Leading by Example and International Collective Action

ANDRÉ ROSSI DE OLIVEIRA
University of Brasilia

JOÃO RICARDO FARIA
University of Texas at Dallas

DANIEL G. ARCE M.
Rhodes College

Abstract

This paper investigates leading by example as a policy prescription for international collective action to provide summation public goods. A country leads by example by committing to a minimal level of provision, and by matching higher contributions there beyond. In an evolutionary game-theoretic setting, we establish conditions for leading by example to be a neutrally stable strategy; i.e., to noncooperatively implement the cooperative outcome. These conditions are related to the degree of concavity of the contributors' utility functions and the incentives for free riding. They can be tested against empirical estimates of the public benefits of an international regime.

1. Introduction

Foreign aid, international treaties, and environmental protocols are increasingly being viewed through the lense of Olson's (1965) classic, *The Logic of Collective Action* (henceforth called *Logic*). For example, Sandler (1997) provides a survey of transnational public goods, identifying how successful treaties and protocols are tailored to recognize the three characteristics of

André Rossi de Oliveira, Department of Economics, University of Brasilia, Campus Darcy Ribeiro, ICC Norte, Brasilia – D.F. 70910-900, Brazil (arossi@unb.br). João Ricardo Faria, School of Social Sciences, University of Texas at Dallas, 1601 North Floyd Rd., Richardson, TX 75083-0688 (jocka@utdallas.edu). Daniel G. Arce M., Robert D. McCallum Professor of Economics & Business, Rhodes College, 2000 North Parkway, Memphis, TN 38112 (arce@rhodes.edu).

Arce gratefully acknowledges research support from the Robert D. McCallum endowment. We thank three anonymous referees for helpful comments.

Received September 2002; Accepted March 2003.

Journal of Public Economic Theory, 7 (1), 2005, pp. 51–63.

publicness: nonexclusivity, nonrivalry, and the underlying aggregation technology. He asserts that the differences in publicness explain in part why the Montreal Protocol on CFCs has been adopted almost globally, while the fate of the Kyoto Protocol on greenhouse gas is less clear, even though the languages of the protocols are strikingly similar.

In the same way, the vast majority of investigations into the theoretical requirements for successful international collective action are based upon tacit cooperation in the iterated Prisoner's Dilemma. Yet, in an international context, actors lack the type of property rights and power that is commensurate with punishment-based strategies that implement cooperation. Further, in many contexts retaliation may not even make sense. For environmental public goods it implies increased pollution or decreased abatement (Sandler and Arce 2003). By contrast, in an initial review of *Logic,* Wagner (1965) suggested that *leaders* who recognize the potential for welfare improvements may undertake convincing actions that cause others to act collectively toward their mutual benefit without the need for punishment, pressure groups, or rent-seeking. Similarly, Kindleberger (1981) opines that hegemonic power is not necessary for successful leadership in providing international public goods; instead, groups of smaller nations can successfully serve as examples through policies of foreign aid, peacekeeping, and active participation in the United Nations initiatives.

In examining the implications of leadership for the provision of international public goods, Arce (2001) and the articles contained in Tuchman Matthews (1991) provide several illustrations of the success of leading by example. A country that leads by example commits to a minimal level of contribution to the provision of the public good, and to match higher contributions there beyond. In an evolutionary game-theoretic environment, leading by example allows other agents to learn and eventually imitate policies that are necessary for successful collective action. For several numeric examples, leading by example is shown to generate welfare improvements for different types of public goods (Arce 2001). The result is novel because leading by example is a noncooperative means for implementing a cooperative solution without the need for retaliation. Our purpose is to generalize this result for the case of summation public goods.[1]

Two important predecessors to leading by example are Guttman (1978) and Sugden (1984), where matching and reciprocity act as constraints on the voluntary contribution decision. For example, Guttman employs a two-stage game where the optimal contribution is characterized for any level of matching behavior, and then he works backward using an iterated dominance argument to find the optimal level of matching. Sugden assumes that there exists a group for which the reciprocity norm holds. By contrast, leading by

[1]As a caveat, we do not consider the Prisoner's Dilemma; unilateral efforts are rarely successful in such an environment.

example does not require iterated dominance, nor any form of rationality based on best responses in order to promote collective action. Further, leaders are not pure reciprocators; they are also committing themselves away from equilibria in which they may be exploited and/or the voluntary level of provision is suboptimal. Leading by example is an effective coordinating device for creating an environment in which others learn the benefits of collective action.

The analysis proceeds as follows. Section 2 presents a general game-theoretic framework for the analysis of international public goods regimes. Section 3 contains a brief overview of evolutionary equilibrium concepts. Section 4 presents the leading by example strategy and our main results. It is found that the success of leading by example is highly tied to the concavity of contributors' utility functions. We relate this result to concrete examples of leading by example. Section 5 offers brief conclusions.

2. The Model

The need for international collective action arises because of the public nature of transboundary externalities. By *public* we mean that a good (or bad) is *nonexclusive*—the benefit or negative externality it endows cannot be limited to the individual purchaser or producer, and *nonrival*—an individual's consumption of the public good or exposure to the public bad is not diminished in any way by another individual's consumption or exposure. It is the public nature of goods such as clean air, ozone protection, etc., that can cause them to be underprovided in the absence of collective action. Similarly, public "bads" such as acid rain, sewage discharge into waterways, etc., are overprovided in absence of a regime of checks and balances. Our interest is in how international agents voluntarily enter into agreements for the provision of international public goods, and how this compares with the level of provision that maximizes societal welfare.

Let there be two agents and let q_i be agent i's contribution to a public good. We will relax the assumption of two agents shortly. The set S_i is agent i's strategy space; hence, $q_i \in S_i$. Agent i's preferences are represented by the payoff function $V_i : \Re \to \Re$ defined as

$$V_i(q_1, q_2) = U(Q(q_1, q_2)) - cq_i, \tag{1}$$

where $q_i \in S_i$; $i \in \{1, 2\}$; $Q : S_1 \times S_2 \to \Re_+$ is a function that determines the amount of the public good provided given individual contributions (the *aggregation technology*), and $U : \Re_+ \to \Re$ is a strictly concave and strictly increasing function where $U(0) = 0$. An agent derives utility from the provision of the public good, which depends on the contributions of all agents and is captured by the term $U(Q(q_1, q_2))$. In keeping with Olson's (1965) analysis of the free-rider problem, the quantity provided of the public good is equal to the *sum* of all individual contributions: $Q(q_1, q_2) = q_1 + q_2$. Finally, there is an individual cost to provision, which is represented by the linear cost function $C(q_i) = cq_i$.

When $Q(q_1, q_2)$ is a summation aggregator the public benefit does not depend upon the distribution of individual contributions. The potential for free riding is great because each agent's contribution is a perfect substitute for another's. The usual assumption is that summation leads to a Prisoner's Dilemma (hereafter, PD), making free riding endemic. A substantial literature exists around this case. Social optima are achieved in a dynamic (typically, repeated) framework where there is a threat of punishment and the shadow of the future is large. There are instances, however, when summation does not lead to the PD and the requirements for cooperation in this case are less well known. For example, the amount of sulfur in the atmosphere is determined by the sum of emissions. Yet, a transportation matrix of acid rainfall reveals that it is concentrated relatively close to the source of emissions, and this creates individual interest in abatement (Sandler 1998). Alternatively, Heckathorn (1996) develops a taxonomy of voluntary provision games around the summation aggregator where the PD is but one of several underlying game forms. He shows that diminishing marginal utility in the consumption of a public good is decisive in determining whether incentives exist for voluntary provision in binary choice models. Our purpose is to examine the *noncoercive* requirements for optimal collective action in non-PD cases of voluntary provision.

In addition, we employ a generalized (rather than numeric) representation of payoffs. Specifically, the image of the function U is assumed to be strictly increasing and to satisfy[2]

$$U(s+1) - U(s) < U(s) - U(s-1) \quad \forall s \in \{1, 2, 3\}. \tag{2}$$

This condition can be interpreted as a diminishing marginal utility or diminishing returns property. Finally, as we are dealing with the summation aggregator, we can express utility from the public good as a function of the sum of contributions:

$$\mathrm{U}(Q(q_1, q_2)) = f(q_1 + q_2). \tag{3}$$

This implies that payoffs in the voluntary contribution game can be expressed as

$$V_i(q_1, q_2) = f(q_1 + q_2) - cq_i. \tag{4}$$

The voluntary contribution game for $S_i \equiv \{0, 1, 2\}$ is given in Box 1. Further, our analysis of this game hinges on two key assumptions about its structure:

(A1) The game is not the Prisoner's Dilemma.
(A2) $(2, 2) \in \arg\max_{q_1, q_2} V_1(q_1, q_2) + V_2(q_1, q_2)$.

---

[2]This condition is the analogue to the requirement that the second derivative of the function U be negative (in the case where U is differentiable), and can be construed as a discrete version of concavity. Recall that $U(0) = 0$.

Box 1: Pairwise matchings of collective action policies

↓**policies (q)**→	**0**	**1**	**2**
0	$f(0), f(0)$	$f(1), f(1) - c$	$f(2), f(2) - 2c$
1	$f(1) - c, f(1)$	$f(2) - c, f(2) - c$	$f(3) - c, f(3) - 2c$
2	$f(2) - 2c, f(2)$	$f(3) - 2c, f(3) - c$	$f(4) - 2c, f(4) - 2c$

One way to satisfy assumption (A1) is to require that the binary 2 × 2 game (in bold) that is nested within Box 1—where $q_i \in \{0, 1\}$—differs from the PD in some way. On a more general level, the requirement is that the game does not have a dominant strategy equilibrium that is Pareto-dominated by the social optimum. Assumption (A2) is due to the fact that leading by example is a strategy ultimately designed for regimes that promote uniform behavior to reach the social optimum. As (2, 2) is the upper bound on uniform contributions in Box 1, this is why we assume that (2, 2) maximizes $V_1 + V_2$. It is not required that (2, 2) be the unique maximand of $V_1 + V_2$.

Several important details of international treaties are embodied within this model. First, the game is not one of binary choice between no provision ($q_i = 0$) and provision ($q_i = 1$).[3] This is because international collective action has been shown to be strategically complex. In particular, there is concern over the phenomenon of *Nash codification* (Murdoch and Sandler 1997, Murdoch, Sandler, and Sargent 1997). Nash codification refers to a situation in which signatories to a treaty merely codify the actions they would take on their own (i.e., a Nash equilibrium). Under such circumstances there may be substantial room for improvement beyond what is specified by the treaty. Enlarging the strategy space from two to three levels of provision is the minimal extension necessary to capture the possibility that agents are voluntarily contributing, but underproviding.

Second, within this context the number of policies under consideration remains discrete. To foreshadow, our policy selection criterion—evolutionary stability—encapsulates *boundedly rational* behavior in that nations need not start out at the equilibrium outcome, but learn and imitate successful policies that they observe and try over time. It follows that if policymakers are boundedly rational, then it is unlikely that they will contemplate or compute an infinite array of alternatives. Deliberations are typically over a discrete set of alternatives.

Third, the situation is symmetric. In an international context all countries have a symmetric set of property rights. Further, it is not out of the question to assume that the benefits derived from clean air, ozone protection, etc., are symmetric.

[3]Palfrey and Rosenthal (1984) and Cavaliere (2001) examine noncoercive strategies for achieving cooperation in *binary choice* models where the number of players varies.

3. Equilibria

Problems requiring international collective action are rarely single-shot. Further, the benefits produced are often designed to be passed on to future generations (Sandler 1978). The concept of evolutionarily stable strategies (Maynard Smith and Price 1973) is designed to capture the notion of a strategy that is robust to selection pressures over time and generations. More specifically, it considers a situation where individuals are repeatedly and randomly drawn from a large *population* to play a symmetric two-player game. In this context, our assumption of a two-player game is immediately relaxed; instead, the model represents the outcome of pairwise interaction of potential parties to an international treaty. Pairwise interaction means that the *strategies/policies* are in long-term competition, and not the individual players. Any pairing of two strategies in this context thereby identifies the effectiveness/fitness of one policy when confronted with another. A policy is in equilibrium when it characterizes the population in question, and satisfies certain stability conditions. Specifically, an equilibrium policy, q_i, ought to be robust against a "mutant" policy, q_i^m. The incumbent policy is said to be evolutionarily stable if it is able to fend off the adoption of the mutant policy by a small proportion of the population.

Formally, we define a *symmetric* two-player game as one that satisfies both (i) $S_1 = S_2 = S$ and (ii) $V_1(q, q^m) = V_2(q^m, q)$ for all $q, q^m \in S$. Such a game is meant to characterize the outcome of pairwise matchings of policies considered by a large population. Condition (i) implies that each player has the same policies at his disposal, and (ii) means that everyone evaluates these policies similarly. Note that symmetry implies that S now denotes the generic individual strategy set, and q an element of that set. The payoff $V(q, q^m)$ is that earned by a q-policy in a pairwise matching with a q^m-policy. Strategy $q \in S$ is an *evolutionarily stable strategy* (ESS) if and only if it satisfies the following first- and second-order conditions (Maynard Smith and Price 1973):

$$\begin{aligned} &(\textbf{FOC}) \quad V(q, q) \geq V(q^m, q)\, \forall\, q^m \in S; \\ &(\textbf{SOC}) \quad \text{If } V(q, q) = V(q^m, q) \text{ then } V(q, q^m) > V(q^m, q^m)\, \forall\, q^m \neq q. \end{aligned} \tag{5}$$

The FOC indicates that an ESS is a best reply against itself. Therefore, every ESS profile (q, q) is a symmetric Nash equilibrium. This means that we can check the main diagonal of any symmetric game to find whether a pure strategy ESS exists. Further, the SOC is an additional stability requirement, thereby implying that ESS is a refinement of Nash equilibrium. It captures the idea that if a mutant strategy (q^m) does equally well against the population strategy (q), the mutant cannot survive and thrive within the population because the population strategy does better in a pairwise matching with the mutant than the mutant does in a pairwise matching with itself.

There is a "natural selection" aspect to the concept of evolutionary stability because it is explicitly linked to the selection criterion known as the *replicator dynamic*. Specifically, if σ_q is the proportion of the population currently contributing at level q, and σ is the vector of population proportions for all contribution levels $q \in S$, then $\mathrm{E}[q, \sigma]$ is the expected payoff for contributing at level q, and $\mathrm{E}[\sigma, \sigma]$ is the average payoff for the population.[4] Then, the differential equation

$$\dot{\sigma}_q = \sigma_q(\mathrm{E}[q, \sigma] - \mathrm{E}[\sigma, \sigma]) \tag{6}$$

is the replicator dynamic with frequency-dependent payoffs. It encapsulates bounded rationality because a strategy grows so long as it does better than average; i.e., it need not be a best reply. We employ ESS as an equilibrium criterion because strategies that satisfy the static characterization in (5) are asymptotically stable with respect to (6) (Taylor and Jonker 1978).

A weaker stability criterion is neutral stability. Formally, $q \in S$ is a *neutrally stable strategy* (NSS) if the FOC holds and the strict inequality in the SOC for ESS is replaced by a weak inequality (Maynard Smith 1982):

$$(\mathbf{SOC}') \quad \text{If } V(q, q) = V(q^m, q) \text{ then } V(q, q^m) \geq V(q^m, q^m)\ \forall\, q^m \neq q. \tag{7}$$

If (q, q) *and* (q^m, q^m) are socially optimal Nash equilibria, it may be the case that q^m can be used to eliminate q as an ESS (and vice versa) because neither does strictly better against the other. In order to rule this possibility out, we focus on the NSS criterion.[5] Also, in keeping with the international relations literature we will restrict our attention to pure strategy equilibria.

If payoffs represent the fitness of policy pairings over time, then ESS and NSS are appropriate equilibrium criteria. Further, when players imitate strategies based on their observed relative merits over time, the process can be shown to reduce to the replicator dynamic (Harley 1981), whose stability our static equilibrium concepts have been shown to characterize.

Finally, the cooperative solution or welfare criterion employed is the standard benchmark of maximizing the sum of player's payoffs.[6] From an international perspective, this means maximization of societal welfare. Such a benchmark is not affected by sidepayments between parties; hence, it is also a point of comparison for regimes that allow for foreign aid, or financial inducements such as the Multilateral Fund arrangement in the Montreal Protocol.

---

[4]These expected payoffs are calculated as if σ_q is the mixed strategy for contribution level q and σ is the joint mixed strategy for all contribution levels.

[5]NSS is the static characterization of Lyapunov stability of the replicator dynamic (Hofbauer and Sigmund 1998).

[6]Assumption (A2) implies that (2, 2) is a social optimum.

Box 2: The extended policy environment

↓policies (q)→	0	1	2	q^*
0	$f(0), f(0)$	$f(1), f(1) - c$	$f(2), f(2) - 2c$	$f(1), f(1) - c$
1	$f(1) - c, f(1)$	$f(2) - c, f(2) - c$	$f(3) - c, f(3) - 2c$	$f(2) - c, f(2) - c$
2	$f(2) - 2c, f(2)$	$f(3) - 2c, f(3) - c$	$f(4) - 2c, f(4) - 2c$	$f(4) - 2c, f(4) - 2c$
q^*	$f(1) - c, f(1)$	$f(2) - c, f(2) - c$	$f(4) - 2c, f(4) - 2c$	$f(4) - 2c, f(4) - 2c$

4. Leadership and Evolution

A notable aspect of many international public goods treaties is that they are often preceded by intermediate actions by nations that constitute a strict subset of signatories to the ultimate agreement. The Montreal Protocol was preceded by a ban on aerosol CFCs in the United States and Canada that was also supported by Sweden and Norway long before the treaty was initially ratified in 1987 by 24 countries. By contrast, other OECD nations have taken the lead when it comes to CO_2 reductions. Leadership in this form is not the usual way in which voluntary provision is described and supported in a repeated-game environment. Specifically, when tacit cooperation is achieved in a Prisoner's Dilemma via a tit-for-tat or (grim) trigger strategy, *all* players start out by cooperating. This is because most supergame strategies are akin to static programs where no learning takes place once the strategies are put into play. By contrast, in an evolutionary environment players can learn the relative merits of actions over time and may indeed switch to voluntary provision when its benefits are learned.

Arce (2001) suggests the name *leading by example* to describe a process whereby the cooperative solution to a voluntary provision game can be attained through the leader's unilateral commitment to an intermediate level of provision, and matching behavior there beyond.[7] The *leading-by-example strategy* (q^*) is defined for agent i by

$$q^* = \begin{cases} 1, & \text{if } q_j = 0 \text{ or } 1 \\ 2, & \text{if } q_j = 2 \end{cases}, \quad j \neq i. \tag{8}$$

Under q^* agent i unilaterally provides one unit of the public good and commits to providing two units if j behaves similarly. When Box 1 is extended to include the leading by example strategy, q^*, it can be represented as in Box 2. The 3×3 game (in bold) nested within Box 2 is identical to Box 1. In the q^* row (or column) of Box 2 the q^* policy functions as $q = 1$ when matched with $q' = 0$ or $q' = 1$. It functions as $q = 2$ when matched with $q' = 2$ (or another leader). The difference between leading by example and reciprocity/matching can be seen by examining the replicator dynamic in (6).

[7]Note that in addition to summation, Arce (2001) considers leadership under best shot, better shot, weakest link, and weaker link aggregators.

The aim of leading by example is to influence the coefficient on the average payoff difference on the right-hand side of (6) by increasing the frequency of collective action through unilateral provision (if need be) and matching behavior there beyond.

Results: We present below a series of propositions that characterize the potential of leading by example for the game in Box 2 that satisfies assumptions (A1) and (A2). Proofs of all these Propositions are provided in the Appendix.

PROPOSITION 1: *Strategy $q = 2$ can never be evolutionarily stable. If $f(1) > c$, then it is neutrally stable if*

$$f(4) - f(3) \geq c, \quad f(3) - f(2) \geq c, \quad f(2) - f(1) \geq c. \tag{9}$$

Proposition 1 essentially states that leading by example is strategically different than providing at level $q = 2$. The only time that $q = 2$ implements the cooperative outcome is if (9) is satisfied. When (9) holds, then it is always a dominant strategy to contribute one more unit of the public good. This is similar to Guttman's (1978) process of using iterated dominance to choose among Pareto-ranked equilibria. Proposition 3 (below) demonstrates that leading by example does not require this condition.

PROPOSITION 2: *q^* can never be an evolutionarily stable strategy.*

This proposition validates NSS as a solution concept; otherwise, under the ESS criterion we would be allowing one socially optimal strategy ($q = 2$) to eliminate another (q^*) because the two are also socially optimal when paired with each other.

PROPOSITION 3: *A sufficient condition for q^* to be a neutrally stable strategy is $f(4) - f(1) > 2c$. If $f(4) - f(1) = 2c$ then q^* is a NSS if and only if $f(1) \geq c$.*

Two conclusions can be drawn from this result. First, the most direct way of ensuring that the game is not the PD is to assume $f(1) \geq c$. This is akin to the idea that a country has sufficient valuation for the public good such that for an initial range of Q its marginal utility exceeds its marginal cost of abatement.[8] Hence, some incentive exists for (suboptimal) unilateral contribution. In terms of international treaties, Sandler (1998) credits the localized transportation matrix for acid rain as being critical for the establishment of the Long-range Transboundary Air Pollution (LTRAP) Convention in Europe during 1979. Localized effects are consistent with $f(1) \geq c$. Prior to its ban on aerosol CFCs, the United States was by far the largest emitter of ozone-destroying chemicals, and its leadership strategy slowed a dangerous trend (Benedict 1991). Slowing a trend in ozone depletion is commensurate with $f(1) \geq c$. Finally, Young (1998) documents that Iceland and some of

---

[8]The Chicken game examined by Cavaliere (2001) is a binary choice example.

the Nordic states received significant environmental benefits from their own actions; consequently, they were willing to take a leadership role ahead of the United States and Soviet Union prior to ratification of the Arctic Environmental Protection Strategy (AEPS) in 1993. Again, this can be interpreted as $f(1) \geq c$ for the AEPS leaders.

A particularly interesting possibility is the case where (1, 1) is Nash, but (2, 2) is not. Leading by example can still support the (2, 2) outcome if the conditions of Proposition 3 are met. In this way, leading by example overcomes Nash codification.

The second conclusion is that the requirements for leading by example to be NSS are linked to the "concavity" of the function U.[9] In fact, the conditions

$$f(4) - f(1) \geq 2c \quad \text{and} \quad f(4) - f(2) \geq c \quad \text{(from (A2))} \tag{10}$$

indicate that it is the behavior of the first differences, $f(s+1) - f(s)$, of the function f (or U) that dictates whether or not the leading-by-example strategy is neutrally stable. Specifically, a function g exhibits a lower degree of concavity (curvature) than a function h if[10]

$$\begin{aligned} &[h(s+1) - h(s)] - [h(s) - h(s-1)] \\ &\quad < [g(s+1) - g(s)] - [g(s) - g(s-1)]\ \forall\, s \in \{1, 2, 3\}. \end{aligned} \tag{11}$$

Under the conditions of our model—the utility function is nonnegative, increasing, and vanishes at zero—to say that g has a lower degree of concavity (curvature) than h is equivalent to saying that g is more steeply sloped than h.

We can therefore remark that when the function f exhibits a relatively high degree of concavity, in the sense that $f(s+1) - f(s)$ is small in comparison to c for all s, but not so high as to invalidate the conditions in (10), then (9) does not hold and leadership is required for successful collective action.[11] Another possibility is that the function has sufficiently low degree of concavity, in the sense that the first differences are larger than c, and then q^* is automatically sustained as a neutrally stable strategy. When the marginal benefit of the public good does not decrease at a steep rate, the gains society obtains from greater contributions to the public good are higher, increasing the chance of success for the leader.

Heckathorn (1996) shows that diminishing marginal utility is decisive in determining whether or not a summation game is the PD. Our generalization demonstrates that diminishing marginal utility is essential for assessing whether optimal provision can occur under leadership in such non-PD

[9]Strictly speaking, this is not concavity, but rather a concavity-related concept because the function U is defined on a discrete domain.

[10]If these function were differentiable, this would amount to requiting $h^{n}(s) < g^{n}(s)\ \forall\, s$ or, since the functions are concave, $|\, h^{n}(s)| > |\, g^{n}(s)|\ \forall\, s$. In other words, the first derivative of h as a function of s should be more steeply sloped than the derivative of g everywhere.

[11]For example, the results obtained by Arce (2001) are for the case of $f(1) > c$, $f(2) - f(1) \geq c$, and $f(3) - f(2) < c$.

games.[12] Leading by example is also a policy whereby unilateral provision can be used to *preclude* a free-riding outcome. For example, the leader unilaterally abates at $q = 1$ to commit itself away from the Nash equilibrium (2, 0). In this way, the leader avoids a situation of exploitation à la Olson (1965) where another country free rides completely. Instead, the leader commits to additional abatement beyond the unilateral level only if such abatement matches that of others. Hence, in order to receive benefits beyond the unilateral level, other countries must abate at the cooperative level. The result is the noncooperative implementation of the cooperative solution without any need for a punishment strategy that may be infeasible in an international context.

What remains to be seen is if leadership can induce collective action for the case where $f(1) < c$. It is straightforward to show that q^* cannot be an NSS and $q = 0$ is the only ESS when $f(1) < c$. Hence, the only relevant restriction for ruling out the PD is $f(1) \geq c$.

5. Conclusion

In this paper, we investigated how cooperative behavior in the provision of international public goods can be obtained noncooperatively as a result of an evolutionary process. More specifically, we examined the conditions under which leading by example can induce collective action in the form of a neutrally stable strategy. Our focus is on games that rule out the Prisoner's Dilemma by assumption. At the same time, we investigate game conditions that are consistent with many situations of international collective action. One example is if the benefit of the first unit of provision of the public good is no smaller than the marginal cost of contribution. This is consistent with the localized effects of sulfur emissions, the deceleration of ozone depletion, and the significant environmental benefits to Nordic countries of protecting the Arctic environment.

In technical terms, the success of leading by example is closely linked to the degree of concavity (curvature) of the public benefit function. Leading by example is neutrally stable if this function has a sufficiently low degree of concavity, in the sense that the first differences do not diminish too fast in comparison with the cost of contributing. It also requires some incentive for unilateral action (to rule out the PD). These conditions can be tested against empirical estimates of the public benefits for any proposed regime.

Appendix

Prior to the proofs of individual theorems it is worthwhile noting that assumption (A2) implies

$$f(4) - f(1) \geq 1.5c \quad \text{and} \quad f(4) - f(2) \geq c.$$

[12]Arce and Sandler (2001) find that the effectiveness of an alternative form of leadership that alters the information structure (via correlated strategies) is similarly influenced by diminishing marginal utility.

Proof of Proposition 1: When the conditions in (9) hold, they imply

$$f(4) - f(2) \geq 2c, \quad f(4) - f(1) \geq 3c, \quad f(3) - f(1) \geq 2c.$$

For $q = 2$ to be a Nash equilibrium it must be the case that $f(4) - f(3) \geq c$, which is why it is required in (9), and $f(4) - f(2) \geq 2c$, which, as indicated above, is implied by (9). By the summation aggregator, we know that $V_i(2, 2) = V_i(q^*, 2)$ and $V_i(2, q^*) = V_i(q^*, q^*)$. Therefore, $q = 2$ cannot be ESS and we proceed to check if it can be NSS.

Case 1: if $f(4) - f(3) = c$, then $V_i(2, 2) = V_i(1, 2)$ and the FOC holds with equality. Checking the SOC′, given $V_i(2, 1) = f(3) - 2c$ and $V_i(1, 1) = f(2) - c$, and $f(3) - f(2) \geq c$, we have $V_i(2, 1) \geq V_i(1, 1)$.

Case 2: if $f(4) - f(2) = 2c$, then $V_i(2, 2) = V_i(0, 2)$ and the FOC holds with equality. As $V_i(2, 0) = f(2) - 2c$, $V_i(0, 0) = f(0)$, $f(2) - f(1) \geq$ c and $f(1) >$ c, we have $V_i(2, 0) > V_i(0, 0)$, thereby satisfying the SOC′; i.e. (2, 2) is a NSS. ■

Proof of Proposition 2: Holds by reason of the summation aggregator, as argued in the first paragraph in the Proof of Proposition 1. ■

Proof of Proposition 3: (q^*, q^*) is a Nash equilibrium iff $f(4) - f(2) \geq c$ and $f(4) - f(1) \geq 2c$. Assumption (A2) guarantees $f(4) - f(2) \geq$ c, and if it holds with weak inequality q^* trivially satisfies the SOC′ for NSS against $q^m = 1$. If $f(4) - f(1) > c$, then the FOC holds with strict inequality for $q^m = 0$. If $f(4) - f(1) = 2c$, then the SOC requires $V(q^*, 0) \geq V(0, 0)$; i.e., $f(1) > c$. ■

References

ARCE M., D. G. (2001) Leadership and the aggregation of international collective action, *Oxford Economic Papers* **53**, 114–137.

ARCE M., D. G., and T. SANDLER (2001) Regional public goods: strategies and institutions, *European Journal of Political Economy* **17**, 493–516.

BENEDICT, R. E. (1991) Protecting the ozone: New directions in diplomacy, in *Preserving the Global Environment. The Challenge of Shared Leadership*, J. Tuchman Mathews, Washington, DC: World Resources Institute, 112–153.

CAVALIERE, A. (2001) Coordination and the provision of discrete public goods by correlated equilibria, *Journal of Public Economic Theory* **3**, 235–255.

GUTTMAN, J. M. (1978) Understanding collective action: Matching behavior, *American Economic Review, Papers and Proceedings* **68**, 251–255.

HARLEY, C. B. (1981) Learning the evolutionary stable strategy, *Journal of Theoretical Biology* **89**, 611–633.

HECKATHORN, D. D. (1996) The Dynamics and dilemmas of collective action, *American Sociological Review* **61**, 250–277.

HOFBAUER, J., and K. SIGMUND (1998) *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge University Press.

KINDLEBERGER, C. P. (1981) Dominance and leadership in the international economy. Exploitation, public goods, and free rides, *International Studies Quarterly* **25**, 242–254.

MAYNARD SMITH, J. (1982) *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

MAYNARD SMITH, J., and G. R. PRICE (1973) The logic of animal conflict, *Nature* **246**, 15–18.

MURDOCH, J. C., and T. SANDLER (1997) The voluntary provision of a pure public good: The case of reduced cfc emissions and the montreal protocol, *Journal of Public Economics* **63**, 331–349.

MURDOCH, J. C., T. SANDLER, and K. SARGENT (1997) A tale of two collectives: Sulfur versus nitrogen oxides emission reduction in Europe, *Economica* **64**, 281–301.

OLSON, M. (1965) *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.

PALFREY, T., and H. ROSENTHAL (1984) Participation in the provision of discrete public goods: A strategic analysis, *Journal of Public Economics* **3**, 171–193.

SANDLER, T. (1978) Interregional and intergenerational spillovers, *Scottish Journal of Political Economy* **25**, 273–284.

SANDLER, T. (1997) *Global Challenges: An Approach to Environmental, Political and Economic Problems*. Cambridge: Cambridge University Press.

SANDLER, T. (1998) Global and regional public goods: A prognosis for collective action, *Fiscal Studies* **19**, 221–247.

SANDLER, T., and D. G. ARCE M. (2003) Pure public goods versus commons: Benefit-cost duality, *Land Economics* **79**, 355–368.

SUGDEN, R. (1984) Reciprocity: The supply of public goods through voluntary contributions, *Economic Journal* **94**, 772–787.

TAYLOR, P. D., and L. B. JONKER (1978) Evolutionary stable sets and game dynamics, *Mathematical Biosciences* **40**, 145–156.

TUCHMAN MATTHEWS, J. (1991) *Preserving the Global Environment. The Challenge of Joint Leadership*. NY: Norton.

WAGNER, R. E. (1965) Pressure groups and political entrepreneurs: A review article, *Papers on Non-Market Decision Making* **1**, 161–170.

YOUNG, O. (1998) *Creating Regimes. Arctic Accords and International Governance*. Ithaca: Cornell University Press.