Journal of Environmental Economics and Management
doi:10.1006/jeem.2000.1171, available online at http://www.idealibrary.com on IDEAL®

# Transboundary Air Pollution, Environmental Aid, and Political Uncertainty[1]

Paul E. Chambers[2]

*Department of Economics and Finance, Central Missouri State University, Warrensburg, Missouri 64093-5074*

and

Richard A. Jensen[3]

*Department of Economics, University of Notre Dame, Notre Dame, Indiana 46556*

Received April 25, 1999; revised June 20, 2000

A two-stage game is used to evaluate the effectiveness of untied aid in reducing transboundary emissions. The donor nation (North) has incomplete information regarding the political willingness of the recipient (South) to enforce emission standards. The South may be tough or weak on pollution. We provide necessary and sufficient conditions for the existence of pooling and separating equilibria. Perversely, untied Northern aid is a potential source of higher pollution, at least in the short run, because it provides an incentive for strategic, reputation-building behavior in the form of excessive Southern emissions. 

## I. INTRODUCTION

Unilateral transboundary pollution poses significant difficulties. The inability of one sovereign nation to impose emission taxes on another suggests that subsidies to foreign regulatory agencies may play important roles in reducing unidirectional emissions.[4] As noted by d'Arge [11], a polluting country has little incentive to unilaterally apply emission taxes to reduce the effects of its pollution on a

[1] We would like to thank Nancy Dyer, James R. Kahn, Ehud Kalai, Robert Kohn, Roger Myerson, Todd Sandler, John Whitehead, and two anonymous referees for helpful comments. We have also benefitted from the comments of seminar participants at the Heartland Environmental and Resource Economics Workshop and the Midwest Mathematical Economics Meetings. Jensen's work was supported by a Summer Research Grant from the Gatton College of Business and Economics of the University of Kentucky. The grant was made possible by a donation of funds to the College by Ashland Oil, Inc. Jensen's work was conducted in part while visiting at the Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, whose support is gratefully acknowledged.

[2] Address correspondence to: Paul E. Chambers, Department of Economics and Finance, Central Missouri State University, Warrensburg, MO 64093-5074 or pchambers@cmsu1.cmsu.edu.

[3] E-mail: rjensen1@nd.edu.

[4] It is well known that in the domestic case with certainty, taxes are the preferred instrument to abate pollution since subsidies may lead to other problems. Specifically, in the long run, subsidies may increase industry pollution as the number of firms in the industry expands due to the entry of new firms (see Kohn [19, 20], Lewis [21]; Polinsky [26]).



0095-0696/01 $35.00

neighboring country (also see Baumol and Oates [5, pp. 278–283]).[5] As an alternative to emission taxes, some international development agencies provide subsidies to foreign regulatory agencies in order to reduce unidirectional emissions since many less developed countries lack basic "institutional compatibilities" for market based incentives (see Russell and Powell [27]). Grossman and Krueger [16] and Sánchez [28] note that a weak regulatory infrastructure combined with the expansion of Mexican maquiladoras industry is a possible source of the pollution in the border region of Mexico and the United States. Sánchez [28] argues that the Secretaría de Desarrollo Urbano y Ecología or SEDUE (the former Mexican equivalent of the U.S. EPA) suffers from a lack of resources to enforce environmental legislation. In the environmental accord of NAFTA, U.S. and Mexican trade negotiators agreed on financing mechanisms for environmental infrastructure projects in the border region (see Baer and Weintraub [3, p. 82]). There is evidence that numerous development organizations provide aid with such intentions. For example, Japan's Overseas Economic Cooperation Fund (OECF) provides untied environmental aid for the financing of "pollution abatement activities." The Inter-American Development Bank (IDB) provides grants specifically to strengthen newly established environmental protection agencies (for examples of IDB support, see Russell and Powell [27]).[6] The success of such untied aid depends on the recipient nation's willingness and ability to implement emission-abatement programs.

Although the subsidies to the regulatory agencies may be desirable due to low incomes and weak regulatory infrastructure in developing countries, another problem may arise since the aid is not tied to the level of emissions in many of these programs.[7] We argue that uncertainty regarding the general ability or will of the recipient nation's government to enforce emissions standards is a key factor in determining the level of success of various environmental aid programs. Russell and Powell [27] argue that political will plays an important role in determining the success of environmental improvement programs. Under certain conditions, the recipient nation may have very limited incentives to enforce environmental standards. When there is uncertainty regarding the recipient nation's ability or willingness to control emissions, the success of the emissions reduction program is also uncertain.[8]

[5] Countries may have a strategic incentive to abate pollution in a cooperative setting (see Barrett [4]; Viejo et al. [31]; Petrakis and Xepapadeas [25]).

[6] For examples of descriptions of recent project loans by the OECF, see http://www.oecf.go.jp. Additional examples of programs by the IDB may be found in the program section of its web site www.iadb.org.

[7] Connolly et al. [8] note that organizational inertia in the form of "a familiar set of preferred solutions" precludes donor institutions from choosing a more optimal form of environmental assistance. Kanbur et al. [18] argue against the conditionality of aid not only because they perceive conditionality to be ineffective, but also because it is very costly to the recipient country if it has a very limited number of skilled administrators. Fairman and Ross [15] summarize the literature concerning the real-world experience of conditionality as "[a] policy instrument that fails more often than it succeeds."

[8] From a related area, environmental aid for habitat preservation programs often yields less than expected preservation levels and in the worst cases may only result in very limited levels of preservation in the form "paper parks" (see McNeely [23]). For example, the first debt-for-nature swap was plagued by problems because the Bolivian government did not fully enforce the nominal property rights (see Deacon and Murphy [13]). Bohn and Deacon [6] and Deacon [12] find political instability to be an important determinant of deforestation (a form of disinvestment). Political instability may also explain the low investment in environmental infrastructure in certain countries.

For simplicity, our focus is on the provision of aid to reduce emissions where the type of uncertainty takes the form of asymmetric information regarding the recipient nation's political will to control emissions. We use the standard North–South framework to represent the higher and lower income countries and model the aid/emissions problem as a two-stage game. We develop a noncooperative model in which a Northern country provides aid to a Southern country with the intention of reducing emissions within the Southern country.[9] However, the Southern country has an incentive to retain high emission levels because a portion of its output is produced in a polluting industry, and the South must engage in costly abatement to reduce its emissions. In the model, we assume that there exists uncertainty regarding the true type of the South, in that it may be tough or weak on pollution, depending on its ability or political will to reduce emissions. In a static situation, this uncertainty leads to lower emissions by a tough type of South, but higher emissions by a weak type. Moreover, this uncertainty leads to less aid than if the North knew the South was weak on pollution, but more aid than if it knew the South was tough. Hence, the tough Southern type gains from the uncertainty about its type, but the weak Southern type loses from the uncertainty.

The effects of this uncertainty on Southern welfare have interesting implications in a dynamic context. In particular, a tough Southern country may have an incentive to choose a high emissions strategy in order to develop a reputation as being weak on pollution. If it successfully develops such a reputation, it will receive higher aid in the future (than if it chose a low emissions strategy and revealed its true type). That is, the Northern aid program may have the unfortunate effect of inducing excessive current emissions by a tough South type intent on disguising its type and reaping future gains from doing so. Even more perversely, the aid may also induce excessive current emissions by a weak South intent on proving its type and thereby reaping future gains. Oddly, it is the North's response to emissions that actually encourages the South to engage in excessively high emissions. Because the North has the goal of reducing emissions, its best response to higher levels of emissions is to provide more aid to the South. This provision of additional aid when confronted with increased emissions reduces the South's cost of reputation building via excessive emissions.

Admittedly, these excessive emissions are a short-run phenomenon because in each case the South must reveal its true type to obtain its future gain, and so the North's uncertainty is resolved. Nevertheless, the same opportunities for excessive emissions return whenever any random shock occurs that can reintroduce uncertainty about the ability or political will of the Southern government to enforce its emissions standards. For example, an election that changes the party in power, a natural disaster, or a currency devaluation could easily cause such renewed uncertainty. Hence, our results about excessive emissions may be particularly pertinent when the polluting country is subject to political instability, since this can lead to the recurrence of uncertainty. Indeed, in this regard our results are consistent with those of Bohn and Deacon [6] and Deacon [12], who find empirical

[9] Buchholz and Konrad [7] consider a cooperative bargaining outcome in which the polluting country chooses irreversible investment in either high or low pollution control technologies in order to affect the threat points.

support for the hypothesis that political instability is a source of environmental degradation.

We proceed as follows. In Section II, we derive a static Bayesian model and show how uncertainty affects the levels of aid and emissions. In Section III, we develop a two-stage Bayesian game in which the Southern country may choose a level of emissions in the first stage in order to influence the stage-two level of aid. In Section IV, we provide some conclusions and a brief sketch of how the model could be extended to an infinitely repeated game in which uncertainty recurs systematically. Proofs and technical details are collected in the Appendix.

## II. A STATIC BAYESIAN GAME OF AID AND EMISSIONS

We use the North–South framework in which polluting industries are located within the South's borders. Environmental quality is assumed to be a normal good for both countries, and the North has a higher income.[10] We assume the North's welfare can be expressed by the utility function $U^n(Y^n, C, q^n)$, where $Y^n$ is its post-transfer income, $C(A)$ is its contributions in the form of aid $A$ to the South, and $q^n(E)$ is its air quality which depends on the level of Southern emissions $E$.[11] We assume Northern utility is an increasing function of its post-transfer income and air quality. Since the North's post-transfer income is $Y^n = y^n - A$, where $y^n$ is its exogenous income, aid transfers reduce North's available income.

Contributions are a "good" for the North, $\partial C/\partial A > 0$ and $\partial U^n/\partial C > 0$, if a "warm glow" exists.[12] In addition to a warm glow effect, the public choice literature suggests that the contribution level may positively affect the welfare of the aid administrators since they may get utility from maximizing the budget they control (see Niskanen [24]). Consistent with this view is the work of Kanbur et al. [18], who find donor countries may have incentives to distribute aid regardless of the results. Since the funding of future projects depends in part on past budgets, delays in the distribution of aid may endanger the livelihood of the staff of the aid agency. Hence, they argue that staff members receive benefits from the distribution of aid that are independent of the project's results. Also, they note that private

[10] Depending on a country's position on its environmental Kuznets curve, income, and therefore aid, may reduce the country's level of emissions. Grossman and Krueger [16] find an environmental Kuznets curve, in that the level of emissions is an increasing function of GDP to a point and then begins to decrease as GDP increases. Selden and Song [30] also find an inverted-U relationship between per-capita emissions and per-capita GDP. López [22] provides a theoretical model which supports the inverted-U relationship between per-capita emissions and per-capita GDP. We assume that the South is on the downward sloping portion of its environmental Kuznets curve. Copeland et al. [9] develop a two country general equilibrium model in which untied aid from the North reduces pollution in the South by affecting relative incomes and therefore affecting relative pollution taxes.

[11] We assume throughout that all functions are continuously differentiable as often as needed.

[12] The existence of a warm glow arises in various contexts. For example, Cornes and Sandler [10] provide an impure joint-product model in which a warm glow occurs as a special case. Kahneman and Knetsch [17] question the validity of the contingent valuation method since the willingness to pay of survey respondents may be based on the moral satisfaction of contributing to the provision of the public good and not the economic value of the good. In an experimental setting, Andreoni [2] finds that subjects are motivated to cooperate due to kindness or a "warm glow."

firms within the donor country may benefit from the distribution of the aid and, therefore, pressure the donor agency to distribute the aid in a timely fashion.[13]

Because air quality is a normal good and aid reduces post-transfer income, it follows that the marginal utility of air quality is reduced by aid expenditures, $\partial^2 U^n / \partial q^n \, \partial A < 0$. And since air quality is reduced by emissions, this implies that $\partial^2 U^n / \partial E \, \partial A > 0$, so aid and emissions are strategic complements for the North. For notational convenience, we rewrite the Northern utility function as $V^n(A, E)$. The North's objective at any date is to choose a level of aid to maximize its utility. Because contributions are a good, the North may choose to provide aid even in the absence of Southern pollution. Of course, the North may have an additional incentive to provide aid in order to encourage the reduction of emissions. This income transfer may influence emissions by affecting the level of enforcement by the regulatory agency and/or by affecting personal income levels since air quality is a normal good.

We assume the South's welfare can be expressed by the utility function $U^s(Y^s, q^s)$, where $Y^s$ is its income and $q^s(E)$ is its air quality, which depends on its emissions. Southern income is $Y^s = y^s + Q(E) + A$, where $y^s$ is its exogenous income and Q(E) is its net income from the polluting industry. We assume that income from this sector is directly related to the level of emissions, $\partial Q / \partial E > 0$. Air quality is a good for the South also, $\partial q^s / \partial E < 0$ and $\partial U^s / \partial q^s > 0$, and since it is a normal good we have $\partial^2 U^s / \partial q^s \, \partial Y^s > 0$. Again for convenience, we rewrite the Southern utility function as $V^s(A, E)$. Because air quality is a normal good and improving it requires the South to forgo income from pollution, it is likely that the marginal utility of emissions is high at low levels of income.

Our objective in this section is to analyze a static game in which the North has incomplete information about the willingness or ability of the South to abate emissions. For simplicity, we assume that, depending upon its political will to abate emissions, the South can be either tough or weak on pollution. We model these two types by assuming there are two possible Southern utility functions, $V_T^s(A, E)$ for the *tough* type and $V_W^s(A, E)$ for the *weak* type. Each type's problem is to choose a level of emissions to maximize its utility. We assume $V_T^s(A,E)$ and $V_W^s(A, E)$ are each strictly concave in $E$ for all $A$, so each type's utility maximization problem has a unique solution for all $A$. The corresponding Southern best-response (reaction) functions are denoted $r_T^s(A)$ and $r_W^s(A)$. For any given level of Northern aid A, the level of emissions that maximizes Southern utility is $r_T^s(A)$ if the South is the tough type, and $r_W^s(A)$ if it is the weak type. These reaction functions are depicted in Fig. 1. They are negatively sloped because aid increases Southern income, and air quality is a normal good, so the South's best response to an increase in the level of aid is to choose a lower level of emissions, whatever its type.[14] Finally, to insure the designations weak and tough make sense,

[13] Scheyvens [29] argues that international prestige is a factor in the determination of both the magnitude and direction of aid flows from Japan's ODA. Alesina and Dollar [1] find that the direction of foreign aid is influenced more by the political factors, such as UN voting patterns, than economic need and policy performance of the recipient nation. Wade [32] shows that the actions of the World Bank were influenced by the environmental movement.

[14] For each type $t = T, W$, the equation $\partial V_t^s(A, r_t^s(A)) / \partial E = 0$ implicitly defines its reaction function, $r_t^s(A)$, with slope $\partial r_t^s / \partial A = -(\partial^2 V_t^s / \partial E \, \partial A) / (\partial^2 V_t^s / \partial E^2)$. Since $\partial^2 V_t^s / \partial E^2 < 0$ by assumption, the sign of $\partial r_t^s / \partial A$ is the sign of $\partial^2 V_t^s / \partial E \, \partial A = (\partial^2 U_t^s / \partial Y^{s2})(\partial Q / \partial E) + (\partial^2 U_t^s / \partial Y^s \, \partial q^s)(\partial q^s / \partial E) < 0$.

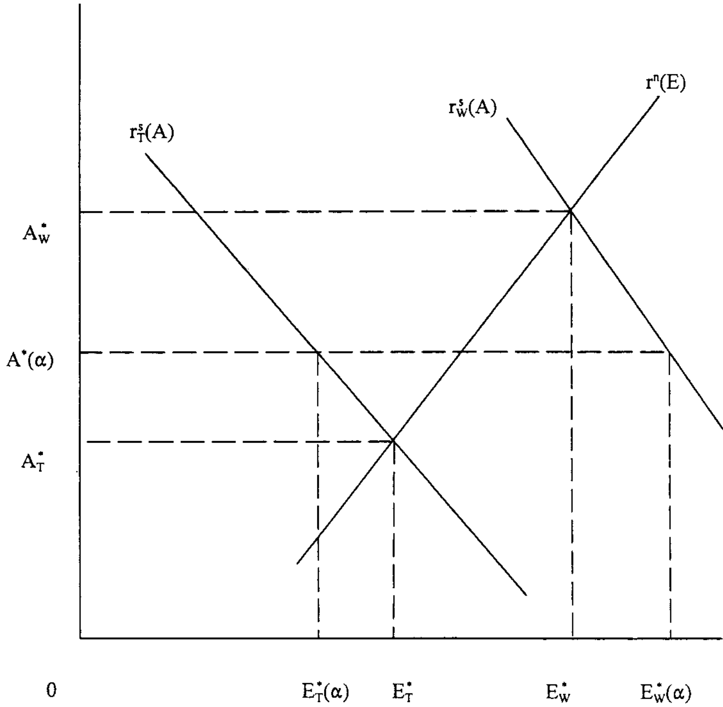


**FIG. 1.** Static Bayesian equilibrium.

we assume that the marginal utility of emissions is greater for the weak type, $\partial V_W^s(A, E)/\partial E > \partial V_T^s(A, E)/\partial E$ for all $A$. This guarantees that the South's best response to any given level of aid is to pollute more if it is the weak type, $r_W^s(A) > r_T^s(A)$, and so that the reaction function of the tough type lies below (to the left) the reaction function of the weak type, as also shown in Fig. 1.

To analyze the North's problem, assume for the moment that it knows the Southern type with certainty. Its problem then is to choose a level of aid to maximize its utility, given the level of Southern emissions. We assume $V^n(A, E)$ is strictly concave in $A$ for all $E$, so there exists a unique solution to this maximization problem for all $E$. The Northern reaction function is denoted $r^n(E)$. For any given level of emissions, $r^n(E)$ is the level of aid that maximizes Northern utility. This reaction function is also shown in Fig. 1. It is positively sloped because the North has a diminishing marginal willingness to pay for air quality, so its best response to an increase in emissions is to increase the aid it provides to the South.[15]

Before presenting the Bayesian equilibrium of this game of incomplete information, it is useful to consider the Nash equilibria which would arise in the two possible games of complete information in which the South's type is known with

[15] The equation $\partial V^n(r^n(E), E)/\partial A = 0$ implicitly defines the North's reaction function, $r^n(E)$, with slope $\partial r^n/\partial E = -(\partial^2 V^n/\partial A\, \partial E)/(\partial^2 V^n/\partial A^2)$. Since $\partial^2 V^n/\partial A^2 < 0$ by assumption, the sign of $\partial r^n/\partial E$ is the sign of $\partial^2 V^n/\partial A\, \partial E = (\partial^2 U^n/\partial A\, \partial q^n)(\partial q^n/\partial E) > 0$.

certainty. First suppose it is common knowledge that the South lacks the political will to abate emissions. Then the Nash equilibrium of this certainty game is the level of aid and level of emissions, $(A_W^*, E_W^*)$, given by the intersection of the Northern reaction function, $r^n(E)$, and the reaction function of the weak Southern type, $r_W^s(A)$. Next suppose it is common knowledge that the South is a tough type. Then the Nash equilibrium of this certainty game, $(A_T^*, E_T^*)$, is given by the intersection of the Northern reaction function and the reaction function of the tough Southern type, $r_T^s(A)$. These two Nash equilibria are also shown in Fig. 1. Given the slopes and relative positions of the types of Southern reaction functions, and the slope of the Northern reaction function, it is evident that $A_T^* < A_W^*$ and $E_T^* < E_W^*$. That is, the levels of aid and emissions are both lower in the certainty game with a Southern country that is tough on emissions.

Now return to the situation where the North does not know whether the South is a weak or tough type. That is, consider the Bayesian game of incomplete information in which, at the time levels of aid and emissions are chosen, the South has private information regarding its type (i.e., only it knows if it is weak or tough). Formally, the strategy of Southern type $t = T, W$ is a choice of emissions, $E_t$, from its strategy set, $S_t^s$, the set of feasible emissions. We assume the strategy sets are $S_T^s = S_W^s = [0, \overline{E}]$, where $\overline{E}$ is just a technological upper bound on the set of feasible emissions. The Southern payoff functions are the utility functions, $V_T^s(A, E)$ for the tough type and $V_W^s(A, E)$ for the weak type. The North's strategy is to choose a level of aid from its strategy set, $S^n = [0, y^n]$. If $\alpha$ denotes the prior probability that the South is a weak type with low political will to abate emissions, then the North's payoff is its expected utility

$$N(A, E_T, E_W) = \alpha V^n(A, E_W) + (1 - \alpha)V^n(A, E_T). \tag{1}$$

Because the North does not know which type it faces, its optimal choice is a best response to both $E_T$ and $E_W$. This Bayesian reaction function is denoted $R(E_T, E_W)$.[16] A Bayesian equilibrium is then a triple $(A^*(\alpha), E_T^*(\alpha), E_W^*(\alpha))$ such that

$$V_W^s(A^*(\alpha), E_W^*(\alpha)) \geq V_W^s(A^*(\alpha), E) \qquad \text{for all } E \in S_W^s,$$

$$V_T^s(A^*(\alpha), E_T^*(\alpha)) \geq V_T^s(A^*(\alpha), E) \qquad \text{for all } E \in S_T^s,$$

$$N(A^*(\alpha), E_T^*(\alpha), E_W^*(\alpha)) \geq N(A, E_T^*(\alpha), E_W^*(\alpha)) \qquad \text{for all } A \in S^n.$$

In terms of Fig. 1, at the time the levels of aid and emissions are chosen, everyone knows the location of the North's reaction function $r^n(E)$. The South knows the location of its reaction function also. However, the North knows only that the South's reaction function is either $r_W^s(A)$ with probability $\alpha$ or $r_T^s(A)$ with probability $1 - \alpha$. This uncertainty leads to greater aid from the North than it would provide if it knew the South was tough, but less aid than it would provide if it knew the South was weak. Note that the lowest emissions, $E_T^*(\alpha)$, occur if uncertainty exists and the South is tough. Uncertainly implies that aid is higher,

[16] The equation $\partial N(R(E_T, E_W), E_T, E_W)/\partial A = 0$ implicitly defines this Bayesian reaction function with slopes $\partial R(E_T, E_W)/\partial E_t = -(\partial^2 N/\partial A\, \partial E_t)/(\partial^2 N/\partial A^2) > 0$ for each $t = T, W$.

$A^*(\alpha) > A_T^*$, and emissions are lower, $E_T^*(\alpha) < E_T^*$, than if the North knows the South is tough. However, uncertainty also implies that aid is lower, $A^*(\alpha) < A_W^*$, and emissions are higher, $E_W^*(\alpha) > E_W^*$, than if the North knows the South is tough. Recalling from our discussion of the two certainty games that $E_T^* < E_W^*$, the ordering of emissions is $E_T^*(\alpha) < E_T^* < E_W^* < E_W^*(\alpha)$. Finally, note that an increase in the probability of South being the weak type increases North's willingness to provide aid since $\partial^2 V^n / \partial A\, \partial E > 0$. At greater levels of uncertainty, the North's expected marginal disutility from providing aid is decreased since $r_T^s(A) < r_W^s(A)$ and $\partial^2 V^n / \partial E\, \partial A > 0$; therefore, it follows that $A^*(\alpha)$ increases. And as $A^*(\alpha)$ increases, both $E_T^*(\alpha)$ and $E_W^*(\alpha)$ decrease due to the negative slopes of $r_T^s(A)$ and $r_W^s(A)$.

THEOREM 1. *For any given* $\alpha \in (0,1)$, *there exists a unique and locally stable Bayesian equilibrium*[17] $(A^*(\alpha), E_T^*(\alpha), E_W^*(\alpha))$ *such that* $E_T^*(\alpha) < E_T^* < E_W^* < E_W^*(\alpha)$ *and* $A_T^* < A^*(\alpha) < A_W^*$ *where* $A^*(\alpha)$ *is an increasing function of* $\alpha$, *and* $E_T^*(\alpha)$ *and* $E_W^*(\alpha)$ *are decreasing functions of* $\alpha$.

For each Southern type $t = T, W$, any movement along its reaction function $r_t^s$ to a higher level of aid results in higher utility. Welfare comparisons of the uncertainty game with the certainty game then follow immediately from this fact and the ranking of aid $A_T^* < A^*(\alpha) < A_W^*$ from Theorem 1. That is, the tough type gains but the weak type loses from uncertainty.

COROLLARY 1. *If the South is a tough type, then compared to the outcome with no private information, uncertainty increases its utility,* $V_T^s(A^*(\alpha), E_T^*(\alpha)) > V_T^s(A_T^*, E_T^*)$. *Conversely, if the South is a weak type, then compared to the outcome with no private information, uncertainty decreases Southern utility,* $V_W^s(A^*(\alpha), E_W^*(\alpha)) < V_W^s(A_W^*, E_W^*)$.

Notice, however, that these results from the static game do not provide a complete or plausible explanation for the failure of certain international environmental aid programs. On the contrary, the static game analysis suggests that the donor nation may actually be pleasantly surprised with the effectiveness of the aid. To explain the partial failure of aid programs, we note that the results of Corollary 1 indicate the possibility of strategic, reputation-building behavior in response to environmental aid programs. In particular, a tough Southern type benefits from the Northern uncertainty about its type and, therefore, has an incentive to maintain this uncertainty. Conversely, a weak Southern type benefits from Northern certainty and, therefore, has an incentive to credibly reveal its type. To fully consider reputation building and the implications for the level of emissions and aid, we develop a two-stage Bayesian game in the next section. By considering this two-stage Bayesian game, we hope to provide an alternative explanation of why certain aid programs may fail.

[17] Existence follows from our assumptions that the utility functions are continuous, strictly concave, and defined on compact and convex strategy sets. Hereafter we also assume $r_W^s(0) < \overline{E}$, $r^n(\overline{E}) < y^n$, $r^n(0) > 0$, and $r_T^s(r^n(0)) > 0$, so that the equilibrium values of aid and emissions are interior to the strategy sets and can be characterized by the first-order necessary conditions.

## III. A TWO-STAGE BAYESIAN GAME OF AID AND EMISSIONS

Now consider the following two-stage game. In stage one, when the South's type is its own private information, and the North's estimate that the South is a weak type is $\alpha$, each Southern type chooses a level of emissions and the North chooses a level of aid. We denote these stage-one choices by $(A_1, E_{T1}, E_{W1})$. The emissions are short-lived in that they are dispersed by the start of stage two, and so do not directly affect stage-two welfare levels. Then in stage two, after the outcome of stage one is observed by all, each Southern type again chooses a level of emissions and the North chooses a level of aid. We denote these stage-two choices by $(A_2, E_{T2}, E_{W2})$. As is standard for multistage games of incomplete information, the equilibrium concept we employ is perfect Bayesian equilibrium.

We focus our analysis on the two common types of perfect Bayesian equilibria, separating and pooling. In a separating equilibrium, the Southern types choose different levels of emissions in stage one, so the observed outcome of that stage reveals the true Southern type to the North. In stage two, therefore, the countries play one of the static certainty games of Section II. The stage-two equilibrium is thus $(A_2, E_{T2}) = (A_T^*, E_T^*)$ if the South was revealed to be the tough type, or $(A_2, E_{W2}) = (A_W^*, E_W^*)$ if it was revealed to be the weak type. In a pooling equilibrium, however, the Southern types take the same action in stage one, so the observed outcome of that stage reveals nothing to the North. Because the North's uncertainty about the true type of South persists in stage two, in this case the countries play the static Bayesian game of Section II with the equilibrium outcome $(A_2, E_{W2}, E_{T2}) = (A^*(\alpha), E_T^*(\alpha), E_W^*(\alpha))$.

Consider the possibility of a pooling equilibrium. In such an equilibrium, one Southern type imitates the other by choosing the same level of emissions in stage one, so the North's uncertainty about the South persists in stage two. The imitator necessarily suffers a loss in stage one because to imitate it must choose a level of emissions that does not maximize its stage-one utility. That is, building a reputation as another type is costly because it requires taking a non-optimal action. Neither Southern type will use such a strategy unless it gains from having the uncertainty about its type persist in stage two. Recalling from Corollary 1 that only the tough type gains from uncertainty, it follows that if there is a pooling equilibrium, then it is one in which the tough type imitates the weak type. Hence, strategic, reputation-building behavior in the form of pooling can only serve to increase the level of emissions.

A natural pooling equilibrium to consider is one in which both types choose the level of emissions in stage one that maximizes the current utility of the weak type, $E_{W1} = E_{T1} = E_W^*(\alpha)$.[18] The North's choice of aid in stage one is its (certain) best response to this, $r^n(E_W^*(\alpha))$. Because the North learns nothing about the South's type from observing the stage-one outcome, its stage-two belief that the South is weak remains $\alpha$. Then from Theorem 1, the stage-two equilibrium is $(A^*(\alpha), E_T^*(\alpha))$ if the South is tough and $(A^*(\alpha), E_W^*(\alpha))$ if it is weak. Given a

[18] We call this pooling equilibrium "natural" because $E_W^*(\alpha)$ is the level of emissions that maximizes the stage-one utility of the weak type if it cannot prevent pooling. However, there are many other pooling equilibria. For example, for any level of emissions $E_1$ "close enough" to $E_W^*(\alpha)$, the same argument that follows can be used to show there is a pooling equilibrium with emissions $E_1$ in stage one, given beliefs as in (3) below that support $E_1$.

Southern discount factor $\rho^s$, the tough type's total discounted payoff from pooling is $V_T^s(r^n(E_W^*(\alpha)), E_W^*(\alpha)) + \rho^s V_T^s(A^*(\alpha), E_T^*(\alpha))$. If, instead, it does not try to imitate, then its true type is revealed and the stage-two equilibrium is $(A_T^*, E_T^*)$. In stage one, it chooses the level of emissions that maximizes its static utility, $E_T^*(\alpha)$, so its total discounted payoff is $V_T^s(A^*(\alpha), E_T^*(\alpha)) + \rho^s V_T^s(A_T^*, E_T^*)$. Hence, a necessary condition for a pooling equilibrium in which the tough type imitates the weak type is

$$\rho^s\left[V_T^s(A^*(\alpha), E_T^*(\alpha)) - V_T^s(A_T^*, E_T^*)\right] \geq V_T^s(A^*(\alpha), E_T^*(\alpha)) - V_T^s\left(r^n(E_W^*(\alpha)), E_W^*(\alpha)\right), \quad (2)$$

which simply says that the discounted future gain from maintaining the uncertainty about its type exceeds the current cost of imitating the weak type.

For this pooling equilibrium, we must also specify Northern beliefs in stage two that are consistent with the stage-one actions taken by both Southern types. In stage two, the North's updated, posterior belief that the South is the weak type depends, in general, upon the level of emissions observed in stage one, $E$. We denote this posterior by $\mu(E)$. A standard specification of consistent beliefs for such a pooling equilibrium is

$$\mu(E) = \alpha \quad \text{if } E = E_W^*(\alpha) \qquad \text{and} \qquad \mu(E) = 0 \quad \text{otherwise.} \quad (3)$$

Under these beliefs, the tough type imitates the weak type if and only if it chooses the same level of emissions $E_W^*(\alpha)$ in stage one. For any other observed level of emissions, including off-the-equilibrium-path mistakes, the North assumes that the South is the tough type with certainty in stage two.

THEOREM 2. *Suppose that* (2) *holds, so the tough type can gain from imitating the weak type. Then the strategies* $(A^*(\alpha), E_T^*(\alpha), E_W^*(\alpha))$ *in stage one,* $(A^*(\alpha), E_W^*(\alpha))$ *in stage two if South is the weak type, and* $(A^*(\alpha), E_T^*(\alpha))$ *in stage two if South is the tough type, and the beliefs in* (3), *constitute a perfect Bayesian equilibrium with pooling.*

Now consider the possibility of separation. The natural separating equilibrium is one in which stage-one strategies are those in the static game, $(A_1, E_{T1}, E_{W1}) = (A^*(\alpha), E_T^*(\alpha), E_W^*(\alpha))$, and stage-two strategies are then $(A_2, E_{T2}) = (A_T^*, E_T^*)$ if the South was revealed to be tough or $(A_2, E_{W2}) = (A_W^*, E_W^*)$ if it was revealed to be weak. Again, we need a specification of Northern beliefs which are consistent with separation. Standard separating beliefs in this model are

$$\mu(E) = 1 \quad \text{if } E = E_W^*(\alpha) \qquad \text{and} \qquad \mu(E) = 0 \quad \text{otherwise.} \quad (4)$$

Under these beliefs, the weak type reveals itself by choosing $E_W^*(\alpha)$ and the tough type reveals itself by choosing anything other than $E_W^*(\alpha)$, including $E_T^*(\alpha)$. The tough type's total discounted payoff when it chooses $E_T^*(\alpha)$ is thus $V_T^s(A^*(\alpha), E_T^*(\alpha)) + \rho^s V_T^s(A_T^*, E_T^*)$. These beliefs also specify that for any other (off-the-equilibrium-path) observed level of emissions, the North assumes that the South is the tough type in stage two. Hence, if the tough type did choose $E_W^*(\alpha)$, then under (4) it would "completely fool" the North into believing that it is the tough type in stage two. In this case, the outcome in stage two would be the Northern level of aid $A_W^*$, and the tough type's best response to this, $r_T^s(\mathrm{A}_W^*)$. The tough type's total discounted payoff in this case is $V_T^s(A^*(\alpha), E_W^*(\alpha)) + \rho^s V_T^s(A_W^*, r_T^s(A_W^*))$. To eliminate the tough type's incentive to imitate the weak type in a separating equilibrium, we therefore need

$$\begin{aligned} &V_T^s(A^*(\alpha), E_T^*(\alpha)) - V_T^s(A^*(\alpha), E_W^*(\alpha)) \\ &\qquad \geq \rho^s\left[V_T^s(A_W^*, r_T^s(A_W^*)) - V_T^s(A_T^*, E_T^*)\right], \end{aligned} \tag{5}$$

so the tough type's current cost of imitation by choosing the weak type's level of emissions exceeds the discounted future gain from completely fooling the North.

THEOREM 3. *Suppose that* (5) *holds, so the tough type cannot gain from imitating the weak type and completely fooling the North at* $E_W^*(\alpha)$. *Then the strategies* $(A^*(\alpha), E_T^*(\alpha), E_W^*(\alpha))$ *in stage one,* $(A_T^*, E_T^*)$ *in stage two if the South is tough, and* $(A_W^*, E_W^*)$ *in stage two if the South is weak, and the beliefs in* (4), *constitute a perfect Bayesian equilibrium with separation.*

Note that the condition to prevent the tough type from imitating and to support a separating equilibrium is not just the converse of the condition to allow the tough type to imitate and support a pooling equilibrium. The reason is that the effect of imitation on Northern beliefs is different in these types of equilibria. Imitation merely maintains the North's uncertainty about the South's true type in a pooling equilibrium, $\mu = \alpha$, but would fool the North into believing the South is the weak type in a separating equilibrium, $\mu = 1$. As a result, the current cost and future benefit associated with imitation also differ in the two types of equilibria, and so (5) is not the converse of (2). In fact, it is simple to rank these current costs and future benefits.

THEOREM 4. *The current cost and future gain of imitation are both higher in the separating equilibrium of Theorem* 3 *than the pooling equilibrium of Theorem* 2. *That is*:

(i) $V_T^s(A^*(\alpha), E_T^*(\alpha)) - V_T^s(A^*(\alpha), E_W^*(\alpha)) > V_T^s(A^*(\alpha), E_T^*(\alpha)) - V_T^s(r^n(E_W^*(\alpha)), E_W^*(\alpha))$ *and*

(ii) $V_T^s(A_W^*, r_T^s(A_W^*)) - V_T^s(A_T^*, E_T^*) > V_T^s(A^*(\alpha), E_T^*(\alpha)) - V_T^s(A_T^*, E_T^*)$.

Both the incentive to imitate and its associated cost are lower in the pooling equilibrium than in the separating equilibrium. Thus, we cannot conclude that one of these equilibria is more likely than the other. We also cannot conclude that they are mutually exclusive. It is therefore possible that both the pooling equilibrium and the separating equilibrium can simultaneously occur for the same parametric values.

Finally, another outcome of interest is that the weak type might prevent pooling by choosing a stage-one level of emissions that is high enough to force a separating equilibrium.[19] For example, suppose that (5) does not hold, so separation does not occur if the weak type chooses $E_W^*(\alpha)$ in stage one. Then it is possible that the weak type can distinguish itself by choosing an even higher level of emissions in stage one, say $E_{W1} > E_W^*(\alpha)$, and thereby increase the tough type's cost of imitation to a prohibitive level. Of course, the weak type also suffers a stage-one

[19] Again, note that such a forcing equilibrium can generally be constructed for any pooling equilibrium, not just the one where the stage-one common level of emissions is $E_W^*(\alpha)$.

loss of utility from this deviation from static utility maximization. However, if this strategy succeeds and separation occurs, then there is a discounted stage-two gain for the weak type from having its type revealed that may outweigh the stage-one loss.

In stage one of such a separating equilibrium, the Southern weak type chooses emissions $E_{W1} > E_W^*(\alpha)$, and the best replies of the North and the Southern tough type are $A_1 = A^*(\alpha; E_{W1})$ and $E_{T1} = E_T^*(\alpha; E_{W1})$.[20] The stage-two strategies are again $(A_2, E_{T2}) = (A_T^*, E_T^*)$ if the South was revealed to be tough, or $(A_2, E_{W2}) = (A_W^*, E_W^*)$ if it was revealed to be weak. The standard specifications of Northern beliefs consistent with this separating equilibrium are

$$\mu(E) = 1 \quad \text{if } E = E_{W1} \qquad \text{and} \qquad \mu(E) = 0 \quad \text{otherwise}. \tag{6}$$

One necessary condition for this equilibrium is

$$\begin{aligned} V_T^s(A^*(\alpha; E_{W1}), E_T^*(\alpha; E_{W1})) - V_T^s(A^*(\alpha; E_{W1}), E_{W1}) \\ \geq \rho^s\left[V_T^s(A_W^*, r_T^s(A_W^*)) - V_T^s(A_T^*, E_T^*)\right], \end{aligned} \tag{7}$$

so the current cost of imitation for the tough type exceeds the discounted future gain at these levels of emissions and aid. Note that there exists a level of emissions $E_{\min} > E_W^*(\alpha)$ such that (7) holds with equality at $E_{W1} = E_{\min}$ and with strict inequality for $E_{W1} > E_{\min}$.[21] The other necessary condition is

$$\begin{aligned} \rho^s\left[V_W^s(A_W^*, E_W^*) - V_W^s(A_T^*, r_W^s(A_T^*))\right] \\ \geq V_W^s(A^*(\alpha), E_W^*(\alpha)) - V_W^s(A^*(\alpha; E_{W1}), E_{W1}), \end{aligned} \tag{8}$$

so the discounted future gain of forcing separation for the weak type exceeds its current cost at these levels of emissions and aid. We assume there exists an $E_{\max} > E_W^*(\alpha)$ such that (8) does not hold for any $E_{W1} > E_{\max}$.[22]

THEOREM 5. *Assume that* (5) *does not hold, so the tough type could gain from imitating the weak type and completely fooling the North at* $E_W^*(\alpha)$. *Also assume that* $E_{\min} \leq E_{\max}$, *so* (7) *and* (8) *both hold for* $E_{W1} \in [E_{\min}, E_{\max}]$, *and let* $E_{W1}^*$ *minimize* $V_W^s(A^*(\alpha), E_W^*(\alpha)) - V_W^s(A^*(\alpha; E_{W1}), E_{W1})$ *for* $E_{W1} \in [E_{\min}, E_{\max}]$. *Then the strategies* $(A^*(\alpha; E_{W1}^*), E_T^*(\alpha; E_{W1}^*), E_{W1}^*)$ *in stage one,* $(A_T^*, E_T^*)$ *in stage two if the South is tough, and* $(A_W^*, E_W^*)$ *in stage two if the South is weak, and the beliefs in* (6) *with* $E_{W1} = E_{W1}^*$, *constitute a perfect Bayesian equilibrium with separation where* $E_{W1}^* > E_W^*(\alpha)$.

[20] If the weak type chooses $E_{W1}$, then the best replies of the tough type and the North are determined in a manner similar to that used to prove Theorem 1. See Section V of the Appendix.

[21] This follows from the fact that (7) cannot hold at $E_{W1} = E_W^*(\alpha)$ if (5) does not hold, but the left-hand side of (7) is increasing in $E_{W1}$ for $E_{W1} \geq E_W^*(\alpha)$.

[22] Notice (8) holds with strict inequality at $E_{W1} = E_W^*(\alpha)$, where its right-hand side, the weak type's current cost of forcing separation, is 0. However, this cost is not necessarily increasing in $E_{W1}$ for all $E_{W1} \geq E_W^*(\alpha)$ because an increase in $E_{W1}$ not only decreases the weak type's current utility directly, but it also increases it indirectly by increasing Northern aid, $A^*(\alpha; E_{W1})$. Nevertheless, it seems evident that increases in $E_{W1}$ must eventually increase the current cost of forcing separation to a prohibitive level, so such an $E_{\max}$ exists. We have constructed a numerical example based on quadratic utility functions in which this occurs (the details of this example are available from the authors upon request).

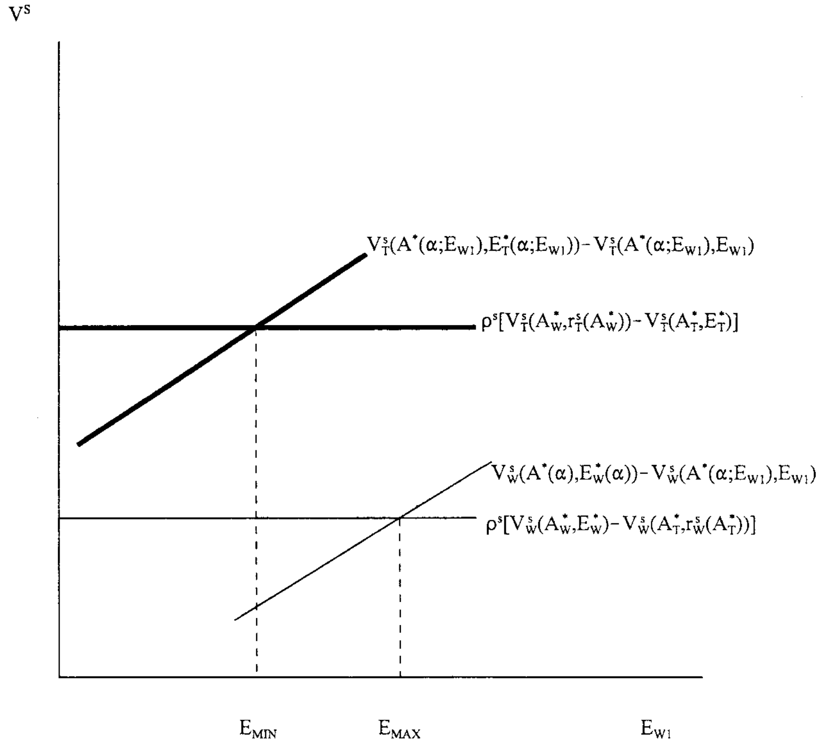


**FIG. 2.** Forced separating equilibrium.

This type of separating equilibrium is depicted in Fig. 2, where we assume the current loss from forcing separation for the weak type is increasing in $E_{W1}$ for $E_{W1} > E^*_W(\alpha)$, and so is less than the future gain for $E_{W1} < E_{\max}$. These costs are also drawn so that $E_{\min} < E_{\max}$, and that, at $E_{\min}$, the current cost of imitation for the tough type exceeds the discounted future gain. Finally, these are drawn so the cost of imitation for the tough type is low enough to allow imitation at $E_{W1} = E^*_W(\alpha)$, so the separation result of Theorem 3 cannot hold in this case.

This result shows that one apparently perverse consequence of aid programs is that there may be circumstances in which a politically weak country has an incentive to excessively increase its emissions just to prove that it really is weak. This strategy of excessive pollution to prove its type significantly reduces the effectiveness of aid. Notice, however, that Southern policy failure is not the source of this ineffectiveness of aid. On the contrary, the Southern weak type is acting optimally by excessively polluting in order to guarantee that the North learns its type, because this also guarantees that the North will provide the higher level of aid associated with a weak type in the future.

## IV. POLICY IMPLICATIONS AND CONCLUSIONS

Since low incomes may be a factor in low environmental quality, pollution may be reduced through the use of aid. However, environmental aid programs often yield uncertain outcomes and many programs are only partially successful. We argue that the uncertainty of the recipient nation's political ability or willingness to abate pollution is an important factor in determining the effectiveness of untied aid related to pollution abatement. Our analysis suggests that policies offering aid that is not tied to observable levels of emissions can be misguided when such uncertainty exists because they induce environment destruction as an optimal policy implementation of the recipient nations due to reputational effects. We model the aid/emissions problem as a two-stage game in which the North has incomplete information regarding the South's ability or political will to enforce emission standards. We demonstrate the South gains from uncertainty if it is a tough type, and so has an incentive to excessively pollute in stage one in order to mimic the weak type and prevent the uncertainty from being resolved for the North in stage two. We provide necessary and sufficient conditions for the existence of both pooling and separating equilibria. We also provide conditions for a separating equilibrium that the weak type forces by excessive pollution. Perversely, aid is the source of higher emission levels in both the pooling and forced separating equilibria.

Bohn and Deacon [6] and Deacon [12] have found empirical evidence that political instability is a source of environmental degradation. Our results are consistent with these findings since the prevalence of political instability implies that uncertainty and the potential for higher emissions may return even after the initial uncertainty has been resolved. For example, consider a stochastic, repeated game in which the constituent game is the two-stage game of Section III. We can think of the two stages in the constituent game as the length of time between elections in the South. Suppose this repeated game begins with nature making a draw $\alpha^*$ from a (common knowledge) distribution $F$ on the unit interval, where $\alpha^*$ is the true probability that a randomly selected Southern government is weak. Suppose no players observe $\alpha^*$, so each estimates it by the mean of $F$. At the end of the first constituent game, after the true type of the first Southern government is observed, the common estimate of $\alpha^*$ is updated according to Bayes rule. In each constituent game thereafter, the state (history of the repeated game) is the current estimate that the South is weak.[23] As long as $\alpha^* \in (0, 1)$, the true Southern type is never known with certainty after any election, so the incentive for excessive emissions as dissipative signaling arises again after every election. Admittedly, such a model of the Southern electoral process is very abstract. However, we think this indicates one possible approach to constructing a more realistic model in which shocks to the political system that re-introduce such uncertainty occur at random intervals.

Perhaps the most important lesson of our analysis is that untied environmental aid can be counter-productive in the presence of recurrent uncertainty (whatever its source) about the will of the recipient nations to abate emissions. Moreover, related extensions of this model suggest that other environmental aid programs such as debt-for-debt swaps, which occur in an environment with uncertainty, may encourage greater deforestation. As a result, we agree with a referee that the most

[23] A natural choice is the beta-Bernoulli family of conjugate distributions (see, for example, DeGroot [14], Chapter 6). That is, if $F$ is a beta distribution with parameters $(a, b)$, then the initial estimate of $\alpha^*$ is $\alpha_1 = a/(a + b)$. Given any current estimate $\alpha_t$ in the $t^{\text{th}}$ constituent game, the posteriors in the $(t + 1)^{\text{th}}$ game are $\alpha_{t+1} = [(a + b)\alpha_t + 1]/(a + b + 1)$ if a weak type was observed in the $t^{\text{th}}$ game, and $\alpha_{t+1} = (a + b)\alpha_t/(a + b + 1)$ if a tough type was observed.

important extension of our work is to try to determine Northern aid policies that eliminate this incentive for excessive emissions as reputation-building behavior.

## APPENDIX: PROOFS OF THEOREMS

### I. *Proofs of Theorem* 1 *and Corollary* 1

Under our assumptions on the payoffs, $(A^*(\alpha), E_T^*(\alpha), E_W^*(\alpha))$ is a Bayesian equilibrium if $\partial V_W^s(A^*(\alpha), E_W^*(\alpha))/\partial E_w = 0$, $\partial V_T^s(A^*(\alpha), E_T^*(\alpha))/\partial E_T = 0$, and $\partial N(A^*(\alpha), E_T^*(\alpha), E_W^*(\alpha))/\partial A = 0$. Using the definitions of the Southern reaction functions, $(A^*(\alpha), r_T^s(A^*(\alpha)), r_W^s(A^*(\alpha)))$ is a Bayesian equilibrium if there exists an $A^*(\alpha)$ such that $f(A^*(\alpha), \alpha) = 0$, where $f(A, \alpha) = \partial N(A, r_T^s(A), r_W^s(A))/\partial A = \alpha[\partial V^n(A, r_W^s(A))/\partial A] + (1-\alpha)[\partial V^n(A, r_T^s(A))/\partial A]$. First note that $f(A_T^*, \alpha) = \alpha[\partial V^n(A_T^*, r_W^s(A_T^*))/\partial A] > 0$ because $r_W^s(A_T^*) > r_T^s(A_T^*) = E_T^*$, $\partial V^n(A_T^*, E_T^*)/\partial A = 0$, and $\partial^2 V^n/\partial E\,\partial A > 0$. Similarly, $f(A_W^*, \alpha) = (1-\alpha)[\partial V^n(A_W^*, r_T^s(A_W^*))/\partial A] < 0$ as $r_T^s(A_W^*) < r_W^s(A_W^*) = E_W^*$, $\partial V^n(A_W^*, E_W^*)/\partial A = 0$, and $\partial^2 V^n/\partial E\,\partial A > 0$. Because $\partial^2 N/\partial A^2 < 0$, $\partial^2 V^n/\partial E\,\partial A > 0$, $\partial r_W^s/\partial A < 0$, and $\partial r_T^s/\partial A < 0$, we have

$$\begin{aligned}\partial f/\partial A = {} & \partial^2 N(A, r_T^s(A), r_W^s(A))/\partial A^2 \\ & + \alpha\left[\partial^2 V^n(A, r_W^s(A))/\partial A\,\partial E\right](\partial r_W^s/\partial A) \\ & + (1-\alpha)\left[\partial^2 V^n(A, r_T^s(A))/\partial A\,\partial E\right](\partial r_T^s/\partial A) < 0.\end{aligned}$$

So, for any $\alpha \in (0,1)$, there exists a unique $A^*(\alpha) \in (A_T^*, A_W^*)$ such that $f(A^*(\alpha), \alpha) = 0$, and thus a unique equilibrium. Moreover, $\partial f/\partial A < 0$ implies that this equilibrium is locally stable (in the Routh–Hurwicz sense). Because $\partial r_W^s/\partial A < 0$, $\partial r_T^s/\partial A < 0$, $r_T^s(A) < r_W^s(A)$, and $A_T^* < A^*(\alpha) < A_W^*$, we have $E_T^*(\alpha) < E_T^* < E_W^* < E_W^*(\alpha)$. Next, because $r_T^s(A) < r_W^s(A)$ and $\partial^2 V^n/\partial E\,\partial A > 0$ imply $\partial f/\partial \alpha > 0$, $A^*(\alpha)$ is an increasing function of $\alpha$. This, $\partial r_W^s/\partial A < 0$, and $\partial r_T^s/\partial A < 0$ imply $E_T^*(\alpha)$ and $E_W^*(\alpha)$ are decreasing functions of $\alpha$. This proves Theorem 1.

Evaluated on $r_T^s(A)$, the tough type's utility is an increasing function of $A$. That is, $\partial V_T^s(A, r_T^s(A))/\partial A = \partial V_T^s(A, r_T^s(A))/\partial A + [\partial V_T^s(A, r_T^s(A))/\partial E](\partial r_T^s/\partial A) = \partial V_T^s(A, r_T^s(A))/\partial A > 0$ because $\partial V_T^s(A, r_T^s(A))/\partial E = 0$. Thus, if the South is a tough type, then uncertainty increases its utility because $A^*(\alpha) > A_T^*$. And because $\partial V_W^s/\partial A > 0$ along $r_W^s$ and $A^*(\alpha) < A_W^*$, uncertainty decreases the South's utility if it is a weak type. This proves Corollary 1.

### II. *Proof of Theorem* 2

For this two-stage game, a perfect Bayesian equilibrium is a specification of strategies for each player at each stage, $(A_1, E_{T1}, E_{W1})$ and $(A_2, E_{T2}, E_{W2})$, and a specification of Northern stage-two beliefs that the South is the weak type, conditional on the outcome of stage one, $\mu(E)$, such that: the strategies are optimal, given these beliefs, and the beliefs are obtained from these strategies and observed actions in stage one using Bayes' rule whenever possible.

First consider stage two. If the observed action in stage one was $E_W^*(\alpha)$, then from (3) the North's stage-two belief that the South is weak is $\mu(E_W^*(\alpha)) = \alpha$. The stage-two game is then the static Bayesian game of Section II with unique equilibrium $(A^*(\alpha), E_T^*(\alpha), E_W^*(\alpha))$. Now consider stage one. Given that both types use $E_W^*(\alpha)$, the North's optimal strategy is to use its best response to that, $r^n(E_W^*(\alpha))$ (any other choice reduces its stage-one utility, and has no impact on stage-two utility). Next, given Northern aid $r^n(E_W^*(\alpha))$, any deviation from $E_W^*(\alpha)$ by the weak type reduces its stage-one utility because $V_W^s(r^n(E_W^*(\alpha)), E_W^*(\alpha)) > V_W^s(r^n(E_W^*(\alpha)), E_{W1})$ for $E_{W1} \neq E_W^*(\alpha)$. Moreover, under (3), a deviation from $E_W^*(\alpha)$ leads the North to believe the South is tough with certainty in stage two and choose the level of aid $A_T^*$. The best the weak type can do in response to this is $r_W^s(A_T^*)$, so its stage-two utility is $V_W^s(A_T^*, r_W^s(A_T^*))$. Thus, a deviation from $E_W^*(\alpha)$ also reduces the weak type's stage-two utility because, as in the proof of Corollary 1, $A_T^* < A^*(\alpha)$ and $E_W^*(\alpha) = r_W^s(A^*(\alpha))$ imply that $V_W^s(A_T^*, r_W^s(A_T^*)) < V_W^s(A^*(\alpha), E_W^*(\alpha))$. Finally, if the tough type deviates from $E_W^*(\alpha)$, then it reveals its true type and the stage-two equilibrium is $(A_T^*, E_T^*)$. The best deviation it can make is thus to $E_W^*(\alpha)$, because this maximizes stage-one utility (any other deviation from $E_W^*(\alpha)$ reduces stage-one utility and has no effect on stage-two utility). The condition in (2) guarantees this deviation from $E_W^*(\alpha)$ to $E_T^*(\alpha)$ reduces the weak type's total discounted utility. Hence, given the beliefs in (3), these strategies are optimal.

Given these strategies, the North always observes $E_W^*(\alpha)$ in stage one for either Southern type. Because this conveys no information, using Bayes rule to update its estimate that the South is the weak type gives $\mu(E_W^*(\alpha)) = \alpha/[\alpha + (1 - \alpha)] = \alpha$. Also, as is well known, any posterior beliefs are admissible when a level of emissions other than $E_W^*(\alpha)$ is observed in stage one, because such an action has 0 probability under these strategies, and Bayes rule gives no information on how to update events that occur with 0 probability. Hence, the beliefs in (3) are consistent with these strategies.

### III. *Proof of Theorem* 3

Consider stage two. If the observed action in stage one was $E_W^*(\alpha)$, then from (4) the North's stage-two belief that the South is weak is $\mu(E_W^*(\alpha)) = 1$. The stage-two game is then the static certainty game with the unique equilibrium $(A_W^*, E_W^*)$. Similarly, if the observed action in stage one was $E_T^*(\alpha)$, then from (4) the North's stage-two belief that the South is tough is $1 - \mu(E_T^*(\alpha)) = 1$. The stage-two game is then the other static certainty game with the unique equilibrium $(A_T^*, E_T^*)$. Now consider stage one. Given that the weak type uses $E_W^*(\alpha)$ and the tough type uses $E_T^*(\alpha)$, the North's optimal strategy is to use its Bayesian best response $R(E_W^*(\alpha), E_T^*(\alpha)) = A^*(\alpha)$ (any other choice reduces its stage-one utility, and has no impact on its stage-two utility). Given Northern aid $A^*(\alpha)$, any deviation from $E_W^*(\alpha) = r_W^s(A^*(\alpha))$ reduces the weak type's stage-one utility. Moreover, under (4), such a deviation leads the North to believe the South is tough with certainty in stage two, and thus to choose aid $A_T^*$. The best the weak type can do then is its best response $r_W^s(A_T^*)$. Thus, any deviation by the weak type also reduces its stage-two utility because, as in the proof of Corollary 1, $A_T^* < A_W^*$ and $r_W^s(A_T^*) > E_W^* = r_W^s(A_W^*)$ imply $V_W^s(A_W^*, E_W^*) > V_W^s(A_T^*, r_W^s(A_T^*))$. Finally, given aid $A^*(\alpha)$, any deviation from $E_T^*(\alpha) = r_T^s(A^*(\alpha))$ reduces the tough type's stage-one utility. If this deviation is to any level of emissions other than $E_W^*(\alpha)$, then there is no change in the tough type's stage-two utility, because under (4) the North believes the South is tough with certainty in stage two for such a deviation. If the deviation is to $E_W^*(\alpha)$, then under (4) the North believes the South is weak with certainty in stage two. But in this case the tough type's total discounted utility from deviating to $E_W^*(\alpha)$ cannot be higher given that (5) holds. Hence, given the beliefs in (4), these strategies are optimal.

Given these strategies, the North always observes $E_W^*(\alpha)$ in stage one if the South is the weak type and $E_T^*(\alpha)$ if the South is the tough type. Thus, using Bayes' rule to update its estimate that the South is the weak type gives $\mu(E_W^*(\alpha)) = 1$ and $\mu(E_T^*(\alpha)) = 0$. Again, any posterior beliefs are admissible when a level of emissions other than $E_W^*(\alpha)$ or $E_T^*(\alpha)$ is observed in stage one because such an action has 0 probability under these strategies. Thus, the beliefs in (4) are consistent with these strategies.

### *IV. Derivation of Best Replies to an Arbitrary Choice* $E_{W1}$

For any $E_{W1}$, let

$$g(A, E_{W1}) = \alpha[\partial V^n(A, E_{W1})/\partial A] + (1-\alpha)[\partial V^n(A, r_T^s(A))/\partial A].$$

Recall from the proof of Theorem 1 that $\partial V^n(A_T^*, r_W^s(A_T^*))/\partial A > 0$, so $g(A_T^*, E_{W1}) = \alpha[\partial V^n(A_T^*, r_W^s(A_T^*))/\partial A] > 0$. Also recall that $\partial V^n(A_W^*, r_T^s(A_W^*))/\partial A < 0$, so $\partial V^n(A, r_T^s(A))/\partial A < 0$ for $A \geq A_W^*$, and that $\partial V^n(r^n(E_{W1}), E_{W1})/\partial A = 0$ by the definition of $r^n(E)$, so $\partial V^n(A, E_{W1})/\partial A < 0$ for $A > r^n(E_{W1})$. Hence, for any given $E_{W1}$, $g(A, E_{W1}) < 0$ for $A > \max\{r^n(E_{W1}), A_W^*\}$. Finally note that

$$\begin{aligned}\partial g/\partial A = \alpha&[\partial^2 V^n(A, E_{W1})/\partial A^2]\\ &+(1-\alpha)\{[\partial^2 V^n(A, r_T^s(A))/\partial A^2]\\ &\qquad+[\partial^2 V^n(A, r_T^s(A))/\partial A\,\partial E](\partial r_T^s/\partial A)\} < 0\end{aligned}$$

because $\partial^2 V^n/\partial A^2 < 0$, $\partial^2 V^n/\partial E\,\partial A > 0$, and $\partial r_T^s/\partial A < 0$. Thus, there exists a unique $A^*(\alpha; E_{W1}) \in (A_T^*, \max\{r^n(E_{W1}), A_W^*\})$ such that $g(A^*(\alpha; E_{W1})) = 0$. The best replies for the North and the Southern tough type are then $A^*(\alpha; E_{W1})$ and $E_T^*(\alpha; E_{W1}) = r_T^s(A^*(\alpha; E_{W1}))$. Finally, it is worth noting that

$$\begin{aligned}\partial A^*(\alpha; E_{W1})/\partial E_{W1} &= -(\partial g/\partial E_{W1})/(\partial g/\partial A)\\ &= -\alpha[\partial^2 V^n(A, E_{W1})/\partial A\,\partial E]/(\partial g/\partial A) > 0\end{aligned}$$

and

$$\partial E_T^*(\alpha; E_{W1})/\partial E_{W1} = [\partial r_T^s(A^*(\alpha; E_{W1}))/\partial A^*][\partial A^*(\alpha; E_{W1})/\partial E_{W1}] < 0.$$

### *V. Proof of Theorem* 4

Recall from Theorem 1 that $E_W^*(\alpha) > E_W^*$ and $A_W^* > A^*(\alpha)$. This and $\partial r^n/\partial E > 0$ imply that $r^n(E_W^*(\alpha)) > A^*(\alpha)$, and so $V_T^s(r^n(E_W^*(\alpha)), E_W^*(\alpha)) > V_T^s(A^*(\alpha), E_W^*(\alpha))$ because $\partial V_T^s/\partial A > 0$. This proves (i). Similarly, because $A_W^*$

$> A^*(\alpha)$ and $\partial[V_T^s(A, r_T^s(A))]/\partial A > 0$ as shown in the proof of Corollary 1, it follows that $V_T^s(A_W^*, r_T^s(A_W^*)) > V_T^s(A^*(\alpha), r_T^s(A^*(\alpha))) = V_T^s(A^*(\alpha), E_T^*(\alpha))$, which proves (ii).

## *VI. Proof of Theorem* 5

First we show that (7) holds for all $E \geq E_{\min}$. As noted in the text, (7) cannot hold for $E_{W1} = E_W^*(\alpha)$. Now note that differentiating the left-hand side of (7) with respect to $E_{W1}$ gives

$$\begin{aligned}&\{[\partial V_T^s(A^*(\alpha; E_{W1}), E_T^*(\alpha; E_{W1}))/\partial A] - [\partial V_T^s(A^*(\alpha; E_{W1}), E_{W1})/\partial A]\}\\&\qquad\times[\partial A^*(\alpha; E_{W1})/\partial E_{W1}] - \partial V_T^s(A^*(\alpha; E_{W1}), E_{W1})/\partial E\end{aligned}$$

(where we have used the fact that $\partial V_T^s(A^*(\alpha; E_{W1}), E_T^*(\alpha; E_{W1}))/\partial E = 0$ by the definition of the best replies to an arbitrary $E_{W1}$). Also note that $\partial^2 V_T^s/\partial A\, \partial E < 0$ and $E_{W1} > E_T^*(\alpha; E_{W1})$ imply the term in braces is positive, while from IV above we have $\partial A^*(\alpha; E_{W1})/\partial E_{W1} > 0$. Because $A^*(\alpha; E_W^*(\alpha)) = A^*(\alpha)$, we have $\partial V_T^s(A^*(\alpha; E_{W1}), E_{W1})/\partial E = 0$ at $E_{W1} = E_W^*(\alpha)$, and therefore $\partial V_T^s(A^*(\alpha; E_{W1}), E_{W1})/\partial E < 0$ for $E_{W1} > E_W^*(\alpha)$ because $\partial^2 V_T^s/\partial E^2 < 0$. That is, the left-hand side of (7) is increasing in $E_{W1}$ for $E_{W1} \geq \mathrm{E}_W^*(\alpha)$. Because the right-hand side of (7) does not depend on $E_{W1}$, it follows that there exists a unique $E_{\min} > E_W^*(\alpha)$ such that (7) holds with equality at $E_{\min}$ and strictly for $E_{W1} > E_{\min}$.

Differentiating the right-hand side of (8) with respect to $E_{W1}$ gives

$$\begin{aligned}&-[\partial V_W^s(A^*(\alpha; E_{W1}), E_{W1})/\partial A][\partial A^*(\alpha; E_{W1})/\partial E_{W1}]\\&\qquad-[\partial V_W^s(A^*(\alpha; E_{W1}), E_{W1})/\partial E],\end{aligned}$$

where $\partial V_W^s/\partial A > 0$ and $\partial A^*(\alpha; E_{W1})/\partial E_{W1} > 0$. However, because $\partial V_W^s(A^*(\alpha; E_{W1}), E_{W1})/\partial E = 0$ at $E_{W1} = E_W^*(\alpha)$ and $\partial^2 V_W^s/\partial E^2 < 0$, we have $\partial V_W^s(A^*(\alpha; E_{W1}), E_{W1})/\partial E < 0$ for $E_{W1} > E_W^*(\alpha)$, so the right-hand side of (8) may be increasing or decreasing in $E_{W1}$ above $E_W^*(\alpha)$.

Under the assumption that there exists a largest level of emissions $E_{\max}$ such that (8) holds, and that $E_{\max} \geq E_{\min}$, the remainder of the proof is analogous to that of Theorem 2 with $A^*(\alpha; E_{W1}^*)$, $E_T^*(\alpha; E_{W1}^*)$, and $E_{W1}^*$ replacing $A^*(\alpha)$, $E_T^*(\alpha)$, and $E_W^*(\alpha)$. Consider stage two. If the observed action in stage one was $E_{W1}^*$, then from (6) the North's stage-two belief that the South is weak is $\mu(E_{W1}^*) = 1$, and the unique equilibrium is $(A_W^*, E_W^*)$. Similarly, if the observed action in stage one was $E_T^*(\alpha)$, then from (6) the North's stage-two belief that the South is tough is $1 - \mu(E_T^*(\alpha)) = 1$, and the unique equilibrium is $(A_T^*, E_T^*)$.

Now consider stage one. Given that the weak type uses $E_{W1}^*$, the optimal strategies for the tough type and the North are $E_T^*(\alpha; E_{W1}^*)$ and $A^*(\alpha; E_{W1}^*)$, their best replies to $E_{W1}^*$. Given this Northern aid, any deviation in stage-one emissions from $E_{W1}^*$ to some other $E_{W1} \in [E_{\min}, E_{\max}]$ reduces the weak type's stage-one utility because, by definition, $E_{W1}^*$ minimizes $V_W^s(A^*(\alpha), E_W^*(\alpha)) - V_W^s(A^*(\alpha; E_{W1}), E_{W1})$ for $E_{W1} \in [E_{\min}, E_{\max}]$. Under (6), any deviation from $E_{W1}^*$ also leads the North to believe South is tough with certainty in stage two, and thus choose aid $A_T^*$. The best the weak type can do then is $r_W^s(\mathrm{A}_T^*)$, in which case its stage-two utility is also lower, $V_W^s(A_W^*, E_W^*) > V_W^s(A_T^*, r_W^s(A_T^*))$. Hence, any deviation to some other $E_{W1} \in [E_{\min}, E_{\max}]$ reduces the weak type's total discounted utility. A deviation to some $E_{W1} > E_{\max}$ must also decrease its total discounted utility because (8) does not hold for $E_{W1} > E_{\max}$. The weak type would do better by deviating from this $E_{W1}$ to $E_W^*(\alpha)$ and allowing pooling, so that North believes it is tough with certainty in stage two. Finally, a deviation to some $E_{W1} < E_{\min}$ would again lead the North to believe South is tough with certainty in stage two, and the weak type's total discounted payoff would again be maximized at $E_W^*(\alpha)$, which must be lower than the total discounted payoff at $E_{W1}^*$ by (8). Hence, given the beliefs in (6), these strategies are optimal.

Given these strategies, the North always observes $E_{W1}^*$ in stage one if the South is weak and $E_T^*(\alpha)$ if it is tough. Thus, using Bayes' rule to update its estimate that the South is weak gives $\mu(E_{W1}^*) = 1$ and $\mu(E_T^*(\alpha)) = 0$. Again, any posterior beliefs are admissible when a level of emissions other than $E_{W1}^*$ or $E_T^*(\alpha)$ is observed in stage one, because such an action has 0 probability under these strategies. Thus, the beliefs in (6) are consistent with these strategies.

## REFERENCES


1. A. Alesina and D. Dollar, "Who Gives Foreign Aid to Whom and Why?" NBER Working Paper W6612 (1998).
2. J. Andreoni, Cooperation in public-goods experiments: Kindness or confusion?, *Amer. Econom. Rev.* **85**, 891–904 (1995).
3. M. D. Baer and S. Weintraub, "The NAFTA Debate: Grappling with Unconventional Trade Issues," Rienner, Boulder, CO (1994).
4. S. Barrett, Self-enforcing international environmental agreements, *Oxford Econom. Pap.* **46**, 878–894 (1994).
5. W. J. Baumol and W. E. Oates, "The Theory of Environmental Policy," Cambridge Univ. Press, Cambridge, UK (1992).
6. H. Bohn and R. T. Deacon, Ownership risk, investment, and the use of natural resources, *Amer. Econom. Rev.* **90**, 526–549 (2000).
7. W. Buchholz and K. A. Konrad, Global environmental problems and the strategic choice of technology, *Z. Nationalökon.* **60**, 299–321 (1994).
8. B. Connolly, T. Gutner, and H. Bedarff, Organizational inertia and environmental assistance to Eastern Europe, *in* "Institutions for Environmental Aid" (R. O. Keohane and Marc A. Levy, Eds.), MIT Press, Cambridge, MA (1996).
9. B. Copeland, R. Sedjo and M. S. Taylor, North–South trade and the environment, *Quart. J. of Econom.* **109**, 755–787 (1994).
10. R. Cornes and T. Sandler, Easy rider, joint production, and public goods, *Econom. J.*, **94**, 580–598 (1984).
11. R. d'Arge, On the economics of transnational environmental externalities, *in* "Economic Analysis of Environmental Problems" (E. Mills, Ed.), Columbia Univ. Press, New York (1975).
12. R. T. Deacon, Deforestation and the rule of law in cross-section of countries, *Land Econom.* **70**, 414–430 (1994).
13. R. T. Deacon and P. Murphy, The structure of an environmental transaction: The debt for nature swap, *Land Econom.*, **73**, 1–24 (1997).
14. M. H. DeGroot, "Optimal Statistical Decisions," McGraw-Hill, New York (1970).
15. D. Fairman and M. Ross, Old fads, new lessons: Learning from economic development assistance, *in* "Institutions for Environmental Aid" (R. O. Keohane and M. A. Levy, Eds.), MIT Press, Cambridge, MA (1996).
16. G. M. Grossman and A. B. Krueger, Environmental impacts of a North American Free Trade Agreement, *in* "The Mexico–U.S. Free Trade Agreement" (P. M. Garber, Ed.), MIT Press, Cambridge, MA (1993).
17. D. Kahneman, and J. Knetsch, Valuing public goods: The purchase of moral satisfaction, *J. of Environ. Econom. Management* **22**, 57–70 (1992).

18. R. Kanbur, T. Sandler, and K. Morrison, The context of development assistance, *in* "The Future of Foreign Assistance: Common Pool Approach and International Public Goods," Overseas Development Council (1999).
19. R. E. Kohn, A General Equilibrium Analysis of the Optimal Number of Firms in a Polluting Industry, *Canad. J. of Econom.* **59**, 347–357 (1985).
20. R. E. Kohn, When subsidies for pollution abatement increase total emissions, *Southern Econom. J.* **59**, 77–87 (1992).
21. T. R. Lewis, Protecting the environment when costs and benefits are privately known, *Rand J. of Econom.*, **27**, 819–847 (1996).
22. R. López, Environment as a factor of production: The effects of economic growth and trade liberalization, *J. Environ. Econom. Management* **27**, 163–184 (1994).
23. J. A. McNeely, The future of national parks, *Environment* **32**, 16–20 and 36–41 (1990).
24. W. Niskanen, "Bureaucracy and Representative Government," Aldine, Chicago (1971).
25. E. Petrakis and A. Xepapadeas, Environmental consciousness and moral hazard in international agreements to protect the environment, *J. Public Econom.* **60**, 95–110 (1996).
26. A. M. Polinsky, Notes on the symmetry of taxes and subsidies in pollution control, *Canad. J. Econom.* **12**, 75–83 (1979).
27. C. S. Russell and P. T. Powell, "Choosing Environmental Policy Tools," Working Paper ENV-102, Inter-American Development Bank, Washington, D.C. (1996).
28. R. A. Sánchez, Health and environmental risks of the Maquiladora in Mexicali, *Natur. Resour. J.* **30**, 163–186 (1990).
29. H. Scheyvens, "The External Politics of ODA: A Case Study of Japan," Working Paper 96/2, Institute of Development Studies, Massey University, New Zealand (1996).
30. T. M. Selden and D. Song, Environmental quality and development: Is there a Kuznets curve for air pollution emissions?, *J. of Environ. Econom. Management* **27**, 147–162 (1994).
31. K. Veijo, K. Göran-Mäler and H. Tulkens, The acid rain game as a resource allocation process with an application to international cooperation among Finland, Russia and Estonia, *Scand. J. Econom.* **97**, 325–343 (1995).
32. R. Wade, Greening the bank: The struggle over the environment, 1970–1995, *in* "The World Bank in Its First Half Century–Volume 2: Perspectives" (D. Kapur, J. P. Lewis, and R. Webb, Eds.), The Brookings Institution, Washington, D.C. (1997).