NBER WORKING PAPER SERIES

PUBLIC GOODS AGREEMENTS WITH OTHER-REGARDING PREFERENCES

Charles D. Kolstad

Working Paper 17017
http://www.nber.org/papers/w17017

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2011

Department of Economics and Bren School, University of California, Santa Barbara; Resources for the Future; and NBER. Comments from Werner Güth, Kaj Thomsson and Philipp Wichardt and discussions with Gary Charness and Michael Finus have been appreciated. Outstanding research assistance from Trevor O'Grady and Adam Wright is gratefully acknowledged. Funding from the University of California Center for Energy and Environmental Economics (UCE3) is also acknowledged and appreciated. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Public Goods Agreements with Other-Regarding Preferences
Charles D. Kolstad
NBER Working Paper No. 17017
May 2011, Revised June 2012
JEL No. D03,H4,H41,Q5

ABSTRACT

Why cooperation occurs 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 concern shifted to 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 and high levels of free riding. 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 involve higher levels of 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 contrast to much of the literature, we treat the wealth of agents as heterogeneous. In such cases, we find that transfers among agents of the coalition may be necessary to sustain cooperation (regardless of the nature of preferences). An example drawn from experiments is provided as an illustration of the effectiveness of social preferences.

Charles D. Kolstad
Department of Economics
University of California
Santa Barbara, CA 93106
and NBER
kolstad@econ.ucsb.edu

I. INTRODUCTION

Why cooperation occurs when noncooperation appears to be individually rational has been an issue in economic theory for at least a half century. Olson (1971) and Becker (1974) examined individual motivation for cooperation in an otherwise selfish world. Contemporaneously, applied game theorists were concerned with 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) – some apparent cooperation (but not much) was seen to be consistent with self-interest. In the 1990's, a small theoretical literature began to develop on the formation of coalitions to facilitate cooperation to solve externality problems – international environmental agreements (IEAs -- eg, Barrett, 1994 and Carraro and Siniscalco, 1993), much in the tradition of Olson's (1971) examination of groups to foster collective action. That theoretical work, to, has been pessimistic on the extent to which free-riding incentives can be overcome.

Empirical evidence that cooperation is more prevalent than simple theory might suggest began to appear around 1980. Abrams and Schitz (1978) analyze time-series data on the effect of increased government provision of public goods on private provision of public goods; they concluded that the observed modest crowding out was inconsistent with pure self-interest (which, they suggest, would result in a one-for-one crowding out of private contributions by public contributions). 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 is common and cooperation is difficult to sustain in the standard public goods problem. Kim and Walker (1984) were one of the first to offer evidence from laboratory experiments suggesting people are 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) are also fraught with disparities between laboratory behavior and experiments.

Partially 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 (though the terminology varies considerably). Andreoni (1990) was concerned with the evidence of weak crowding out of charitable donations and suggested that agents receive utility from public goods through several routes. In addition to direct benefits from the public good, he suggested that agents receive utility from the *act of contributing* to the public good – something he terms "warm glow." The reason for that warm glow is ambiguous – is it from an altruistic concern about the well-being of fellow man or simply guilt-aversion, as is sometimes the case in making a routine tip in a restaurant?

This perspective that simple utility is inadequate to explain cooperation led to a burgeoning literature on social preferences (eg, Fehr and Schmidt, 1999; Charness and Rabin, 2002). The 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. Although some criticize the experimental social preference literature as *ad hoc*, the fact is that virtually all of the literature seeking to explain cooperation lacks an axiomatic foundation.

Despite this progress, 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 can endogenous institutions facilitate contributions; ie, how should the theory of coalitions to provide public goods (as in the IEA literature) be modified when agents have social preferences? Put differently, how does the nature of preferences shape public goods contributions and the formation of coalitions to facilitate public goods provision?

This paper is one of the first papers to address these unanswered questions.[1] Rather than offer our own theory of social preferences, we start with 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

[1] Kosfeld et al (2009) examine the formation of institutions to facilitate public goods provision.

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 with empirical consequence emerge from this paper. We find that theoretically, and in the standard linear public goods game, social preferences do enhance contributions to the public good and expand cooperation. Furthermore, income is not relevant to the decision to contribute, though it is relevant to the aggregate provision of public goods (the richer contribute more). Although this is intuitive, the mechanism for expansion of cooperation is new and is confirmed in our experimental illustration.

The use of coalitions to expand cooperation is a relatively new addition to the standard public goods literature. Allowing the formation of coalitions does increase the aggregate provision of public goods, regardless of the nature of social preferences. Furthermore, Social preferences tend to shrink the minimum viable contributing coalition, as might be expected. But this also results in a smaller stable coalition. 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. However, if wealth transfers may occur within the coalition, the the surplus gained by the coalition is sufficient to equalize the incentives to defect within the coalition and thus neutralize the results on income disparities.

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, a 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_i and g_i to maximize utility, subject to a budget constraint:

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

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 (though not the average *among the contributors*) and the size of the contributing group approaches zero. However, aggregate contributions approach a limit which is finite and nonzero. He points out that this result 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). This is consistent with the crowding out literature (Abrams and Schitz, 1978) mentioned in the introduction.

A significant body of experimental work has accumulated on this issue as well. 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 (Smith, 1980). 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 "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 to the individual of that provision. 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.[2] 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 several decades, 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 (see Sobel, 2005). One of the first extensions of this nature is the model of "impure altruism" by Andreoni (1989,1990), building in part on Olson (1971) and Becker (1994). Impure altruism holds that there are two avenues for personal utility

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

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 the individual contribution. The individual may appear altruistic but that is because the individual obtains utility from giving. Thus the utility function in Eqn. (1) 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.

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 partial contradiction to this result, suggesting that efficiency also plays a role in outcomes in prisoner's dilemma games. To illustrate, in the Prisoner's Dilemma game shown in Fig. 1, theory would suggest defection will repeatedly occur. 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; $2<x<7$.

A number of authors have suggested that groups of people interacting strategically typically fall into at least two groups: self-interested and cooperators. This participant heterogeneity is said to explain the levels of cooperation observed in experimental and empirical data. Andreoni and Miller (2002) report experimental results for a dictator game and find 23% of respondents behave selfishly; the remainder show some degree of altruism. Fischbacher and Gachter (2010) conduct experiments with linear public goods game (similar to this paper) and find that 23% of participants are free-riders, contributing nothing; Fischbacher et al (2001) report a third are free-riders. Many of the remaining subjects are "conditional cooperators," cooperating conditional on others cooperating. Fehr and Schmidt (1999) review a number of papers with experimental results on public good games and conclude the papers, on average, involve a much higher fraction of pure free-riders (no contributions to the public good) – 73%. However, even with such a high fraction of free-riders, one can infer that the remaining subjects are behaving at variance with the purely selfish model. The upshot is that in groups, one might expect some heterogeneity in preferences—some agents are purely selfish and others display some sort of altruistic behavior.

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 endogenous formation of groups or coalitions of players to coordinate the provision of public goods. This is in fact the thrust of Olson's (1971) seminal treatise on 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 returned to addressing 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[3]) is that the noncooperative behavior is represented as a two stage game (and countries are posited to act as rational utility-maximizing agents). In the first stage (the "membership game") agents decide whether they wish to be in a coalition (an IEA). Specifically, each agent 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 (though he assumes a leader-follower structure rather than a two stage Nash equilibrium). Unfortunately, he is unable to come up with many analytic results without simplifying;[4] 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 coalitional outcome and a noncooperative outcome).[5] The intuition behind this is straightforward. An equilibrium coalition is held together by the fact that if any one player defects, the cooperative strategy of the coalition will fall apart. Thus the equilibrium coalition is the smallest one

[3] See d'Aspremont et al (1983) and Donsimoni et al (1986).

[4] Subsequent authors have refined this model, solving it model analytically. See Finus (2001), Diamantoudi and Sartzetakis (2006), and Rubio and Ulph (2006).

[5] 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 (2001).

possible for which contributing to the public good (abatement) is collectively rational for the coalition. When there is more to be gained from cooperation, then the minimal viable abating coalition is smaller; when there is less to be gained, it takes more coalition members for abating to be collectively rational.

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 agents free-riding ($g_i = 0$) and coalition members undertaking some pollution reduction (abatement), providing the coalition is large enough. Let n^* be the smallest size of a coalition which chooses to abate. A basic result is that $n^* = 1/MPCR$ and all stable contributing coalitions are of this size. One can also show that the benefits from cooperation increase in MPCR whereas the size of a coalition decreases in MPCR, following the same logic as in the previous paragraph. The assumption of identical endowments is almost universal in this literature, despite the fact that the wealth of a country seems, in the real world, 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), who explores the significance of equity in reaching agreement in an IEA. In fact, in one of the few empirical papers on this issue, Lange et al (2007) survey attitudes of individuals involved in climate negotiations and find a strong preference for equity. Whether this translates into countries caring about equity is another matter.

A number of authors focus on incentives to hold coalitions together, whether these be punishments or transfers among agents. 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 an agent defect.

One of the first recent 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 (applied to coalition members). Their theoretical 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 than predicted by theory.

III. THEORY: PRIVATE PROVISION

A. Basic Conditions.

Let there be $i=1,\ldots,N$ agents, each with wealth w_i, choosing between private goods consumption, x_i, and contributions to public goods, g_i. Building on the terminology of Andreoni (1990), agents are *impurely altruistic;* i.e., they have an individual utility function (u_i) that is additively separably into an *egoistic* or *self-regarding* component (π_i) and an *altruistic* or *other-regarding* component (α_i). Utility can be viewed as a convex combination of these two components:

$$u_i(x_i,g) = \lambda_i \; \pi_i(x_i,G) + (1-\lambda_i)\, \alpha_i(\boldsymbol{\pi}) \qquad (2)$$

where $\lambda_i \varepsilon$ [0,1] is the parameter reflecting the extent to which the agent is selfish vs. altruistic. Note that in Eqn (2), egoistic (self-centered) utility is given by the function π, which is the utility from the payoff from consumption ignoring the well-being of others. The altruistic (other-regarding) component is given by the function α, which depends on the vector of egoistic utility for the other agents. This is a somewhat paternalistic representation of altruism in that agents care only about the egoistic payoffs of others, not the overall utility of others. Additive separability is a modest restriction, though consistent with the literature, as is seen below.

Rather than postulate a general egoistic utility function which depends on individual and aggregate contributions (as in Bergstrom et al, 1986), we linearize the egoistic payoffs. This is consistent with a much of the literature, particularly in the experimental realm. It also allows us to consider heterogeneous preferences, often assumed away in more general models. In the absence of the altruistic component, this is a standard homogeneous preferences linear public good game. Specifically, the egoistic component (which we will also refer to as the monetary payoff) is given by

$$\pi_i = x_i + aG \quad \text{where } x_i + g_i = w_i,\ G = \Sigma_i\, g_i \tag{3a}$$

$$= w_i - g_i + aG \quad \text{where } G = \Sigma_i\, g_i \text{ and } 0 \le g_i \le w_i \tag{3b}$$

Here w_i is wealth, g_i 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 constraint $x_i+g_i=w_i$. The parameter 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 the MPCR, a, to be in the open interval (1/(N-1),1). Thus we are excluding a=1, for which the agent would be indifferent between unilaterally contributed or not. We also exclude very small a, for which contributing may not even be collectively rational. 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.

Although a linear payoff is common in the empirical and experimental literature, this is equivalent to assuming gains from the private good are a perfect substitute with gains from the public good. And this 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 more difficult to provide an analytic representation of the Nash equilibrium, particularly when heterogeneous other-regarding preferences are treated.

There are several representations of the altruistic component of utility in Eqn. 2. We consider two of the major representations in the literature: Fehr and Schmidt (1999) and Charness and Rabin (2002). Andreoni (1990) in the empirical implementation of his model assumes altruism is manifest by introducing g_i into utility – the agent receives a warm-glow from giving. The notion of warm-glow would appear to be closely related in particular to Charness and Rabin (2002), though the mechanism whereby giving to the public good generates utility is more ambiguous and undefined in Andreoni (1990).

Fehr and Schmidt (1999) stipulate that $\alpha_i(\boldsymbol{\pi})$ depends on weighted average payoffs (for other agents):

$$\alpha_i(\boldsymbol{\pi}) = - \gamma_i/(N\text{-}1)\ \Sigma_{j\neq i}\max(\pi_j - \pi_i,0) - \beta_i/(N\text{-}1)\ \Sigma_{j\neq i}\max(\pi_i - \pi_j,0) \qquad (4)$$

where the γ_i and β_i are in the interval [0,1) and reflect aversion to personally disadvantageous (people doing better than i) and advantageous (people doing worse than i) inequality, respectively (it is assumed that $\gamma_i \leq \beta_i$). 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. (2-4) 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 the altruistic component of utility depends on 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):

$$\alpha_i(\boldsymbol{\pi}) = [\, \delta_i \min_{j\neq i} \pi_j + \varepsilon_i \Sigma_j \pi_j \,]/(1 - \lambda_i) \qquad \text{where } \delta_i , \varepsilon_i \geq 0 \text{ and } \delta_i + \varepsilon_i + \lambda_i = 1 \qquad (5)$$

In Eqn. (5), δ_i reflects the relative importance to agent i of distribution/equity and ε_i reflects the importance of efficiency. We refer to Eqn. (2-3, 5) as C-R preferences, though the equity term in Eqn. (5) 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. Some readers may find the restrictions implicit in preferences given by Eqn. (2-3,5) troubling – preferences are linear and altruism is narrowly defined. However, that is how the literature, primarily experimental, has approached this problem. We adopt a form of preferences (C-R) widely used and cited in the literature.

B. Efficient and Noncooperative Outcomes

Assume C-R preferences as characterized in Eqn. (2-3) and (5), where individual λ, δ and ε may vary from one agent to another. By assumption, a>1/N; thus the aggregate monetary (egoistic) payoff is maximized when everyone is contributing their entire wealth to the public good. Similarly, the minimum monetary (egoistic) payoff over all agents 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. 7b below). Only if δ=0 would g and G be the only arguments of the utility function.

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

$$\pi_i = w_i - g_i + a\,(G_{-i} + g_i) = w_i - (1-a)g_i + aG_{-i} \tag{6}$$

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

$$u_i(g_i) = \lambda_i[\, w_i - (1-a)g_i + aG_{-i})] + \delta_i[w_m - g_m + a(\Sigma_j g_j)] + \varepsilon_i \Sigma_k[w_i - g_k + a(\Sigma_m g_m)] \tag{7a}$$

$$= \lambda_i[\, w_i - (1-a)g_i + aG_{-i}] + \delta_i[w_m - g_m + a(G_{-i} + g_i)] + \varepsilon_i\, [W + (aN-1)(G_{-i} + g_i)] \tag{7b}$$

where $$W = \Sigma_k w_k;\ \lambda_i,\ \delta_i,\ \varepsilon_i \geq 0;\ \lambda_i + \delta_i + \varepsilon_i = 1 \tag{7c}$$

To simplify, let $\Delta_i(g) \equiv u_i(g) - u_i(g=0)$, which can be written, after some simplifying, as:

$$\Delta_i(g_i) = g_i\{a[1+(N-1)\varepsilon_i]-1+\delta_i\} \equiv g_i\{a + \varepsilon_i\, [a(N-1)-1] - \lambda_i\} \tag{8}$$

Clearly utility in Eqn. (8) is maximized at either g_i=0 or g_i=w_i, depending on the sign of the term in braces in Eqn (8), which leads to the following proposition:

Prop. 1. Assuming the N homogeneous player public goods game with C-R social preferences (Eqn. 2-3,5), then

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

(2) The Non-cooperative Nash equilibrium involves each agent either contributing nothing (g_i=0) or everything (g_i=w_i) to the public good according to

$g_i = 0$ if $a < \bar{a}_i$ (9a)

$g_i = w_i$ if $a > \bar{a}_i$ (9b)

where $\bar{a}_i = (1-\delta_i)/[1+\varepsilon_i(N-1)]$ (9c)

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

The intuition behind Prop. 1 is straightforward. With standard egoistic preferences, the cutoff between contributing or not is a=1. With social preferences, the cutoff is lower: $\bar{a} \leq 1$. Eqn. (9c) simply defines how social preferences reduce this cutoff. We assume that where there is ambiguity ($a = \bar{a}_i$), agents contribute fully.

Note in Prop. 1 that if $\lambda_i=1$ (standard preferences—all weight is on egoistic payoffs), then $\delta_i = \varepsilon_i = 0$ (by Eqn. 7c) and $\bar{a}_i=1$: the Nash equilibrium is for all agents to contribute nothing to the public good, since by assumption $a_i < 1$. 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 with some concern for efficiency ($\varepsilon_i > 0$), keeping δ_i constant, is to lower $\bar{a}_i$, effectively expanding the levels of MPCR wherein cooperation takes place. This result is quite similar to Proposition 4 in Fehr and Schmidt (1999), a theorem in which the authors expand the set of MPCRs for which contributing is individually rational. Note further that in Prop. 1 above, when efficiency is of some concern, then increasing N has the effect of lowering $\bar{a}_i$. 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 $\varepsilon_i=0$ (utility does not depend on efficiency), N drops out of $\bar{a}_i$ and the effect of N on $\bar{a}_i$ disappears:

<u>Corollary 1</u>. Assuming the N homogeneous player public goods game with C-R social preferences as represented in Eqn. (2-3,5), then the cutoff level between noncooperation and cooperation for an individual agent ($\bar{a}_i$) as defined in Prop 1, exhibits the following comparative statics:

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

b. If $\varepsilon_i = 0$ (no concern for efficiency), then changing the number of players (N) has no effect on the cutoff MPCR level between cooperation and noncooperation ($\bar{a}_i$).

C. Aggregate Provision of the Public Good.

One of the interesting findings from Bergstrom et al (1986) is that there is a critical 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* for each of these individuals). A further result is that an income redistribution among contributors (or among noncontributors) does not change the aggregate provision of the public good. Does this also hold in case of social preferences and the linear model?

Let $P(a)=\{i|a\geq\bar{a}_i\}$. The set P(a) consists of all the agents who will contribute to the public good in a Nash equilibrium (of course P may be empty). The population is divided into noncontributors and contributors who contribute their entire wealth. Clearly the total provision of public goods is given by $G = \Sigma_{i\varepsilon P(a)}\, w_i$. This leads to the following analog of the neutrality result (Theorem 5) in Bergstrom et al (1986):

Corollary 2 (Neutrality). Assume the N homogeneous player noncooperative public goods game with C-R social preferences as represented in Eqn. (2-3,5). Then redistributions of wealth among contributors or among non-contributors will not change the aggregate provision of the public good. Further, a redistribution of wealth from contributors to noncontributors will decrease the aggregate provision of public goods.

This result is of course due to the fact that contributors are contributing all of their wealth (because of linear preferences) and the choice of whether or not to contribute is independent of wealth. Bergstrom et al (1986) obtain the same result though the rationale is somewhat more subtle – private good consumption is unaffected by redistribution so that all of the change in wealth goes into the public good. The result is the same however.

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, $\bar{a}_i$ for an incumbent, i, approaches a limit of 0, provided $\varepsilon_i > 0$. 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. One would expect the marginal utility for an individual of either private consumption or public welfare to diminish with increased levels of either variable, which does not happen in the linear model.[6]

IV. ENDOGENOUS COALITIONS FACILITATING PROVISION

We now consider a slightly more complicated institution. We allow a subset of agents to endogenously form a coalition for the express purpose of coordinating contributions to the public good. This is in the spirit of the groups explored by Olson (1971). 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 an allocation

[6] If Charness and Rabin (2002) had represented a concern for efficiency differently, in terms of average per capita egoistic utility rather than aggregate egoistic utility, then the result here may not not hold.

wherein it is not in any agent's individual interest to unilaterially change behavior. Some agents find themselves in the coalition in equilibrium (and some not), with no incentive to unilaterally defect. It is not relevant what path an agent took to find him or herself in the coalition or in the fringe (or even, in fact, if such a path exists).

As is standard in the literature on cartels and international environmental agreements, 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 within the coalition. 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} and the size of the coalition 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 game, 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 (i) of the fringe to contribute everything or nothing to the public good, depending on the value of $\bar{a}_i$ relative to a:

Contribute w_i if $a > \bar{a}_i$ (10a)

Contribute 0 if $a < \bar{a}_i$ (10b)

where $\bar{a}_i$ is defined by Eqn 9c.

Pf: Identical to proof of Prop. 1 □

Without loss of generality, we assume that in the case of ties ($a = \bar{a}_i$), agents contribute fully. We can similarly examine the incentives of the coalition, though we need to slightly restrict the minimum payoff in the C-R preferences. As defined, the minimum is over all agents in the economy. However, the coalition will be aggregating utility over members of the coalition. Thus it makes more sense (and is more tractable mathematically) to view equity from the point of view of the coalition with respect to the minimum payoff agent *outside of the coalition*. It is unlikely that this is a significant restriction:

Prop. 2. Assume 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 of coalition members are with respect to the minimum payoff of agents outside the coalition. Then, conditional on the size of the coalition being n, with coalition members indexed by C, it is a dominant strategy for the coalition to contribute G_C:

$$G_C = \Sigma_{k\epsilon R}\, w_k, \text{ where } R = \{k\epsilon C \mid \lambda_k \leq na\,[1+\varepsilon_{avg}(N-1)]\}, \tag{11a}$$

Which implies:

$$\text{(a)} \quad \text{if } \lambda_{max} \leq na\,[1+\varepsilon_{avg}(N-1)] \quad \text{then } G_C = W_C \tag{11b}$$

$$\text{(b)} \quad \text{if } \lambda_{min} > na\,[1+\varepsilon_{avg}(N-1)] \quad \text{then } G_C = 0 \tag{11c}$$

where

$$\lambda_{min} = \min(\lambda_k | k \in C\}; \quad \lambda_{max} = \max(\lambda_k | k \in C\}; \quad \varepsilon_{avg} = [\Sigma_{k \in C}\, \varepsilon_k]/n \tag{11d}$$

Pf: The aggregate utility for the members of the coalition when individual contributions are g_k for $k\varepsilon C$ is

$$\Pi_C(\mathbf{g}) = \Sigma_{k\varepsilon C}\{\lambda_k(w_k - g_k + a(G_C+G_F)) + \delta_k(w_m - g_m + a(G_C+G_F)) + \varepsilon_k \Sigma_i(w_i - g_i + a(G_C+G_F))\}$$

$$= \Sigma_{k\varepsilon C}\{\lambda_k(w_k - g_k + a(G_C+G_F))\} + (w_m - g_m + a(G_C+G_F))\,\Sigma_{k\varepsilon C}\,\delta_k + n\varepsilon_{avg}\,[W + a(N-1)\,(G_C+G_F)] \tag{12}$$

and thus the payoff for the coalition contributing nothing is

$$\Pi_C(\mathbf{0}) = \Sigma_{k\varepsilon C}\{\lambda_k(w_k+aG_F) + (w_m-g_m+aG_F)\,\Sigma_{k\varepsilon C}\,\delta_k + n\varepsilon_{avg}[W+a(N-1)\,G_F] \qquad (13)$$

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

$$\Delta(\mathbf{g}) = -\Sigma_{k\varepsilon C}\,\lambda_k g_k + a\,n\,G_C[1 + \varepsilon_{avg}(N-1)] \qquad (14)$$

Clearly the rhs of Eqn. (14) is zero when $G_C=0$ (at $\mathbf{g}=\mathbf{0}$). Payoffs will increase for contributions from any k for which $d\Delta/dg_k > 0$:

$$d\Delta/dg_k = -\lambda_k + a\,n\,[1 + \varepsilon_{avg}(N-1)] > 0 \qquad (15a)$$

$$\Leftrightarrow \lambda_k < a\,n[1 + \varepsilon_{avg}(N-1)] \qquad (15b)$$

$$\Leftrightarrow n > \{\lambda_k/[1 + \varepsilon_{avg}(N-1)]\}/a \leq 1/a \qquad (15c)$$

Thus for any k for which Eqn. (15b) holds, net payoffs will be increased by contributing, implying that $g_k=w_k$. For any k for which Eqn. (15b) fails to hold (leaving aside cases of equality), contributions will only reduce welfare; thus $g_k=0$. This completes the proof. □

Note in Prop. 2 that the result is conditional on the size of the coalition, n. Note also that if standard preferences apply ($\lambda=1$, $\varepsilon_{avg}=0$), then Eqn (15c) simplifies to $n>1/a$, which is the standard result in the endogenous coalition formation literature. Also note from Eqn (15c) that social preferences reduce the lower bound on n, since the term in braces in Eqn. 15c is less than or equal to unity.

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 associated with 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

Furthermore, define the function ceil(x) as mapping x into the smallest integer greater than or equal to x.

These lead to the following corollary to Prop.2 on the size of the Nash coalition.

Corollary 3. Assume C-R preferences such that a non-cooperative equilibrium yields no contributions to the public good. Consider a two-stage public goods game with coalitions, a coalition, C, of size ñ (the number of agents in the coalition), with all coalition members contributing something to the public good in equilibrium. This implies

$$\tilde{n} \geq \text{ceil}\{\lambda_k / [1 + \varepsilon_{avg}(N-1)]/a \mid k \varepsilon C\} \tag{16}$$

where ε_{avg} is defined in Eqn. 11d and.

Pf: The proof follows from a direction application of Prop. 2.□

Nash equilibrium in the membership game is easy to define but hard to connect to fundamental characteristics of the game. In particular, a coalition C is a Nash equilibrium of the membership game if (a) the payoff attained by any member of the coalition, d, inside the coalition is as great as that agent can obtain outside the coalition when the coalition is reduce to C-{d}; and (b) no member of the fringe, f, outside the coalition can do better by joining the coalition, making it C U {f}. This is simply the definition of a Nash equilibrium. We have not been able to compute the size of an equilibrium coalition from the fundamental charaterteristics of the problem ($\lambda_, \varepsilon_, a, w$). It is more constructive to restrict preferences somewhat, which we do in the next section and then derive general results.

V. IDENTICAL PREFERENCES, HETEROGENEOUS WEALTH

In the previous sections, we developed theory for the general case of heterogeneous agents with heterogeneous wealth levels. Although it is possible to obtain conditions charcterizing an equilibrium in the coalition game, clarity is not well served. It is common (eg, see Bergstrom et al, 1986) to assume homogeneous preferences and let heterogeneity be manifest through different levels of wealth. We consider that special case here. Thus we assume all agents share the same λ, δ and ε, though have different wealth levels, w_i.

These lead to the following proposition on the size of the Nash coalition.

<u>Prop. 3.</u> Assume homogeneous 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 (and with no side-payments within the coalition), if an equilibrium coalition exists, providing positive amounts of public goods, it will be of size $\tilde{n}$,

$$\tilde{n} = \text{ceil}[1/\bar{e}], \tag{17a}$$

$$\text{where } \bar{e} \equiv \{a+ \varepsilon[a(N-1)-1]\}/\lambda \tag{17b}$$

Furthermore, for any such coalition C, of size $\tilde{n}$, if the richest member of the coalition (r) is not too wealthy in the sense of satisfying

$$w_r/W_C \leq \bar{e} \tag{18}$$

then the coalition is a Nash equilibrium.

Pf: From Eqn. (15c), in Prop. 2,

$$n > \{\lambda /[1 + \varepsilon_{avg}(N-1)]\}/a \tag{19}$$

which, for homogeneous preferences, equivalent to

$$n < \lambda/\{a+\varepsilon[a(N-1)-1]\} \tag{20}$$

since $\varepsilon_{avg} = \varepsilon$. Thus the first part of the Proposition follows directly from Prop. 2.

All that remains to show is the condition for internal stability for coalitions of size $n = \tilde{n}$. We examine the incentives for an arbitrary 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

$$u_C = \lambda a W_C + \delta(w_m+aW_C)+\varepsilon[W+W_C(aN-1)] \tag{21}$$

The payoff in the fringe involves no contributions to the public good by anyone:

$$u_F = \lambda w_d + \delta w_m+ \varepsilon W \tag{22}$$

which implies

$$\begin{aligned}\Delta = u_C - u_F &= \lambda a W_C + \delta(w_m+aW_C)+\varepsilon[W+W_C(aN-1)] - \lambda w_d - \delta w_m - \varepsilon W \\ &= \lambda(aW_C-w_d) + \delta a W_C + \varepsilon W_C(aN-1) \\ &= -\lambda w_d + W_C\{\lambda a + \delta a + \varepsilon(aN-1)\} \qquad (23)\end{aligned}$$

which implies

$$\Delta > 0 \quad \text{iff} \quad w_d/W_C < \bar{e} = \{\lambda a + \delta a + \varepsilon(aN-1)\}/\lambda \tag{24a}$$

$$= 1+ \{\lambda(a-1) + \delta a + \varepsilon(aN-1)\}/\lambda \tag{24b}$$

The right hand side of Eqn. (24a) is positive and the portion in braces in Eqn (24b) 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. 24 must hold for all members of the coalition, which is satisfied if it holds for the richest member of the coalition.□

Prop. 3 defines the size of the minimal coalition size for which contributing is collectively rational. This is an extension of the result on standard preferences, which states that the size of a minimal contributing coalition is ceil(1/a). Social preferences make that size smaller ($\bar{e} \geq a$). Furthermore, the proposition states that if there is too much wealth inequality among members of the coalition then the incentives for the richest member to defect to the fringe are too strong to hold the coalition together.

Interestingly, the window of wealth disparities that does not destroy the equilibrium is quite narrow. We know $\tilde{n} = \text{ceil}(1/\bar{e})$ which implies $\tilde{n} \geq 1/\bar{e}$, which implies $\bar{e} \geq 1/\tilde{n}$. But there is precious little space between $w_r/W_C = 1/\tilde{n}$ (perfect wealth equality) and $w_r/W_C = \bar{e}$. This suggests that very little wealth inequality would be tolerated in an equilibrium.

Note that for standard preferences which are not other-regarding (λ=0), $\bar{e}$=a and $\tilde{n}$=ceil(1/a).[7] When all agents have the same initial endowment, $w_r/W_C = 1/\tilde{n} \leq a$. 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.

This suggests that “sharing the wealth” within the coalition might be appropriate. However, wealth transfers among members of the coalition is often suggested as being politically infeasible. But if the gains are divided up with the goal of equalizing the incentives to defect, then stability can be achieved, even with significant wealth disparities:

<u>Prop. 4</u>. Assume homogeneous 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, with transfers among coalition members, any coalition (C), which providing positive amounts of public goods, and is ofsize $\tilde{n}$,

[7] This is a standard result in the international environmental agreements literature. As was discussed earlier, IEA models are somewhat more restrictive (eg, homogeneous endowments and binary choice). See Barrett (2003) and Ulph (2004).

$\tilde{n}$ = ceil[1/ $\bar{e}$], (25a)

where $\bar{e} \equiv \{a + \varepsilon[a(N-1)-1]\}/\lambda$ (25b)

is a Nash equilibrium in the two stage game.

Pf: Let each agent in the coalition be taxed at the amount t, equal to the gain from staying in the coalition, as defined by Eqn. (23). A positive tax is a transfer from the agent to the coalition and a negative tax is a transfer to the agent. Clearly as defined, such a tax neutralizes incentives to defect, as shown in the proof for Prop. 3. All that remains is to show that the sum over all taxes collected (and disbursed) within the coalition is non-negative (ie, there is enough surplus to support the transfers).

As described above, let the tax for agent d be defined by

$$t_d = -\lambda w_d + W_C\{\lambda a + \delta a + \varepsilon(aN-1)\} \quad (26)$$

which implies

$$\Sigma_{d\varepsilon C}\, t_d = -\lambda W_C + nW_C\{\lambda a + \delta a + \varepsilon(aN-1)\} \quad (27)$$

$$= W_C\,\{-\lambda + n(a + \varepsilon[a(N-1)-1])\} \quad (28)$$

By the assumption in Eqn. (25), the term in braces in Eqn. (28) is positive. □

In essence, there is enough surplus among members of the coalition to equalize the incentives to defect among coalition members.

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). 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:

<u>Lemma 2</u>. Assuming homogeneous C-R preferences as in Prop. 4, with 0 < λ < 1, then

(1) $\tilde{n}$ is non-decreasing in λ, holding ε constant;

(2) $\tilde{n}$ is non-decreasing in λ, holding δ constant;

(3) $\tilde{n}$ is non-increasing in ε, holding δ constant;

and (4) $\tilde{n}$ is non-increasing in ε, holding λ constant.

Pf: Combine Eqn (17a) and Eqn (17b) and then totally differentiate, ignoring the ceil function. Further, totally differentiate the identity in Eqn (8). Solve for $d\tilde{n}/d\lambda$, $d\tilde{n}/d\varepsilon$, or $d\tilde{n}/d\delta$, holding the third parameter fixed proves the lemma. The role of ceil is to moderate the signs of the derivatives (changing > to ≥ and < to ≤). □

This Lemmas indicates that introducing C-R social preferences (lowering λ), keeping endowments identical among agents, tends to lower the size of stable coalitions. Introducing heterogeneity of agents does not change the size of coalitions but suggests that even modest wealth disparity can cause a coalition to be unstable (Prop. 3) unless transfers are involved (Prop. 4).

V DISCUSSION

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 2: 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.

It is appropriate to compare these results with those of the simple public goods game without coalition formation but with heterogeneous agents, some of whom wish to unilaterally contribute and some of whom do not. In this case, there will be a subgroup of contributors and a subgroup of noncontributors. The contributors may appear to be a coalition of agents with the same agenda; however the theory really only suggests that individuals in the group will individually contribute. For example, many argue that the Montreal Protocol for protecting stratospheric ozone is a good example of a coalition of nations to provide abatement. However, Murdoch and Sandler (1997) have suggested that the Montreal Protocol is largely an association of countries which find reducing emissions individually rational.

VI AN EMPIRICAL EXAMPLE

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 an example, with parameters of the C-R preferences drawn from a laboratory experiment. The experiment is of a standard public goods game (as in section III of this paper). The experiment allows us to estimate the parameters of social preferences, under some simplifying assumptions. It is important to underscore that this is an example and not in any way a test of the theory, since we make no claim that this particular laboratory experiment is generalizable. The value of the example is to demonstrate the extent to which cooperation may be enhanced in a social preferences setting by permitting the formation of coalitions.

In this example, we are interested in several testable questions based on the experimental data: (a) can we reject the null hypothesis of self-interested actors vs. the alternative hypothesis of social preferences with an altruistic term? (b) is the level of voluntary contributions in the experimental setting consistent with theoretical predictions, given the social preferences revealed in the experiment? and (c) given the social preferences statistically inferred in the experiment, to what extent can allowing a coalition of agents increase the aggregate provision to the public good?

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. We are grateful to KOR making their data available to us, through the *American Economic Review*'s policy on archiving data from published empirical papers.[8] Although the data are not perfect as a test of our model, they are useful as an example, for estimating some of the parameters of C-R preferences for the subjects in their experiments.

In the KOR 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 experimental outcomes to the individual players.[9] Each group of four played 20 rounds with the same set of players (no random rematching between rounds).

Because of the similarity between their pure public goods games and the theoretical structure of the pure public goods game in this paper (section III), we can use the results of the public goods game to estimate some of the values of the parameters of C-R preferences.

---

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

[9] The other experiments reported in 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 ("implementation") wherein the coalition votes on whether to contribute or not (contributing requires unanimity); and (b) there is a fixed cost of implementing the coalition (there are none in our model). In all of their experiments, agents have the same endowment. We are unable to use their coalition experiments here because of the different institutional structure.

Because the players in their experiments are only told the aggregate contributions, and not the individual contributions, it is difficult or impossible to econometrically identify the distributional parameter (δ); we thus set it to zero, focusing on the parameter for egoistic payoffs (λ) and its complement, the parameter on social efficiency (1-λ). Thus the C-R preferences we estimate involve utility depending on a weighted sum of private benefits and group benefits.[10]

A. Data

Individuals play 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. The implicit assumption is that after 10 or 15 rounds, players should have "settled in" to a Nash equilibrium. There are 10 sessions of four for the PG40 game and 9 sessions of four for the PG65 game. In other words, there are four individuals to a group, there are 20 rounds for each group within a session, there are ten sessions for one treatment (PG40) and nine sessions for the other treatment. 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.

One potential problem is that the players were not randomly rematched every round, as is sometimes done in repetitions (eg, Fischbacher and Gächter, 2010). Thus there is the potential for strategic play. This is not considered a significant issue here for several reasons. In comparing the case of MPCR = 0.40 with a simple game in Fischbacher and Gächter (2010), results are qualitatively similar, suggesting that randomizing players is not a major issue. Secondly, we focus on only one round (round 15) of the twenty rounds, ignoring differences among rounds. Third, this is only an example, not a test of theory.

[10] Admittedly, setting δ to zero is less than ideal. We did attempt to estimate the full model with a distributional term. We were unable to obtain convergence. Since this is an example and not a rigorous test of theory, this approximation would appear acceptable.

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 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 standard deviation 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).

B. Analysis

Since we are interested in characteristics of the Nash equilibrium, not the path to equilibrium, we focus on a single round for each player. Somewhat arbitrarily, we chose round 15; one would expect similar results choosing another late round. This results in 76 different observations from KOR, from which we estimate two versions of C-R utility, Eqn. (2-3,5). 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.[11] 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

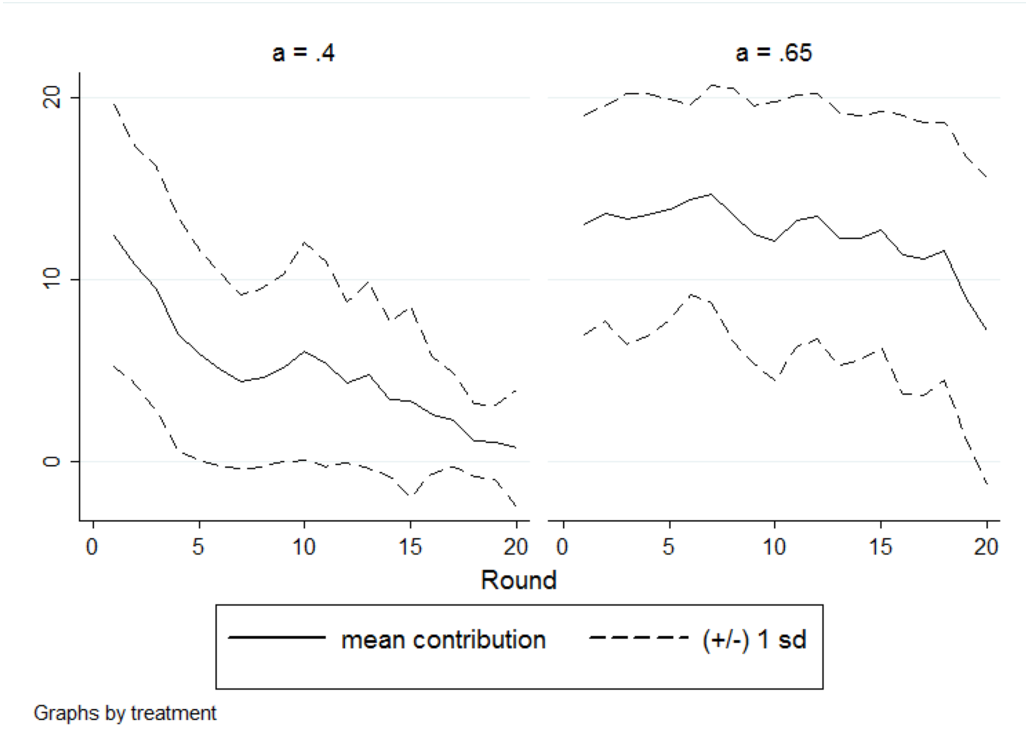
[11] KOR suggest in their conclusions that issues of fairness appear to be driving the tendency to form larger coalitions, though without utilizing F-S preferences. Although participants in the KOR experiments may have F-S preferences based on expectations of behavior of others, the fact is that they do not know the behavior of others. KOR only provide participants with information on their own payoff, the number of people in the coalition and the aggregate payoff. F-S preferences are clearer when participants know 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.

participants lognormally (Case II) in a random coefficients framework. 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,[12] 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 as 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 multinomial logit (conditional 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). The reason for using multinominal logit is that we wish to reflect the fact that chosen levels of contribution must give more utility than the non-chosen levels of contributions (by revealed preference). In such cases, either one must estimate a continuous model with a side restriction of the first order conditions for the choice problem or use discrete choice methods which explicitly reflect this optimizing behavior.

[12] As mentioned earlier, 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.

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



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 ε is restricted to be zero. Note that the mean value of λ in the random coefficients is close but slightly lower than for the fixed coefficient logit. A likelihood ratio test of the hypothesis that preferences are purely self-interest against the alternative hypothesis of Case I (ie, ε=0) is clearly rejected.

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 the distribution of the random coefficients; Figure 4a converts that into a pdf over ā, using Eqn. 7c. Figure 4b shows the associated cumulative density function for ā. Recall that ā is the minimum MPCR (a) for which it is individually rational for agents to contribute to the public good and that for standard preferences, ā=1.

Table I: Estimated Coefficient from KOR Data

	Reference	Case I	Case II
λ	1	0.767	
Std error		0.02	
mean(λ)			0.715
Std error on mean			0.036
Observations	380	380	380
Log likelihood	-113	-97	-94

Note: Reference is standard preferences with ϵ restricted to be zero; estimation is of λ, which is freely allowed to vary and then normalized to unity. Case I is a conditional logit, with λ and ε freely allowed to vary and then normalized to sum to unity; Case II is a random coefficients logit (mixed logit), with normalization following the estimation. For Case II, the mean and variance of the lognormal distribution are given, reflecting the dispersion over the population. Number of observations reflects 1520 data points with five possible choices for each data point.

Figure 3: Estimated density for ϵ = 1-λ in mixed logit (random coefficients)

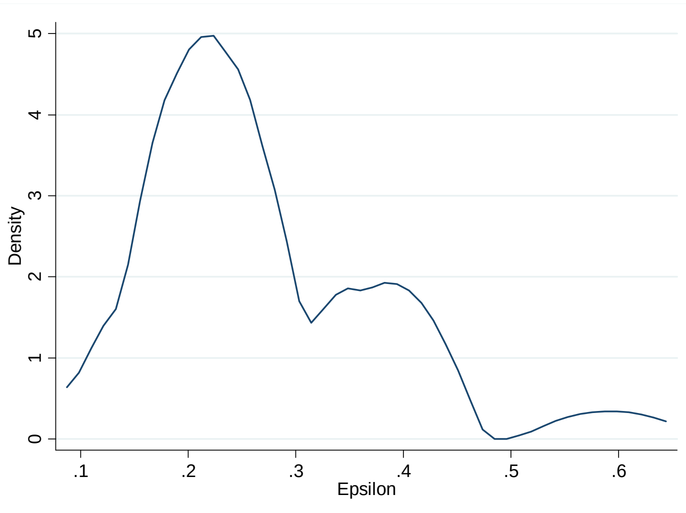
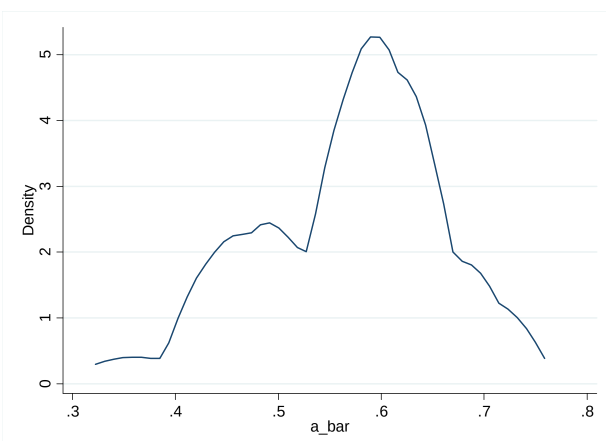
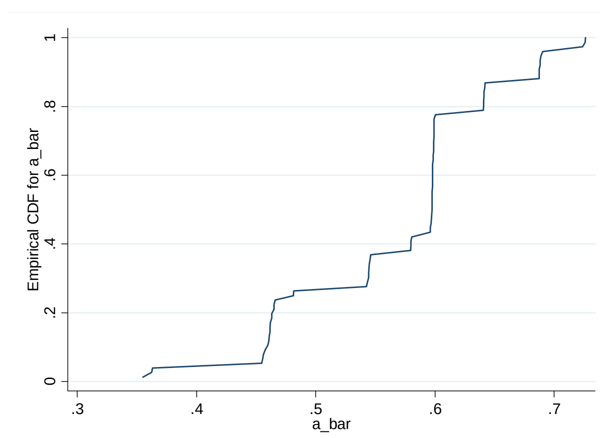


Figure 4: PDF (a) and CDF (b) for ā induced by density on λ in mixed logit.

(a)



(b)



C. Implications

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, *should a coalition have been allowed to form*, 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 (from theory—Eqn 9c & 20) of ā and ñ, using data from Table I

	ā	ñ
Standard Preferences (λ=1)		
MPCR = 0.4	1	3
MPCR = 0.65	1	2
C-R Preferences (λ=0.749)		
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.[13]

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

[13] In the KOR experiments, there are also coalition treatments with these same MPCRs. Although in most sessions contributing coalitions do not form, average contributions are higher with the higher MPCR. Because of the differences in how KOR treat coalitions, a direct comparison with our computations is difficult.

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 ($\tilde{n}$), 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.

We next 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. With 80% of the agents finding it individually rational to contribute to the public good, observationally it may appear that a large coalition supports contributing. But this will occur in the absence of any provision for the formation of coalitions.

Returning to the three questions posed at the beginning of this section, we can reject the hypothesis of pure self-interest; social preferences are much more consistent with the experimental data. We are also able to conclude that the aggregate level of contribution to public goods is roughly consistent with theoretical predictions (though this is not a statistical conclusion). Whether allowing the formation of a coordinating coalition increases overall public goods contributions is ambiguous. As Table II indicates, for an MPCR of 0.65, no coalition will form because contributing is individually rational for most.

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 collectively rational for the coalition to provide the public good. By implication, the smallest collectively rational contributing coalition is smaller than it would be with standard preferences. 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 incentives for defection from the coalition are sensitive to wealth – more wealthy members of a coalition have a stronger incentive to defect. When coalitions members have similar wealth the coalition is more likely to be stable than when the distribution of wealth is more unequal. However, we are also able to show that there exists a set of transfers among members of the coalition that can equalize the incentives to defect among coalition members; further, that the gains from cooperation within the coalition are sufficient to finance these transfers. This result has not been reported elsewhere in the literature, in my knowledge.

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. Further, we can examined a counterfactual using these data – what behavior might one expect from coalitions with the social preferences inferred from the experiments. These results are only illustrative but they do indicate that the theory presented here is empirically implementable.

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).

Becker, Gary, "A Theory of Social Interactions," *J. Pol. Econ.*, **82**:1063-93 (1974).

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).

Diamantoudi, E. and E.S. Sartzetakis, “Stable International Environmental Agreements: An Analytical Approach,” *J. Public Econ. Theory*, **8**:247-63 (2006).

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).

Fischbacher, Urs, Simon Gächter, and Ernst Fehr, "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment," *Econ. Letters*, **71**:397-404 (2001).

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).

Olson, Mancur, *The Logic of Collective Action*, 2nd Ed. (Harvard University Press, Cambridge, Mass., 1971).

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

Rubio, S. J. and A. Ulph, “Self-Enforcing International Environmental Agreements Revisited,” *Oxford Econ. Papers*, **58**:233-63 (2006).

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

Smith, Vernon L., “Experiments with a Decentralized Mechanism for Public Goods Decisions,” *Amer. Econ. Rev.*, **70**:584-99 (1980).

Sobel, Joel, “Independent Preferences and Reciprocity,” *J. Econ. Lit.*, **43**:392-436 (2005).

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