

These lead to the following proposition on the size of the Nash coalition. In this proposition, we simplify by assuming the endowments of all agents are the same. Obviously this need not be the case:[4]

Prop. 2. Assume C-R preferences such that a non-cooperative equilibrium yields no contributions to the public good. Then for a two-stage public goods game with coalitions, if an equilibrium coalition exists, providing positive amounts of public goods, it will be of size ceil($\tilde{n}$), where $\tilde{n}$ is defined in Eqn (10c) and ceil(x) is the function which maps x into the smallest integer greater than or equal to x. For any coalition C, of size ceil($\tilde{n}$), if the richest member of the coalition (r) is not too wealthy in the sense of satisfying

$$w_r/W_C < \bar{e} \equiv \{(1-\lambda)a + \lambda[\delta a + (1-\delta)(aN-1)]\}/(1-\lambda) \qquad (15)$$

then the coalition is a Nash equilibrium.

Pf: We need not concern ourselves with coalitions of size $n < \text{ceil}(\tilde{n})$, since we know from Lemma 2 that they will provide no public goods. External stability of $n = \text{ceil}(\tilde{n})$ means that no member of the fringe wishes to join the coalition, expanding its size to $n = \text{ceil}(\tilde{n}) + 1$. This is identical to claiming all coalitions with $n \geq \text{ceil}(\tilde{n}) + 1$ are not internally stable—members of large coalitions prefer to defect to the fringe. We show this first.

Let n be the size of a coalition with $n \geq \text{ceil}(\tilde{n}) + 1$. Thus for both n and n-1 sized coalitions, it will be in the coalition's interest to contribute its entire wealth to the public good (by Lemma 2). Consider an arbitrary member of the coalition, d. We compare d's utility staying in the coalition (u_C) versus defecting to the fringe (u_F):

$$u_C = (1-\lambda)aW_C + \lambda\delta(w_m+aW_C)+\lambda(1-\delta)[W+W_C(aN-1)] \qquad (16a)$$

[4] This is a common result in the international environmental agreements literature, though in those cases the models are somewhat more restrictive (eg, homogeneous endowments and binary choice). See Barrett (2003) and Ulph (2004).

$$u_F = (1\text{-}\lambda)\,[w_d + a(W_C\text{-}w_d)] + \lambda\delta(w_m + a(W_C\text{-}w_d)) + \lambda(1\text{-}\delta)[W + (W_C\text{-}w_d)\,(aN\text{-}1)] \quad (16b)$$

where, as before, W_C is the wealth of the coalition (before the defection), w_d is the wealth of the defector, and w_m is the minimum wealth outside the coalition (assumed to remain the same before and after the defection). This implies that the gains from staying in the coalition are given by

$$\Delta = u_C - u_F \;=\; w_d\,\{(1\text{-}\lambda)\,(a\,\text{-}1) + \lambda\delta a + \lambda(1\text{-}\delta)(aN\text{-}1)\} \quad (17)$$

The term in braces is the same as the term in braces in Eqn. (9c). Since we are assuming $a < \bar{a}$, the term is negative. Thus it is optimal to defect. This demonstrates that no coalition with size $n > \text{ceil}(\tilde{n})$ is internally stable – defection occurs, independent of the wealth distribution.

All that remains to show is the condition for internal stability for coalitions of size $n = \text{ceil}(\tilde{n})$. As above, we examine the incentives for a random defector in the coalition, d. The defector compares payoff in the coalition (u_C) with payoff in the fringe with the coalition one member smaller (u_F), which we know from Lemma 2 involves no contributions to the public good. The payoff from remaining in the coalition is as defined in Eqn. (16a). The payoff in the fringe however is different, since in this case the fringe involves no contributions to the public good by anyone:

$$u_F = (1\text{-}\lambda)w_d + \lambda\delta w_m + \lambda(1\text{-}\delta)W \quad (18a)$$

which implies

$$\begin{aligned}\Delta = u_C - u_F \;&=\; (1\text{-}\lambda)aW_C + \lambda\delta(w_m + aW_C) + \lambda(1\text{-}\delta)[W + W_C(aN\text{-}1)] - (1\text{-}\lambda)w_d - \lambda\delta w_m - \lambda(1\text{-}\delta)W\\ &=\; (1\text{-}\lambda)(aW_C\text{-}w_d) + \lambda\delta aW_C + \lambda(1\text{-}\delta)W_C(aN\text{-}1)\\ &=\; -(1\text{-}\lambda)w_d + W_C\{(1\text{-}\lambda)a + \lambda[\delta a + (1\text{-}\delta)(aN\text{-}1)]\} \quad (18)\end{aligned}$$

which implies

$$\Delta > 0 \quad \text{iff} \quad w_d/W_C \;<\; \bar{e} \;=\{(1\text{-}\lambda)a + \lambda[\delta a + (1\text{-}\delta)(aN\text{-}1)]\}/(1\text{-}\lambda) \quad (19a)$$

$$= 1 + \{(1\text{-}\lambda)(a\text{-}1) + \lambda[\delta a + (1\text{-}\delta)(aN\text{-}1)]\}/(1\text{-}\lambda) \quad (19b)$$

The right hand side of Eqn. (19a) is positive and the portion in braces in Eqn (19b) is negative since $a < \bar{a}$ (see Eqn. 9c) which means that $\bar{e}$ lies between zero and one. For the coalition to be stable, Eqn. 19 must hold for all members of the coalition, which is satisfied if it holds for the richest member of the coalition.■

The fact that $\bar{e}$ lies between 0 and 1 and is completely determined by parameter values (not choices), implies that for w_r/W_C small enough, stability is assured. On the other hand, if the wealth of agent r dominates the coalitions so that w_r/W_C is close to one, then agent r will choose to defect. The implication is that coalitions of the right size with more uniform distributions of wealth are more likely to be stable than coalitions with great disparities in wealth. Disparities in wealth increase the possibility that a wealthier agent will find it better to leave the coalition, consuming his endowment rather than contributing it to the public good.

Note that for standard preferences which are not other-regarding (λ=0), $\bar{e}$=a and $\tilde{n}$=1/a. When all agents have the same initial endowment, w_r/W_C = 1/ceil($\tilde{n}$) < 1/$\tilde{n}$ = a. Thus such a coalition is stable. It would not take very much wealth variation to destroy the stability of this coalition. An alternative way of looking at this is that coalitions are more likely to occur among the subset of agents which share similar endowments.

Another special case is other-regarding preferences which are concerned with overall efficiency but not equity (δ=0). In this case, $\bar{e}$ is larger: $\bar{e} = a + \lambda(aN-1)/(1-\lambda) > a$. Thus there will be more tolerance of wealth disparities within a stable coalition.

C. Discussion

The standard linear model for coalition formation involves identical endowments and monetary payoffs. The main result from that literature is that all stable coalitions have size ceil(a). We have extended that result in a number of directions.

Introducing C-R social preferences, keeping endowments identical among agents, tends to lower the size of stable coalitions (Lemma 3). Introducing heterogeneity of agents does not change the size of coalitions but suggests that too much wealth disparity can cause a coalition to be unstable (Prop. 2).

The result on the size of coalitions may appear to contradict the results of Kosfeld et al (2009) who show that with Fehr-Schmidt preferences (in which players dislike payoff inequality), larger coalitions may be equilibria (even the grand coalition), depending on parameter values. However, this is not inconsistent with Lemma 3: provided a > 1/(N-1), then increases in concern for inequality (increases in δ) have the result of increasing the value of $\tilde{n}$, which is the size of the equilibrium coalition.

It is common to interpret the result on the size of stable coalitions as a “dismal” result in the theory of coalition formation for public goods – dismal in the sense that many MPCRs lead to very small coalitions and the larger the MPCR, the smaller the coalition (Barrett, 1994). We would argue that the more appropriate interpretation of the size of stable coalitions is the size of the *smallest effective coalition*. In reality, other tools will be used to keep a coalition together – incentives, punishments and interconnecting coalitions, to name a few. However, it is difficult to overcome the situation wherein a coalition is so small that it is not in the self-interest of the coalition to contribute to the public good.

Thus the fact that social preferences tend to shrink the size of an effective coalition to provide public goods is good news, at least in terms of public goods provision. Viewing preferences as social can expand the set of viable coalitions.

V EXPERIMENTAL EVIDENCE

So far, our discussion of coalitions and C-R preferences has been pure theory. We have no sense of how significant these other-regarding preferences may be. We thus turn to empirical measurement of the nature of C-R preferences in a laboratory setting.

Kosfeld, Okada and Riedel (2009) [abbreviated KOR here] conduct experiments on the linear public goods game and the formation of coalitions to provide public goods. One of the benefits of the *American Economic Review*'s policy on archiving data from published empirical papers is that the raw data from the KOR study are freely available.[5] Although the data do not fit our needs perfectly they are useful for estimating some of the parameters of C-R preferences for the subjects in their experiments.

In their experiments, they examine groups of four agents under two different institutional conditions. One is the pure public goods game as described here (used as a "control" by KOR); they consider two values of the MPCR – 0.40 and 0.65, though they only report limited information on play to the individual players. The other experiments (the primarily ones for KOR) involve the formation of coalitions, using the same two values of the MPCR. Their coalition experiments are similar to the structure described above in that there are multiple stages, with the membership stage preceding the contributions stage. However, there are some significant differences from our theoretical coalition model: (a) they have an intermediate step wherein the coalition votes on its contribution levels (difficult to avoid in an experiment); (b) there is a fixed cost of joining the coalition (there are none in our model); and (c) the coalition punishes members who do not contribute their entire endowment (this is absent from our model). In all of their experiments, agents have the same endowment.

Because of the similarity between their pure public goods games and the theoretical structure in this paper, we can use the results of the public goods game to estimate the values of the parameters of C-R preferences. Because the players in their experiments are only told the aggregate contributions, and not the individual contributions, we are unable to identify the distributional parameter (δ) and thus set it to zero, focusing on the social efficiency parameter (λ). Thus the C-R preferences we estimate involve utility depending on a weighted sum of private benefits and group benefits.

A. Data

Individuals played the public goods game in groups of four individuals, randomly assigned. Two values of the MPCR were used, 0.40 and 0.65. The two public goods games are referred to as PG40 and PG65, respectively. Each group repeats play for 20 rounds, in an effort to move to a Nash equilibrium. There are 10 sessions of four for the PG40 game and 9 sessions of four for the PG65 game. Thus there are 800 observations for PG40 (10 sessions x 20 rounds x 4 players) and 720 observations for PG65. The experiments were performed at the University of Amsterdam.

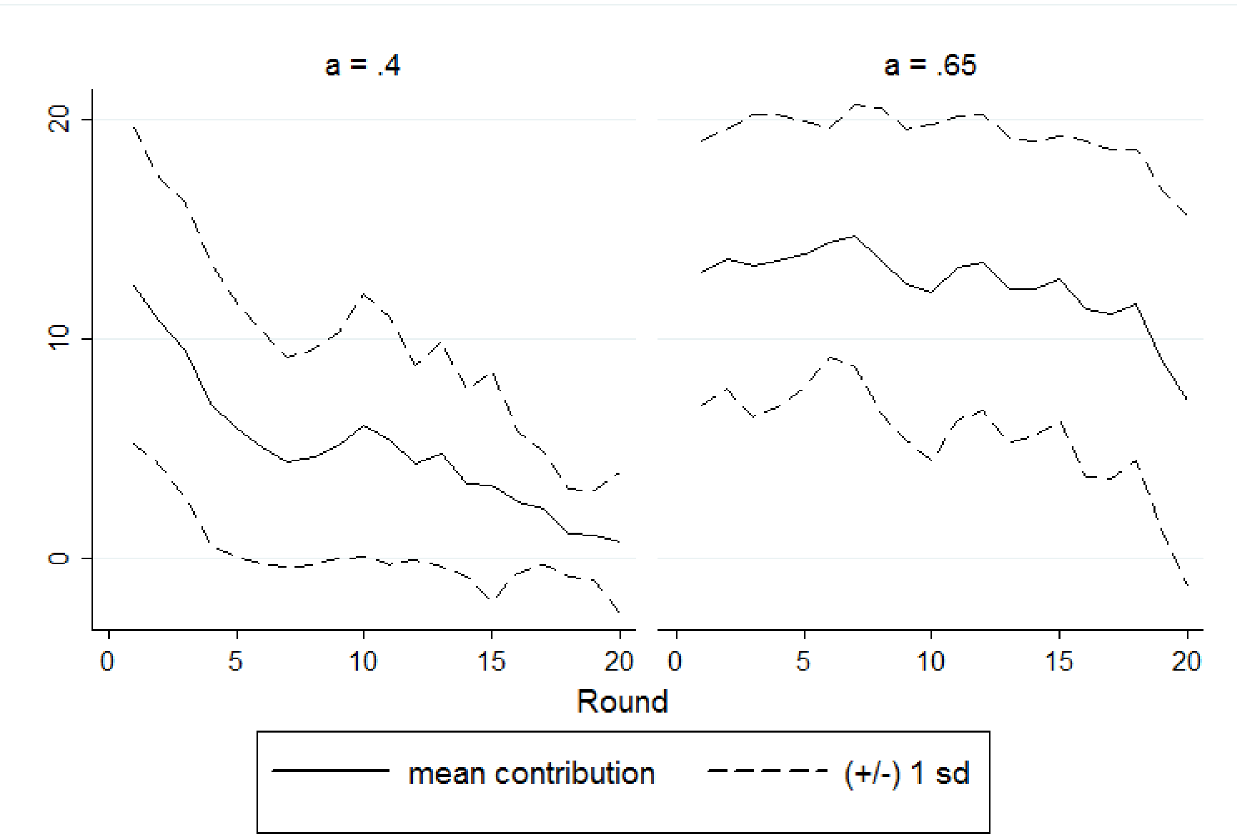
The currency of the experiment is "points" which may be converted into euros at the end of the experiment (40 points = 1€). In each round, each agent is given 20 points to use – either to contribute

[5] See http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.4.1335

to the public good or to convert to a private payoff at the end of the experiment. Thus each agent may contribute anywhere from 0 to 20 points to the public good. That is the only action an agent takes in each round. At the end of the round the agents are told what the other players have contributed (in aggregate, not individually).

Figure 2 is a summary of the total contribution in each game, by round. Shown is the average over the 9-10 sessions for each MPCR as well as the plus or minus one sigma for total contributions, to give an idea of the dispersion over the different sessions. It is interesting that the average contributions drop over the course of the different rounds much more rapidly for the lower MPCR. The higher MPCR seems to have a more pronounced end-effect (contributions drop in the last round).

Figure **2: Contributions by round from two public goods games in KOR**



B. Analysis

Using the 1520 observations from KOR, we estimate two versions of C-R utility, Eqn. 3. In both cases we restrict δ to be 0 since the experiments did not communicate enough information to participants for them to know the minimum payoff.[6] Despite this, we can learn how participants trade-off their own personal payoffs with aggregate payoffs to the group, an important component of social preferences. In one of the two cases we let λ_i vary over participants (Case II). In the other case, we fix λ to be the same for all participants (Case I). We have not attempted to utilize other variables (such as expectations, beliefs,[7] or time dummies) to explain the behavior over the different rounds of the game (such as the significant decline in contributions over time in the left hand panel of Fig. 2).

In both cases, we approximate the continuous choice for contributions from 0 to 20 to a discrete choice so that we may use multinomial logit to estimate the model. In particular, we divide the interval 0-20 into five bins (0-4, 4-8, etc.), considering the midpoint of the bin (2, 6, etc) to be the choice made. Case I is the standard logit and case II is a mixed logit, with λ varying over the population lognormally (by assumption, to avoid negative λ), as described by Revelt and Train (1998).

In each model there is really only one parameter – λ. Table I shows the parameter estimates and other characteristics of the estimation, along with a reference model in which λ is restricted to be zero. Note that the mean value of λ in the random coefficients is slightly higher than for the fixed coefficient logit. A likelihood ratio test on the hypothesis that all of the λ_i are the same is clearly rejected. Furthermore, the hypothesis that $\lambda=0$ is also clearly rejected.

[6] A main goal of KOR is examine the formation of coalitions to provide public goods and the role of fairness. The paper is in two parts – theory and experiments. In the theory section, they develop a theory of coalition formation for public goods with F-S preferences (Eqn 2 above) and show that coalitions will be larger with such preferences. Their experimental results show that participants tend to establish larger coalitions than simple payoff theory would suggest. They suggest in their conclusions that issues of fairness appear to be driving the tendency to form larger coalitions, though without utilizing F-S preferences. In fact, the KOR experiments do not appear to be sufficient to estimate F-S preferences (or even for subjects to utilize F-S preferences). KOR only provide participants with information on their own payoff, the number of people in the coalition and the aggregate payoff. F-S preferences require knowledge of the contribution or payoff vector over all players and that information is not provided in their experiments. This is not to suggest there are errors in KOR – the authors only indicate that their experimental results suggest the importance of fairness, without reference to F-S.

[7] See Fischbacher and Gächter (2010) attempt to explain the decline in cooperation over rounds in a public goods game, utilizing conditional cooperation, beliefs and self-identified social preferences.

Table I: Estimated Coefficient from KOR Data

	Reference	Case I	Case II
λ	0	0.251	
Std error		0.006	
mean(λ)			0.330
Std error on mean			0.021
Observations	7600	7600	7600
Log likelihood	-2354	-2149	-1878

Note: Reference is standard preferences with λ restricted to be zero. Case I is a conditional logit and Case II is a random coefficients logit (mixed logit). For Case II, the mean of the estimated λ is shown along with the standard error on the estimated mean λ (lognormal distribution), reflecting the dispersion over the population. Number of observations reflects 1520 data points with five possible choices for each data point.

The mixed logit model estimate of λ is quite similar to that in the fixed coefficient model. Figure 3 shows the probability density function on how the random coefficients are distributed; Figure 4a converts that into a pdf over $\bar{a}$, using Eqn. 7c. Figure 4b shows the associated cumulative density function for $\bar{a}$. Recall that $\bar{a}$ is the minimum MPCR (a) for which it is individually rational for agents to contribute to the public good and that for standard preferences, $\bar{a}=1$.

Figure 3: Estimated density for λ in mixed logit (random coefficients)

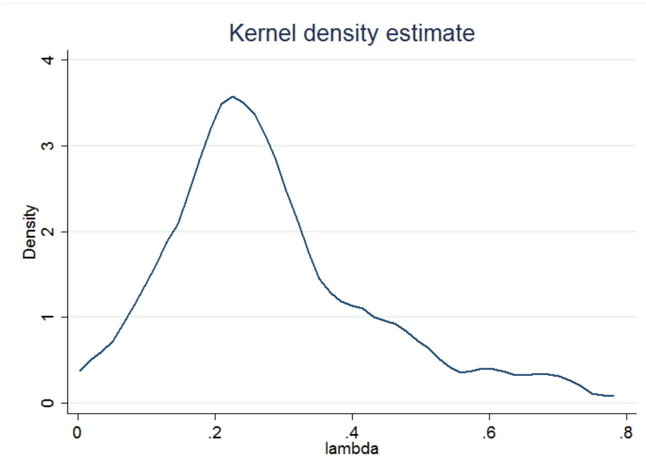
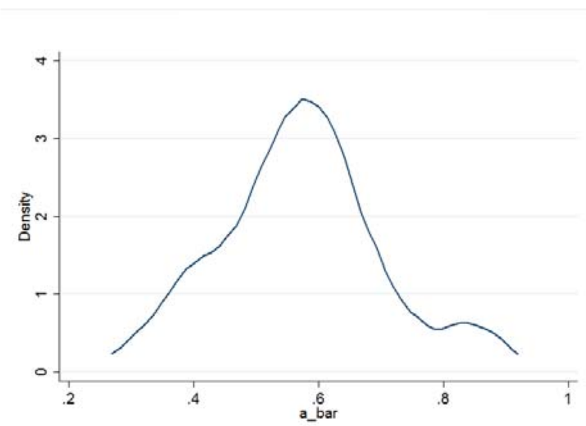
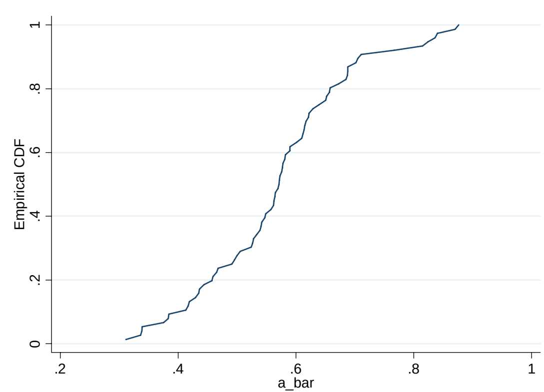


Figure 4: PDF (a) and CDF (b) for ā induced by density on λ in mixed logit.

(a)



(b)



C. Implications

First consider the random coefficients mixed logit (Case II). Inspecting Fig. 4b, we see that for MPCR=a=0.4, approximately 10% of the agents find their $\bar{a}_i < a$ (which is the condition for contributing to the public good)—not many (consistent with contributions in later rounds as shown in Fig. 2). In contrast, for MPCR=a=0.65, Fig 4b indicates that approximately 80% of agents find their $\bar{a}_i < a$, and thus will find it individually rational to contribute to the public good. This is also roughly consistent with Figure 2.

Although we cannot utilize the coalition experiments of KOR directly, we can use the above estimates of the value of λ in C-R preferences to compute the expected coalition sizes using the theory develop in this paper, and to compare those coalitions with those with standard preferences. Table II shows the values of ā and ñ associated with both standard preferences and C-R preferences.

Table II: Computed values of ā and ñ (from Eqn. 7c and 10c)

	ā	ñ
Standard Preferences (λ=0)		
MPCR = 0.4	1	3
MPCR = 0.65	1	2
C-R Preferences (λ=0.244)		
MPCR = 0.4	0.577	2
MPCR = 0.65	0.577	1 (no coalition)

Note in Table II that ā with C-R preferences is considerably smaller than with standard preferences. Recall that ā represents the cutoff value for MPCRs between contributing and non-contributing being individually rational. Thus for the case of C-R preferences and MPCR=0.65, there is no need for a coalition, since agents will individually contribute all of their assets to the public good. This is consistent with the result that ñ drops from 2 to 1 – no coalition develops with C-R preferences because no coalition is necessary.

Thus utilizing C-R preferences tends to expand the values of MPCR which lead to unilateral cooperation, beyond what would be expected with standard preferences. That could be viewed as a positive result for the provision of public goods. The other effect of C-R preferences is to lower the minimal size of contributing coalitions (ñ), which is an ambiguous finding. On the one hand, this suggests that the power of coalitions to contribute to public goods will be reduced. On the other hand, it implies that smaller coalitions, which may be easier to coordinate, can have impact.

VI CONCLUSIONS

In this paper we revisit the important question of voluntary provision of pubic goods. In particular, we are interested in two issues. One is the role of social preferences. How do social preferences change the received wisdom on contributions? The second issue is the role of voluntary coalitions in coordinating the provision of public goods. This institution is important in the literature on international environmental agreements (where the public good is pollution abatement). Little is known of how social preferences modify what we know about such coalitions and by extension, international environmental agreements (viewed as abstract economic coordination entities).

We adopt the specification of social preferences due to Charness and Rabin (2002), primarily because it contains three important ingredients that characterize many discussions of social preferences: private gain, equity and social efficiency. Using a linear public goods model with a fixed MPCR but an arbitrary distribution of wealth, we find that a major consequence of using social preferences is to lower the threshold for contributing to the public good being individually rational. This is in contrast to theory with standard preferences where for any MPCR less than one, free-riding is individually rational. We also confirm the neutrality theorem in the context of social preferences: any redistribution of wealth among contributors leaves the aggregate provision of public goods unchanged.

In extending the analysis to voluntary coalitions, we show that social preferences tend to reduce the size of an equilibrium coalition. Another interpretation of that result is that social preferences expand the set of coalitions for which it is rational for the coalition to provide the public good. We do find that a factor that tends to destabilize coalitions is inequality of wealth. This is not because of preferences for equity but rather because defection is more likely. When coalitions members have similar wealth the coalition is more likely to be stable than when the distribution of wealth is more unequal.

In order to illustrate the significance of these results, we have used experimental data on public goods experiments – not our own but those of Kosfeld et al (2009). Using these data we are able to estimate the parameters of the social preference function and demonstrate how results differ when using social preferences. Such empirically estimated preferences appear to explain the experimental results better than standard preferences.

Although this paper is couched in terms of the problem of voluntary provision of public goods, it is motivated in part by the literature on the international environmental agreement (IEA). Conceptually an IEA is simply a group of agents (countries) with utility functions attempting to provide a public good (pollution abatement). Thus the results here have significant implications for both the theoretical IEA literature as well as IEA policy.

REFERENCES

Andreoni, James, "Privately Provided Public Goods in a Large Economy: The Limits of Altruism," *J. Pub. Econ.*, **35**:57-73 (1988).

Andreoni, James, "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," *J. Pol. Economy*, **97**:1447-58 (1989).

Andreoni, James, "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *Economic J*., **100**:464-77 (1990).

Barrett, Scott, "Self-Enforcing International Environmental Agreements," *Oxford Economic Papers*. **46**:878-94 (1994).

Barrett, Scott, "A Theory of Full International Cooperation," *J. Theoretical Politics*, **11**:519-41 (1999).

Barrett, Scott, *Environment and Statecraft* (Oxford University Press, Oxford, 2003).

Bergstrom, Ted, Larry Blume and Hal Varian, "On the Private Provision of Public Goods," *J. Pub. Econ*., **29**:25-49 (1986).

Bliss, C and B. Nalebuff, "Dragon Slaying and Ballroom Dancing – The Private Supply of a Public Good," *J. Pub. Econ.*, **25**:1-12 (1984).

Burger, Nicholas and Charles D. Kolstad, "Voluntary Public Goods Provision, Coalition Formation and Uncertainty," NBER Working Paper 15543, Cambridge, Mass. (Nov. 2009).

Carraro, C. and D. Siniscalco, "Strategies for the International Protection of the Environment." *J Public Econ.*, **52**:309-28 (1993).

Charness, Gary and Mathew Rabin, "Understanding Social Preferences with Simple Tests," *Quart. J. Econ*., **117**:817-69 (2002).

Charness, G., L. Rigotti, and A. Rustichini, "Cooperation rates as a function of payoffs for mutual cooperation," UC Santa Barbara Working Paper (2008).

Charness, Gary and Chun-Lei Yang, "Efficient Public Goods Provision with Endogenous Group Size and Composition: An Experiment," *J. Pub. Econ*. (forthcoming, 2011).

Chaudhuri, Ananish, "Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature," *Exp. Econ*. (forthcoming, 2010).

d'Aspremont, C., A. Jacquemin, J. Jaskold-Gabszewicz and J. Weymark, "On the Stability of Collusive Price Leadership," *Canadian J. Economics*, **16**:17-25 (1983).

Dannenberg, Astrid, Andreas Lange and Bodo Sturm, "On the Formation of Coalitions to Provide Public Goods – Experimental Evidence from the Lab," NBER Working Paper 15967, Cambridge, Mass. (May 2010).

Donsimoni, M.-P., N.S. Economides, and H.M. Polemarcharkis, "Stable Cartels," *International Economic Rev*., **27**:317-27 (1986)

Fehr, Ernst and Klaus M. Schmidt, "A Theory of Fairness, Competition and Cooperation," *Q. J. Econ*., **114**:817-68 (1999).

Finus, Michael, *Game Theory and International Environmental Cooperation* (Edward Elgar, Cheltenham, 2001).

Fischbacher, Urs and Simon Gächter, "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments," *Amer. Econ. Rev*., **100**:541-56 (2010).

Isaac, R.M. and J. M. Walker, "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism." *Quart. J. Econ.*. **103**:179-99 (1988).

Isaac, R.M., J.M. Walker and A.W. Williams, "Group Size and the Voluntary Provision of Public Goods: Experimental Evidence Utilizing Large Groups." *J Pub. Econ.,*. **54:**1-36 (1994).

Kahneman, Daniel, Jack L. Knetsch and Richard Thaler, "Fairness as a Constraint on Profit Seeking: Entitlements in the Market," *Amer. Econ. Rev*., **76**:728-41 (1986).

Kim, O. and J.M. Walker, "The Free Rider Problem: Experimental Evidence." *Public Choice*. **43**:3-24 (1984).

Kosfeld, Michael, Akira Okada, and Arno Riedl, "Institution Formation in Public Goods Games," *Amer. Econ. Rev.,* **99**:1335-55 (2009).

Lange, Andreas, "The Impact of Equity Preferences on the Stability of International Environmental Agreements," *Env. Res. Econ*., **34**:247-67 (2006).

Ledyard, John O., "Public Goods: Some Experimental Results," Ch. 2 in J. Kagel and A. Roth (Eds), *Handbook of Experimental Economics* (Princeton University Press, Princeton, NJ, 1995).

Murdoch, J.C. and T. Sandler, "The Voluntary Provision of a Pure Public Good: The Case of Reduced CFC Emissions and the Montreal Protocol," *J. Pub. Econ*., **63**:331-49 (1997).

Revelt, David and Kenneth Train, "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level," *Rev. Econ. Stat*., **80**:647-57 (1998).

Schelling, Thomas C., "Hockey Helmets, Concealed Weapons, and Daylight Saving: A Study of Binary Choices with Externalities," *J. Conflict Resolution*, **17**:381-428 (1973).

Ulph A., "Stable international environmental agreements with a stock pollutant, uncertainty and learning," *J. Risk and Uncertainty*, **29**:53-73 (2004).