

MORALLY-MOTIVATED SELF-REGULATION

David P. Baron

Northwestern University and Stanford University

October 2007

ABSTRACT

Some individuals and firms voluntarily mitigate the harmful consequences of their economic activities in situations in which they could free ride. In the context of a random matching model where citizens play a private provision of public goods game, this paper examines the scope of self-regulation motivated by altruism or warm glow preferences. Moral preferences are represented as stronger among neighbors than among strangers, and those preferences may be unconditional or reciprocal. The focus is on the role of organizations in increasing the scope of self-regulation. Social label, certification, and enforcement organizations are considered, as are public regulation and social pressure applied by an activist NGO funded by voluntary contributions by citizens. Social label and certification organizations can exist with reciprocal but not unconditional altruism, and they expand the scope of self-regulation by mitigating the free-rider problem, but their effect is limited. Enforcement organizations expand the scope of self-regulation for both unconditional and reciprocal altruism, and for-profit enforcement is more aggressive than non-profit enforcement. Voluntary self-regulation, however, can crowd out support for public regulation, weaken the demand for private organizations, and reduce contributions to fund social pressure.

Morally-Motivated Self-Regulation

David P. Baron

Northwestern University and Stanford University

October 2007

I. Introduction

Individuals take a variety of actions to mitigate externalities, redistribute wealth, and provide public goods that benefit others. Some have reduced their environmental impact, contributed to relief efforts, supported environmental NGOs, bought fair trade products, and shown a willingness to pay a premium for green products, such as green electricity. Organizations have also been formed to facilitate voluntary measures. Individuals can purchase offsets for the carbon footprint of their households through Climate Trust, Atmosfair, and NativeEnergy. Air travelers can purchase carbon offsets through Expedia and Travelocity. Firms have established programs for environmental protection, the sustainability of inputs, and the assurance of credence attributes of their products, and corporate social responsibility is increasingly embraced by firms. Google, for example, has pledged to become carbon neutral. Carbon trading is available on the Chicago Climate Exchange, working conditions in overseas factories are strengthened and monitored by the Fair Labor Association, wealth is transferred to growers and local producers through fair trade organizations, and private organizations and charities provide education and medical care to address pandemics such as HIV/AIDS. These activities are examples of self-regulation – the voluntary private provision of public goods – as an alternative to public provision and public regulation.

Self-regulation takes place outside the institutions of government and hence is in the realm of private rather than public politics.¹ Self-regulation can have a number of motivations. Self-regulation by a firm could be motivated by profit incentives as when it produces a green product because consumers are willing to pay a premium for it. A firm could also self-regulate to deter public regulation as in Lyon, Maxwell, and Hackett (2004) and Lyon and Maxwell (2004). A firm could also self-regulate to deter private politics, as in Baron (2007c) and Baron and Diermeier

¹ Public politics takes place in the institutions of government, whereas private politics occurs outside, but often in the shadow, of those institutions. Private politics pertains to individual and collective action to influence the conduct of private agents as in the case of NGOs that apply social pressure to change the conduct of firms.

(2007) where self-regulation can lead an activist to select a different target for a boycott. Some self-regulation thus can be explained by the private benefits from a citizen's actions and the supply by firms of products with attributes desired by citizens.

This paper focuses instead on self-regulation motivated by moral concerns. The context is one in which citizens or firms can voluntarily provide a local public good in the presence of incentives to free-ride. The public good may be thought of as pertaining to an environmental externality or to unobservable credence attributes of a product that consumers cannot learn through search, experience, or consumption. Such attributes could include the conditions under which a product is produced, including any unregulated environmental externalities associated with production, how well workers are treated and paid, how and where the product is marketed, hidden hazards associated with consumption of the product, how recyclable it is, and whether it is made from sustainable inputs.

The approach in this paper is to investigate self-regulation in a static rather than repeated setting. This may be thought of corresponding to changing circumstances in which long-run behavior is not relevant or where convergence requires a large number of repetitions. The approach is to consider the implications for the scope of self-regulation of alternative assumptions about motivation as represented by egoistic and other-regarding preferences for conduct in a free-rider setting. The focus is on the opportunities for public and particularly private organizations to increase the scope of self-regulation. One public organization is government regulation, but private self-regulation can crowd out public regulation (Calveras, Ganuza, and Llobet 2007). One type of private organization considered simply screens citizens through the use of social labels that allow citizens with similar preferences to interact among themselves. A second type is a certification organization that provides future trading partners with information about the past conduct of citizens. A third type of organization strengthens self-regulation by enforcing informal pledges to contribute to the public good. A fourth approach is to rely on for-profit firms to provide enforcement to expand the scope of self-regulation. Finally, expanding the scope of self-regulation could be externalized by relying on activists and NGOs to monitor the conduct of citizens and publicly disclose their failure to provide the public good. Although organizations can increase the private provision of the public good, voluntary self-regulation can crowd out support for public regulation, weaken the demand for private organizations, and reduce contributions to fund social pressure.

The economic actors considered can include firms that act based on moral preferences. Those preferences could come from the preferences of shareholders, as in Baron (2007a)(2007b)(2007c)

and Graff Zivin and Small (2005). Morally-based preferences could also reflect the preferences of the managers of the firm when there is a separation of ownership from control. A firm, for example, could undertake costly actions under the framework of corporate social responsibility that benefit a community or targeted recipients. Shareholders could also value those actions and sacrifice a financial return by accepting a lower return on the shares of the firm. Corporate social responsibility could vary depending on the industry or the goods produced by a firm and the associated external benefits and costs.²

The scope of self-regulation is characterized by identifying the factors that expand and those that limit it. The theory provides a framework that can explain why some societies deal with free-rider problems more effectively than others. The objective of this paper is more limited and pertains to identifying preference attributes that give rise to self-regulation and to organizations that expand the scope of self-regulation. More specifically, the paper relates the scope of self-regulation to the context in which the public goods game is embedded.³ Four contextual features are considered. The first is the direct costs and benefits from providing the public good. The second is the socioeconomic distance between the citizens involved in the public goods game. The third is the nature and extent of moral preferences. The fourth is organizations that affect the behavior of citizens and the opportunities available to them.

A necessary condition for the private provision of the public good in the model is the presence of moral preferences or warm glow preferences. Moral preferences are other-regarding and take the form of altruism in which a citizen takes into account the benefits she provides to others when she contributes to the public good. Warm glow preferences (Andreoni 1988, 1990) are egoistic and reflect satisfaction from the action of providing the public good. Moral preferences thus pertain to the well-being of others, whereas warm glow preferences reflect the private goods aspect of an action. Altruistic and warm glow preferences can have the same effect on behavior.

Moral preferences can be pure or impure and may be conditional on the actions of others or limited by socioeconomic distance. Generalized morality is a pure form of altruism in which a citizen's willingness to provide the local public good is independent of how distant a trading partner is, where distance could be geographic or socioeconomic. For example, a citizen could provide the

² Siegel and Vitaliano (2007) found that firms producing credence goods were more likely to engage in corporate social responsibility than firms producing search goods. They argue that this finding is consistent with a product differentiation strategy, but it is also consistent with the concept of self-regulation.

³ This is consistent with the perspective in List (2007) that context is important for understanding behavior in the laboratory and the real world.

public good regardless of whether her trading partner is a neighbor or stranger. Whether the trading partner is a neighbor or a stranger matters with limited morality. As modeled here, with limited altruism the utility of a citizen from providing the public good decreases the more distant is the trading partner. A citizen then may provide the public good when trading with a neighbor but not when trading with a stranger. The theory predicts that the scope of self-regulation is greater in close-knit than disparate societies and greater with generalized than limited morality.

Altruistic preferences could also be unconditional or conditional. Unconditional preferences are independent of the action of a trading partner, as in the case of pure altruism. The conditional preferences considered here are reciprocal in the sense that citizens care more about the benefits provided to a trading partner if the partner also provides the public good than if he does not.⁴ Reciprocal altruism results in a smaller scope of self-regulation in a heterogeneous citizenry, but it also provides an opportunity for organizations to expand self-regulation.

The form, in addition to the strength, of moral preferences matters for the extent to which organizations can expand the scope of self-regulation. Consider a social label organization that is open to all citizens for a fee. The purpose of the organization is to identify citizens with similar preferences so that they can interact with each other. For example, citizens can purchase fair trade products produced by suppliers that comply with fair trade requirements. If in the model citizens have preferences exhibiting unconditional altruism, adverse selection prevents separation; i.e., green citizens cannot interact only among themselves, since red citizens will also join the organization to free ride on the green citizens.⁵ If, however, citizens have preferences exhibiting reciprocal altruism, there exists a membership fee such that green citizens join the club and red citizens do not. This increases the scope of self-regulation by eliminating the effect on green citizens of the free riding by red citizens.⁶ Social label organizations could also facilitate self-regulation through other instruments such as enforcement (Prakash and Potoski 2006, 2007).

⁴ Rabin (1998) discusses the economics and psychology of reciprocal altruism and related experimental evidence.

⁵ An example of a failed social label organization is Responsible Care formed by firms in the chemical industry to improve safety and environmental protection in factories in the aftermath of the Bhopal tragedy. The firms joining Responsible Care included those with good and bad safety and environmental records, and the subsequent performance of the firms that joined the organization was no better than those firms that did not join (King and Lenox, 2000, 2002). The performance of Responsible Care participants subsequently improved after enforcement mechanisms were put in place. Whether participating firms had altruistic preferences remains an open question.

⁶ Consequently, if the model is correct, the presence of effective social label organizations suggests that citizens have reciprocal rather than unconditional altruistic preferences.

The demand for public regulation is greater in a heterogeneous society when moral preferences reflect reciprocal altruism than unconditional altruism. This results because regulation eliminates free-riding by citizens with weaker moral preferences. The demand for uniform public regulation is increasing in the quality of the public good and in the strength of moral preferences. The same is true for regulation on demand, unless all citizens either voluntarily provide the public good or demand regulation. Self-regulation, however, can crowd out public regulation.

Both social label organizations and certification organizations expand the scope of regulation with reciprocal altruism, but if citizens have preferences reflecting unconditional altruism, neither organization affects behavior. The maximal scope, however, is bounded above by the self-regulation with unconditional altruism. Enforcement organizations are different, however, since the equilibria are the same for both reciprocal and unconditional altruism. Both forms of enforcement organization expand the scope of self-regulation beyond that with unconditional altruism.

Pledges to contribute to public goods can be enforced by non-profit organizations or by for-profit firms. In the model enforcement takes the form of harm if a citizen fails to contribute to the public good. This enforcement is costly, so the organization must cover its costs. Both non-profit and for-profit enforcement can increase the scope of self-regulation, but the policies chosen are different. A non-profit organization maximizes the expected utility of those who avail themselves of its enforcement services, whereas a for-profit firm chooses its enforcement policy to maximize its profits with revenue from those citizens that use its services. For the case of enforcement on demand; i.e., enforcement voluntarily requested by citizens before they play the public goods game, the for-profit firm provides more aggressive enforcement than does the non-profit organization, but the price charged for enforcement is higher than the fee charged by the non-profit organization.

Another form of organization provides information to a trading partner about how a citizen played in the past. In a two-period model a certification organization expands the scope of self-regulation with reciprocal altruism by inducing citizens with more limited altruism to pool with citizens who are more willing to provide the public good. The citizens with the more limited altruism do so not because they gain from contributing but instead because they find themselves in a dilemma. If they do not contribute to the public good in accord with their single-period preferences, they will be identified as having more limited altruism. Their period-two trading partners then will not provide the public good for some matches because they know the citizen will free-ride. This provides an explanation for self-regulation in which peer conduct matters not because of sociological influence but instead because the citizen will be identified as one who will

free ride.

To explore the scope of self-regulation, an abstract rather than descriptive model is used. The model is based on the random matching model developed by Dixit (2003)(2004) to examine the effect of distance and contract enforcement on trade. Here citizens play a public goods game and have morally-based preferences. The model is related to that of Tabellini (2007), who considers a version of Dixit’s model in which people are in a prisoners’ dilemma and experience guilt if they do not cooperate. The equilibria have properties similar to those in Section II in this paper. He also considers a two-period overlapping generations version in which parents embed their children’s welfare in their own and can transmit values or norms to their children.

Levy and Razin (2007) provide a theory in which a religious organization arises endogenously when people have heterogeneous beliefs about being punished if they defect in a prisoners’ dilemma game. The emergence of the religious organization relies on the prisoners’ dilemma exhibiting strategic complements, and if that is not the case, as in the basic public goods game, the organization does not emerge. Their result is analogous to that for the social label organization considered in Section V, where the organization arises with reciprocal but not with unconditional altruism.

The next section introduces the basic model with unconditional moral preferences, and Section III considers reciprocal altruism and characterizes its impact on the scope of self-regulation. From both normative and positive perspectives Section IV considers public regulation as an alternative to self-regulation. Section V considers private means of increasing the scope of self-regulation. These include social label organizations formed by citizens to facilitate self-selection, a certification organization, and enforcement both by a non-profit organization and a profit-maximizing firm. Section VI considers social pressure on citizens to self-regulate, where the social pressure is applied by an activist funded by voluntary contributions from citizens. Conclusions are offered in the final section.

II. Generalized and Limited Morality

A. The Basic Model

The basic model is constructed with few population characteristics to focus on preference-induced conduct. Consider a symmetric model of a society in which citizens are uniformly distributed on a circle with circumference $2L$. The parameter L may be thought of as how disparate is the society, so the more fractionalized the society the larger is L . Each citizen is randomly matched with another citizen at a socioeconomic distance y with probability $\frac{\alpha e^{-\alpha y}}{2(1-e^{-\alpha L})}$, $\alpha > 0$.⁷ Fig-

⁷ Dixit provides an interpretation of this formulation as resulting from search activities that are

ure 1 illustrates the set-up with citizens A and B matched at a distance y . The higher is α the greater is the probability that a match is local, so a higher α can be interpreted as reflecting how close-knit is a society. The matching may be thought of as representing the everyday activities of citizens, and it is more likely that those activities involve citizens who are close to rather than far from each other. A citizen is thus more likely to interact with a neighbor than a stranger.⁸ If the citizens are firms, distance could be geographic or pertain to a product or technology space. Firms in the chemical industry are likely to be closer to oil companies in technology space than they are to information technology firms. Similarly, the technology used by real estate brokers is closer to that of the information technology industry than to the technology used in the pharmaceutical industry.

The matched citizens trade resulting in a surplus for each. This trade is assumed to be governed by the law and involve an exchange in which there is no chance of cheating. Since both citizens gain, they will trade, so the surplus will be suppressed. Associated with the trade is a harmful externality, and each citizen can mitigate a portion of the externality by providing a local public good. To simplify the model and exposition, the public good is assumed to provide a benefit b both to the citizen providing it and to her matched partner.⁹ The benefit could be from a reduction in pollution emitted by a producer, the purchase of an environmental offset by an airline passenger (Kotchen 2006a), improving working conditions, strengthening public education, improving public safety, etc. A citizen, however, is assumed to have an incentive to free ride, since providing the benefit has a private cost c , where $c > b$. So that the local public good is socially beneficial, assume that $2b - c > 0$. The basic free-rider problem could be resolved through contracts, but the situations considered here are assumed to be either noncontractable or require costly enforcement. Instead, the influence of moral preferences on private provision is considered.

A citizen is assumed to care about her own payoffs, and she also may care about the effect of her actions on the well-being of her trading partner. Alternatively, she may have warm glow

not modeled here.

⁸ Ellison (1993) and Eshel, Samuelson, and Shaked (1998) considered complete information, repeated games with random matching in which players are distributed on a circle and can provide local public goods that benefit only immediate neighbors. Eshel, Samuelson, and Shaked allow players to choose to be an altruist or an egoist based on comparing the average payoffs to each type among neighbors. They characterize the limiting distributions of types and conclude that players are primarily altruists. Ellison showed that although in the limit players played the risk dominant equilibrium the rate of convergence can be sensitive to the matching model. Convergence is rapid when players are matched only with their neighbors.

⁹ The model can be extended in a straightforward manner to public goods that benefit all citizens, as shown in Section II.D.

preferences for the act of providing the local public good. In the former case, preferences are altruistic, and in the second case the citizen cares about the private return from her action. To simplify the exposition the term altruism will be used to encompass both altruistic and warm glow preferences, and both will be represented by the same expression.¹⁰ Andreoni and Miller (2002) conclude from experiments that most subjects exhibited altruism and those who did behaved in accord with revealed preference theory. Hence, their revealed preferences could be represented by a utility function.

These preferences may be stronger the closer the trading partner is to the citizen, since she may care more about neighbors than those who are distant from her (Banfield 1958). In a dictator game experiment Bohert and Frey (1999) found that dictators' offers to other players were decreasing in the social distance between the players, where social distance corresponded to identifiability and familiarity. Similarly, in a voluntary public goods experiment Keser and van Winden (2000) found that contributions were greater and free-riding less among those who interacted repeatedly than among those who were strangers. La Ferrara (2003) provided an overlapping generations model of credit in a "kin group" and found support for the model from data from Ghana. Credit terms were better (e.g., no interest) and default rates were lower for intra-kin loans and for households that contributed funds for lending in the past. These results are consistent with the importance of socioeconomic distance and also with the importance of reciprocity as considered in Section III.

Altruistic preferences are thus represented by a utility $xe^{-\eta y}$, $\eta \geq 0$, where y is the socioeconomic distance to the matched partner and x is a parameter. The parameter η reflects the degree of limited morality with $\eta = 0$ corresponding to pure altruism and $\eta \rightarrow \infty$ corresponding to no altruism nor warm glow preferences. Consequently, lower values of η correspond to stronger moral preferences. The parameter x could equal the benefits b provided to the trading partner in the case of altruism or could be different in the case of warm glow preferences. This utility may be independent of the action of the partner, or it may depend on that action. For example, a citizen may abandon her altruistic or warm glow preferences if her partner is not expected to reciprocate in providing the local public good. The utility from altruism is specified as $\theta xe^{-\eta y}$, $\theta \in [0, 1)$, when the trading partner does not reciprocate, where the parameter θ indexes the extent to which preferences are unconditional with $\theta = 0$ corresponding to pure reciprocal altruism. Initially, preferences are

¹⁰ The distinction between altruistic and warm glow preferences is less a philosophical one and more one of positive implications. Andreoni has shown that if citizens have altruistic preferences government provision of public goods financed by lump-sum taxes crowds out personal giving to finance public goods but does not do so with warm glow preferences. In the present paper there is no government provision, so public crowding out does not arise.

assumed to be independent of whether the partner reciprocates, i.e., unconditional altruism, and then reciprocal altruism is considered in Section III. Moral preferences thus can be generalized or limited and can be unconditional or reciprocal, as illustrated in Figure 2.

The basic game played by the matched citizens is presented in Figure 3. The timing in the game is that nature first draws a match for each citizen, and then the matched pairs choose their actions. Initially, information is assumed to be complete. The equilibrium concept is Nash, and the focus is on symmetric equilibria. A strategy S is a mapping from the match distance to the action set $\{C, N\}$, where C represents contributing or providing the local public good and N represents not contributing. The game is played only once, so citizens have no opportunity to develop a reputation.

If a citizen provides the public good, her utility U_C with unconditional altruism is given by

$$U_C = B - c + xe^{-\eta y}, \quad (1)$$

where $B = 2b$ is the benefits when the other player also provides the public good and $B = b$ otherwise. If a citizen does not provide the public good, the utility is $U_N = B$, where $B = b$ if the other player provides the public good and $B = 0$ otherwise.

In addition to the assumptions above, assume that $b - c + xe^{-\eta L} < 0$, so that a citizen with limited morality does not prefer to contribute in all matches. Similarly, assume that $b - c + x > 0$, so a citizen prefers to contribute when matched with a citizen at her own location.

B. Unconditional Altruism

To characterize the equilibrium with unconditional, limited altruism, note that the public goods game is a dominant strategy game in which the utility difference $\Delta U = U_C - U_N = b - c + xe^{-\eta y}$ is independent of the partner's action. A citizen thus has a dominant strategy S^* of choosing C if $\Delta U \geq 0$ and choosing N otherwise; that is,

$$S^* = \begin{cases} C & \text{if } y \leq y^o \\ N & \text{if } y > y^o, \end{cases}$$

where

$$y^o = \frac{1}{\eta} \ln\left(\frac{x}{c-b}\right) \quad (2)$$

is the boundary of contributions. For a match at the boundary y^o the utility in (1) from contributing is $U_C = b$, which is the benefit received from the contributions of the partner. For a given y the equilibrium is unique.

A citizen thus contributes when matched with a partner no farther away than y^o and otherwise plays N .¹¹ For matches closer than y^o the limited morality is sufficient to overcome the incentive to free ride, and all citizens contribute. For more distant matches the incentive $c - b$ to free ride prevails, and no citizen contributes.

The parameter η can be interpreted as the degree of limited as opposed to generalized morality, so more limited moral preferences (higher η) result in contributions for a smaller set of matches. Also, $\lim_{\eta \rightarrow \infty} y^o = 0$, so in the limit as η increases the scope of self-regulation goes to 0. Altruistic or warm glow preferences are thus necessary for the provision of the local public good, but the impact is limited compared to that with generalized utility, which involves contributions for every match (with $x = b$). The boundary y^o is strictly convex in η , so more limited morality results in relatively smaller decreases in the scope of self-regulation.

Since the public goods game takes place after the match has been drawn, the boundary y^o is independent of the parameter α of the match probability and of the dispersion L in society. Conditional on the socioeconomic distance between them citizens with warm glow or unconditional altruistic preferences behave the same regardless of how dispersed the citizenry is. The scope of self-regulation is given by $\frac{y^o}{L}$, so the greater is the dispersion of the citizenry the lower is the scope of self-regulation. The boundary y^o is strictly increasing in b and x and strictly decreasing in c and η . Consequently, the more beneficial, or higher quality, the local public good relative to its cost the greater is the scope of self-regulation. Also, the stronger are the altruistic or warm glow preferences for providing the public good (higher x , lower η) the greater is the scope of self-regulation.¹²

The ex ante expected utility EU^* of a citizen is

$$\begin{aligned} EU^* &= \int_0^{y^o} (2b - c + xe^{-\eta y}) \left(\frac{\alpha e^{-\alpha y}}{1 - e^{-\alpha L}} \right) dy \\ &= \frac{1 - e^{-\alpha y^o}}{1 - e^{-\alpha L}} (2b - c) + \left(\frac{x\alpha}{\alpha + \eta} \right) \frac{1 - e^{-(\alpha + \eta)y^o}}{1 - e^{-\alpha L}}. \end{aligned} \quad (3)$$

¹¹ Tabellini considers a model in which a citizen experiences guilt if she chooses N but less guilt the more distant is the other citizen in the match. Let the disutility from guilt be additive and represented by $ge^{-\gamma y}$, $0 < \gamma < \alpha$, and assume that g and γ are common knowledge. Then, the citizen plays C for a wider set of matches. That is, the boundary $y^o(g, \gamma)$ of contributions satisfies

$$b - c + xe^{-\eta y^o(g, \gamma)} + ge^{-\gamma y^o(g, \gamma)} \equiv 0$$

and is strictly increasing and strictly concave in g with $y^o(0, \gamma) = y^o$. The boundary is strictly decreasing in γ . Guilt has an effect similar to that of unconditional altruism in the sense that the scope of self-regulation increases with stronger (higher x , lower η) altruistic preferences.

¹² The limit as x decreases is zero contributions; i.e., $\lim_{x \rightarrow c-b} y^o = 0$.

The first term is the expected utility from the benefits and costs of the public good, and the second term is the expected utility from altruism. The expected utility is increasing in y^o , since contributions occur for a broader set of matches. The greater is the dispersion L of society the lower is the expected utility, since other citizens are farther away and hence the public good is provided for fewer matches. The expected utility is increasing in b and x and decreasing in c and η , as expected.

C. Heterogeneous Preferences

Citizens can differ in the extent to which they care about the well-being of other citizens or receive a warm glow from the act of providing the public good. The heterogeneity introduced in this section is in the rates at which their unconditional altruism or warm glow preferences decline with socioeconomic distance. This heterogeneity has no effect on the nature of equilibrium strategies when altruistic or warm glow preferences are unconditional.

Suppose that citizens are one of two types with parameters $\eta_i, i = 1, 2, \eta_1 < \eta_2$, which are private, soft information and cannot be revealed to others. Since the preferences of a citizen for C versus N are independent of the actions of her matched partner, the citizens of each type have dominant strategies. The dominant (Bayesian Nash equilibrium) strategy S_i^* of a citizen of type i is given by¹³

$$S_i^* = \begin{cases} C & \text{if } y \leq y_i^o \\ N & \text{if } y > y_i^o, \end{cases} \quad (4)$$

where

$$y_i^o = \frac{1}{\eta_i} \ln\left(\frac{x}{c-b}\right), \quad i = 1, 2. \quad (5)$$

The equilibrium with private information is thus qualitatively the same as the equilibrium in the case in which citizens are of one type. The same is true for any number of types.

In the equilibrium the type 2s free ride on the contributions of the type 1s. Because of unconditional altruism, however, this free riding does not affect the behavior of the 1s, since the utility $xe^{-\eta_1 y}$ results only from their own contributions.

This, however, has distributive effects. If the proportion of type 1 citizens is β , the expected utilities $EU_i^\beta, i = 1, 2$, are

$$EU_1^\beta = \frac{1}{1 - e^{-\alpha L}} \left[(2b - c) \left(1 - e^{-\alpha y_1^o} \right) - (1 - \beta)b \left(e^{-\alpha y_2^o} - e^{-\alpha y_1^o} \right) + \frac{\alpha x}{\alpha + \eta_1} \left(1 - e^{-(\alpha + \eta_1)y_1^o} \right) \right] \quad (6)$$

¹³ If citizens experience guilt from not providing the public good, the equilibrium is analogous to that in (4) and (5) with $y_i^o(g, \gamma)$ replacing y_i^o .

and

$$EU_2^\beta = \frac{1}{1 - e^{-\alpha L}} \left[(2b - c) \left(1 - e^{-\alpha y_2^o} \right) + \beta b \left(e^{-\alpha y_2^o} - e^{-\alpha y_1^o} \right) + \frac{\alpha x}{\alpha + \eta_2} \left(1 - e^{-(\alpha + \eta_2) y_2^o} \right) \right]. \quad (7)$$

The term $-(1 - \beta)b(e^{-\alpha y_2^o} - e^{-\alpha y_1^o})$ in (6) represents the loss to a type 1 from the possibility of being matched with a type 2 on the set $y \in (y_2^o, y_1^o]$, since for these matches the type 2s do not provide the public good. The term $\beta b(e^{-\alpha y_2^o} - e^{-\alpha y_1^o})$ in (7) represents the gain to a type 2 from the possibility of being matched with a type 1 on that interval and free riding on her contribution.

Although heterogeneity and incomplete information about the preferences of other citizens have no effect on strategies, the expected scope $\frac{1}{L}(\beta y_1^o + (1 - \beta)y_2^o)$ of self-regulation depends on the distribution of types. In the case considered here, the expected scope of self-regulation is increasing in β , since a greater proportion of citizens contribute for matches $y \in (y_2^o, y_1^o]$.¹⁴ Also, aggregate utility $EU_1^\beta + EU_2^\beta$ is increasing in β , since there are more contributions the more citizens there are with stronger moral preferences.

The results of this section are summarized in the following proposition.

Proposition 1: With unconditional moral preferences citizens provide the local public good only in matches with $y \in [0, y^o]$. The scope of self-regulation is increasing in the quality of the public good and the strength of moral (or warm glow) preferences. Heterogeneity of moral preferences results in equilibria with the same qualitative properties, and the scope of self-regulation for each type is unaffected by the proportions of types. The expected scope of self-regulation is increasing in the proportion of citizens with stronger moral preferences.

D. A Pure Public Good

The model can be extended to pure public goods for which a contribution provides benefits to all citizens, as in the case of mitigating global warming. This is equivalent to the local public goods model. To simplify the notation, suppose there is a finite number M of citizens. The expected utility $EU_{C_j}^M$ of a citizen j if she contributes and all other citizens contribute is

$$EU_{C_j}^M = 2b - c + x e^{-\eta y} + \sum_{i \neq j} \left(b + x e^{-\eta y} \right).$$

The expected utility $EU_{N_j}^M$ if she does not contribute is $EU_{N_i}^M = \sum_{j \neq i} b$, and the difference is

$$EU_{C_i}^M - EU_{N_i}^M = b - c + x^M e^{-\eta y}, \quad (8)$$

¹⁴ Since y_i^o is strictly convex in η_i , the expected scope of self-regulation is greater than the scope of self-regulation at the mean $\beta \eta_1 + (1 - \beta) \eta_2$.

where $x^M = (M - 1)x$. Then, (8) is the same as (1) with x^M replacing x . The model thus includes pure as well as local public goods.

III. Reciprocal Altruism

A citizen may have altruistic but conditional preferences. Altruism may extend only to other citizens who contribute to the public good or warm glow preferences may extend only to the act of providing benefits to someone who deserves them; i.e., who earned them by also providing a public good. For example, in the context of corporate social responsibility a concern of firms is that by undertaking costly social actions, such as providing public goods, they will have a cost disadvantage relative to competitors that do not take such actions. A firm then could be willing to provide the public good if it were confident that other firms would do the same, in which case the playing field would be level. Such preferences can be represented as reciprocal altruism in the sense that firms have altruistic preferences but only when others also provide the public good.

Reciprocal altruism could be represented in a number of ways. Levine (1998) represented it through preferences in which a citizen is “more altruistic to an opponent who is more altruistic toward them.” Rabin (1993, p. 1282) considered a concept of fairness in which “people are willing to sacrifice their own material well-being to help those who are being kind.” He represented this by a “kindness function” that depends on strategies and beliefs. Here, reciprocal altruism is conditional only on (anticipated) actions.¹⁵ Reciprocity pertains to actions, so a citizen must have beliefs about whether her trading partner will contribute to the public good. Since information is complete in the basic model, a citizen understands which action her partner will take. A citizen’s altruistic preferences thus are conditional and weaker by a factor θ when (C, N) is played.

Reciprocal preferences introduce both complexity and opportunities. The complexity arises because of multiple equilibria, and the opportunity is for organizations to expand the scope of self-regulation. Multiple equilibria result because the utility difference between playing C and N depends on the action of the partner. Reciprocal (or conditional) altruism also provides an opportunity for an organization to affect the scope of self-regulation, as considered in Section V. Both the complexity and the opportunity arise because with reciprocal altruism the public goods game has strategic complements and hence is a coordination game.

¹⁵ Tabellini considers reciprocity similar to that considered here, but his basic model assumes strategic complements, so the qualitative properties of his equilibria are unchanged. His formulation of reciprocity corresponds to shame as considered in Section VI and results in a larger maximal scope of self-regulation. Reciprocity as considered here results in a smaller scope of self-regulation compared to unconditional altruism when citizens’ preferences are heterogeneous.

To characterize the equilibria with reciprocal altruism, let δ denote the probability that the partner plays C . The difference in the expected utilities from playing C rather than N then is

$$EU_C - EU_N = b - c + (\delta + \theta(1 - \delta))xe^{-\eta y},$$

and define

$$y^r(\delta; \theta) = \begin{cases} 0 & \text{if } (\delta + \theta(1 - \delta))x \leq c - b \\ \frac{1}{\eta} \ln\left(\frac{(\delta + \theta(1 - \delta))x}{c - b}\right) & \text{if } (\delta + \theta(1 - \delta))x > c - b. \end{cases} \quad (9)$$

If $y^r(0; \theta) > 0$, the unique equilibrium for matches $y \in [0, y^r(0; \theta)]$ is for a citizen to play C , since she has a dominant strategy as in Section II. Similarly, for $y > y^r(1; \theta) = y^o$ the dominant strategy equilibrium is (N, N) , since even if the partner plays C , a citizen cannot gain from playing C . The maximal scope of self-regulation is thus the same as with unconditional altruism. When $y^r(0; \theta) > 0$, it is strictly increasing in θ , so as altruism becomes less conditional, the minimum scope of self-regulation increases.

For matches with $y \in (y^r(0; \theta), y^r(1; \theta)]$, the game is a coordination game with three best-response equilibria. In the Pareto dominant equilibrium both citizens play C , and neither has an incentive to deviate. In this equilibrium the scope of self-regulation is the same as with unconditional altruism. In another, both citizens play N , since if the partner will play N , the citizen by playing C can gain only $b - c + \theta xe^{-\eta y}$, which is negative for $y \in (y^r(0; \theta), y^r(1; \theta)]$. The third equilibrium is in mixed strategies, with both citizens playing C with probability $\delta(y; \theta)$ given by¹⁶

$$\delta(y; \theta) \equiv \frac{c - b - \theta xe^{-\eta y}}{(1 - \theta)xe^{-\eta y}}. \quad (10)$$

That is, given that the partner plays $\delta(y; \theta)$, a citizen is indifferent between playing C or N and hence is willing to play $\delta(y)$.

If $y^r(0; \theta) = 0$, the minimal probability δ^o of contributing even for a match $y = 0$ is

$$\delta^o = \frac{c - b - \theta x}{(1 - \theta)x}.$$

Then, the equilibrium mixed strategy $\delta^*(y; \theta)$ is

$$\delta^*(y; \theta) = \begin{cases} \delta^o & \text{if } y = 0 \\ \delta(y; \theta) & \text{if } y \in (0, y^r(\theta; 1)]. \end{cases}$$

The probability $\delta(y; \theta)$ in (10) of playing C is strictly increasing in θ , since

$$\frac{d\delta(y; \theta)}{d\theta} = \frac{c - b - xe^{-\eta y}}{(1 - \theta)^2 xe^{-\eta y}} > 0, \quad y \in (y^r(0; \theta), y^r(1; \theta)].$$

¹⁶ Note that $\delta(y^r(0; \theta)) = 0$ and $\delta(y^r(1; \theta)) = 1$.

Consequently, the scope of self-regulation is increasing in the extent to which morality is unconditional rather than reciprocal. In the limit as $\theta \rightarrow 1$ contributions from all citizens are induced for matches $y \in [0, y^r(1; \theta)]$, as characterized for the case of unconditional altruism in Section II.

The probability $\delta(y; \theta)$ is strictly increasing and strictly convex in y , i.e., because of limited morality a higher probability of a partner contributing is required to induce reciprocal contributions for more distant matches. For matches $y > y^r(1; \theta)$ even a partner contributing with probability 1 is insufficient to induce the citizen to provide the local public good.

The mixed strategy equilibrium and the (N, N) equilibrium identify a role for culture to the extent that it fosters unconditional altruism (as well as generalized morality) or it selects an equilibrium with a greater scope of self-regulation. Culture changes slowly, however, and citizens have the alternative of forming an organization to increase the scope of self-regulation, as considered in Section V. As shown in that section an organization can form with heterogeneous types even when in the absence of the organization all citizens would play the Pareto dominant equilibrium.

With heterogeneous preferences the equilibria are analogous to those with one type. Let the type 1s play C with probability γ and the type 2s play C with probability ρ . Then, $\delta = \beta\gamma + (1-\beta)\rho$ is the probability that a partner in a match at y plays C . Define the boundaries $y_i^r(\delta; \theta)$, $i = 1, 2$, by

$$y_i^r(\delta; \theta) = \begin{cases} 0 & \text{if } (\delta + \theta(1 - \delta))x \leq c - b \\ \frac{1}{\eta_i} \ln\left(\frac{(\delta + \theta(1 - \delta))x}{c - b}\right) & \text{if } (\delta + \theta(1 - \delta))x > c - b. \end{cases}$$

As above if $y_2^r(0; \theta) > 0$, the dominant strategy equilibrium for $y \leq y_2^r(0; \theta)$ is for all citizens to play C . For $y > y_2^r(1; \theta) = y_2^o$, the type 2's have a best response of playing N . If $y_2^r(0; \theta) = 0$, contributing is a best-response for the type 2s and the type 1s for matches $y \in [0, y_2^o]$.

For the type 1s playing C is a best response if $y \in (y_2^o; y_1^r(0; \theta)]$, which is nonempty if $\theta x > c - b$ and

$$\left(\frac{c - b}{x}\right)^{1 - \frac{\eta_1}{\eta_2}} \leq \theta. \quad (11)$$

For matches $y \in (\max\{y_2^o, y_1^r(0; \theta)\}, y_1^r(\beta; \theta)]$, which is nonempty for

$$\left(\frac{c - b}{x}\right)^{1 - \frac{\eta_1}{\eta_2}} \leq \beta + \theta(1 - \beta), \quad (12)$$

there are three best-response equilibria. In one, all type 1s play C , and in another all type 1s play N . The third is a mixed strategy equilibrium analogous to that characterized above for one

type.¹⁷ When (12) is satisfied there are sufficient type 1s that their reciprocal altruism induces them to contribute. The type 2s then free ride on the contributions of the type 1s for matches $y \in (y_2^o, y_1^r(\beta; \theta)]$.

The inequality in (12) is satisfied, for example, if δ or θ is large and η_2 is large, in which case the left side is approximately $\frac{c-b}{x}$. It is also satisfied if $c - b$ is small relative to x . If (12) is not satisfied, there are too few type 1s to induce contributions by the 1s for $y > y_2^r(1; \theta)$. There is then a single equilibrium for each match distance y . For $y \leq y_2^r(1; \theta)$ all citizens play C , and for $y > y_2^r(1; \theta)$, all citizens play N . In this case the type 2s cannot free ride on the type 1s because the free riding that the type 2s would do if a type 1 were to contribute for a match $y > y_2^r(1; \theta)$ is sufficient to cause the type 1s not to contribute. The scope of self-regulation is then limited by the (potential) free-riding by the 2s. This is due to reciprocal altruism.

In all three equilibria no type 1 contributes for $y \in (y_1^r(\beta; \theta), L]$, since a type 1 citizen can only count on a contribution from the other type 1s with whom she might be matched. That is, the type 2s free ride, which limits the scope of self-regulation by the type 1s when altruism is reciprocal. In contrast, with unconditional altruism the free riding by the 2s has no effect on the strategy of the 1s, but with reciprocal altruism a type 1 receives utility $xe^{-\alpha y}$ only when matched with another type 1. This occurs only with probability β , so for matches $y \in (y_1^r(\beta; \theta), y_1^o]$ the best response is not to contribute. Consequently, when the citizenry is heterogeneous, the scope of self-regulation is smaller with reciprocal than with unilateral altruism. The scope of self-regulation is strictly increasing in β , since then there are fewer type 2s to free ride and more type 1s to reciprocate.

¹⁷ In one mixed strategy equilibrium, if the type 1s play C , the type 2s have a mixed strategy

$$\rho(y; \theta) = \frac{c - b - (\beta + \theta(1 - \beta))xe^{-\eta_2 y}}{(1 - \beta)(1 - \theta)xe^{-\eta_2 y}}, \quad y \in (y_2(\beta; \theta), y_2^o],$$

which satisfies $\rho(y_2(\beta; \theta)) = 0$ and $\rho(y_2^r(1; \theta)) = 1$. The type 1s play C when the type 2s play $\rho(y)$ for $y \leq y_2^o$ provided that $y_1^r(\beta; \theta) \geq y_2^o$. In this equilibrium, the type 1s play C on $[0, y_2^o]$. The probability $\rho(y; \theta)$ is strictly increasing in θ , so the probability of a contribution by a type 2 is greater the more unconditional (higher η_2) is the altruism of citizens. There is also an equilibrium in which the type 1s play a mixed strategy $\gamma(y; \theta)$ given by, for the case in which $y_2^r(1; \theta) < y_1^r(\beta; \theta)$,

$$\gamma(y; \theta) = \frac{c - b - \theta xe^{-\eta_1 y}}{\beta(1 - \theta)xe^{-\eta_1 y}}, \quad y \in [y_2^r(1; \theta); y_1^r(\beta; \theta)].$$

The probability $\gamma(y; \theta)$ is increasing in y , so a greater likelihood of contributions by the partner is needed to induce contributions for more distant matches. The probability is increasing in θ , so the more unconditional is the altruism of citizens the higher is the probability of contributing for a given y .

The expected utility $EU_1^r(\beta)$ for a type 1 in the Pareto dominant equilibrium with the greatest scope of self-regulation for the case in which (12) is satisfied is

$$\begin{aligned} EU_1^r(\beta) &= \left[\int_0^{y_2^o} \left(\frac{2b - c + xe^{-\eta_1 y}}{1 - e^{-\alpha L}} \right) + \int_{y_2^o}^{y_1^r(\beta; \theta)} \left((1 + \beta)b - c + (\theta + \beta(1 - \theta))xe^{-\eta_1 y} \right) \right] \frac{\alpha e^{-\alpha y}}{1 - e^{-\alpha L}} dy \\ &= \frac{1}{1 - e^{-\alpha L}} \left[(2b - c) \left(1 - e^{-\alpha y_1^r(\beta; \theta)} \right) - (1 - \beta)b \left(e^{-\alpha y_2^o} - e^{-\alpha y_1^r(\beta; \theta)} \right) \right. \\ &\quad \left. + \frac{\alpha x}{\alpha + \eta_1} \left((1 - \theta)(1 - \beta)e^{-(\alpha + \eta_1)y_1^r(\beta; \theta)} - (1 - \beta)(1 - \theta)e^{-(\alpha + \eta_1)y_2^o} \right) \right], \end{aligned} \tag{13}$$

where $-(1 - \beta)b \left(e^{-\alpha y_2^o} - e^{-\alpha y_1^r(\beta; \theta)} \right)$ is the effect of the free-riding by the type 2s on a type 1. The expected utility $EU_2^r(\beta)$ for a type 2 for the case in which (12) is satisfied is

$$\begin{aligned} EU_2^r(\beta) &= \left[\int_0^{y_2^o} \left(2b - c + xe^{-\eta_2 y} \right) + \int_{y_2^o}^{y_1^r(\beta; \theta)} \beta b \left(\frac{\alpha e^{-\alpha y}}{1 - e^{-\alpha L}} \right) \right] dy \\ &= \frac{1}{1 - e^{-\alpha L}} \left[(2b - c) \left(1 - e^{-\alpha y_2^o} \right) + \beta b \left(e^{-\alpha y_2^o} - e^{-\alpha y_1^r(\beta; \theta)} \right) + \frac{\alpha x}{\alpha + \eta_2} \left(1 - e^{-(\alpha + \eta_2)y_2^o} \right) \right], \end{aligned} \tag{14}$$

where the term $\beta b \left(e^{-\alpha y_2^o} - e^{-\alpha y_1^r(\beta; \theta)} \right)$ is the gain to a type 2 from free-riding on the type 1s.

The results of this section are characterized in the following proposition.

Proposition 2: Reciprocal altruism transforms the dominant strategy game into a coordination game for some matches. With a homogeneous citizenry the scope of self-regulation in the Pareto dominant equilibrium is the same for all $\theta \in [0, 1)$ as with unconditional altruism, but the scope is smaller in the other equilibria. In a heterogeneous citizenry the scope of self-regulation in the Pareto dominant, best-response equilibrium is smaller with reciprocal than unconditional altruism because of free riding. The expected scope of self-regulation is increasing in θ and β .

IV. Public Regulation

To increase the private provision of the local public goods, citizens could demand public regulation to induce or compel themselves to self-regulate. Public regulation, however, can be crowded out by voluntary self-regulation. The first-best requires contributions by all matched citizens, but individual self-regulation occurs only for matches such that $y \in [0, y^o]$. Self-regulation thus is second-best when citizens have limited morality. Public regulation can mitigate the free-rider problem by compelling contributions. For example, the government could require citizens to purchase a share of wind-generated electricity, ride a bicycle rather than drive a car, purchase a carbon offset when flying, or for firms improve the working condition in overseas factories. Regulation and its enforcement, however, is costly, and the cost could exceed the benefits. The

regulation considered does not involve the public provision of the public good but instead involves requiring private provision. The regulation thus does not affect the private cost c of providing the public good.¹⁸ In addition, the availability of regulation is not assumed to affect the altruism of citizens.

Consider first the case of uniform regulation that compels contributions in all matches at a cost t per citizen, which could result from operating the regulatory agency. To give public regulation its best chance, assume that all citizens comply with the regulation.¹⁹ The ex ante expected utility EU^R of a citizen subject to uniform regulation is

$$EU^R = \int_0^L (2b - c + xe^{-\eta y}) \frac{\alpha e^{-\alpha y}}{1 - e^{-\alpha L}} dy - t,$$

and the expected gain from regulation for both unconditional and reciprocal altruism is, using (3),

$$\begin{aligned} EU^R - EU^* &= \int_{y^o}^L (2b - c + xe^{-\eta y}) \frac{\alpha e^{-\alpha y}}{1 - e^{-\alpha L}} dy - t \\ &= (2b - c) \left(\frac{e^{-\alpha y^o} - e^{-\alpha L}}{1 - e^{-\alpha L}} \right) + \left(\frac{\alpha x}{\alpha + \eta} \right) \left(\frac{e^{-(\alpha+\eta)y^o} - e^{-(\alpha+\eta)L}}{1 - e^{-\alpha L}} \right) - t. \end{aligned} \quad (15)$$

Letting t^r equate to 0 the expression in (15), regulation is unanimously preferred if the cost is less than t^r , and otherwise self-regulation is preferred. The boundary t^r is increasing in b and x and decreasing in c , η , and L . Higher benefits b from the public good has two effects. First, it increases the gain $2b - c + xe^{-\eta y}$ from regulation for matches $y \in (y^o, L]$. Second, it increases the scope of self-regulation which reduces the set on which regulation is beneficial. The net effect, however, is to increase the demand for regulation, so regulation is a normal good. The gain in (15) is also increasing in the strength of moral preferences (higher x , lower η), so regulation is a moral normal good.

Dixit (2003) considers enforcement of this form, and focuses on whether self-enforcement or external enforcement analogous to public regulation is preferred in a society as a function of L . He finds that for small L self-enforcement is better from a welfare perspective, whereas for large L external enforcement is better. This is equivalent to t^r in (15) being decreasing in L . That is, for a given $t > 0$ the difference in (15) is negative if self-regulation would be extensive (y^o sufficiently close to L) in the absence of regulation. In this case voluntary self-regulation crowds out public regulation. This provides an explanation for why a society would not support public regulation

¹⁸ Tabellini considers the effect of government enforcement on the educational choices of parents for their children in a model in which players experience guilt if they do not cooperate.

¹⁹ Alternatively, regulation could involve enforcement, as considered in Section V.C.

yet deal effectively with the externality. Conversely, when self-regulation is less extensive, as in a disparate society (high L) with little self-regulation, the demand for uniform regulation could be high. Even if all citizens ex ante prefer uniform regulation, however, some citizens are worse off ex post than in the absence of regulation.

The gain from regulation in (15) is the same for both unconditional and reciprocal altruism when citizens are homogeneous and coordinate on the Pareto dominant equilibrium. The gain for public regulation when the citizenry is heterogeneous, however, is greater with reciprocal altruism than with unconditional altruism because with the former the free riding by the type 2 citizens causes the type 1 citizens to restrict their self-regulation, as shown in Section III. Consequently, in a heterogeneous society the demand for uniform public regulation is greater when altruistic preferences are reciprocal than unconditional.

From a positive perspective regulation must be adopted by the citizenry. If a vote were taken ex ante, all citizens would vote for regulation if the cost were less than t^r , and otherwise would vote against it. Citizens, however, can control when they vote, so suppose that the vote takes place after the matching but before citizens execute their trades. To give regulation its best chance, suppose it is selective rather than uniform; i.e., regulation is supplied only when demanded by one of the citizens in a match. Regulation, however, must be approved by a majority of citizens before it can be made available.

Citizens with matches $y \in [0, y^o]$ would not support regulation, whereas some citizens with matches $y \in (y^o, L]$ would vote in favor of it depending on the fee, which is assumed to equal the cost t .²⁰ The gain to a citizen in such a match is

$$2b - c + xe^{-\eta y} - t,$$

where t is incurred only when the matched citizens demand regulation. The cost may be viewed as a user fee, as in the case of a carbon surcharge on air travel. Those citizens with matches $y \in [0, y^o]$ are unaffected, whereas citizens with matches $y \in (y^o, y^t]$ benefit from regulation, where $y^t(\delta; \theta) = L$ if $t \leq 2b - c$, and otherwise

$$y^t(\delta; \theta) = \begin{cases} 0 & \text{if } (\delta + \theta(1 - \delta))x \leq c - 2b + t \\ \frac{1}{\eta} \ln\left(\frac{(\delta + \theta(1 - \delta))x}{c + t - 2b}\right) & \text{if } (\delta + \theta(1 - \delta))x > c - 2b + t. \end{cases} \quad (16)$$

²⁰ Only those citizens who strictly benefit from regulation are assumed to vote for it, since regulation could also involve fixed costs covered by taxes rather than fees on those who use the regulation.

To be effective in inducing contributions ($\delta = 1$) to the public good, the cost t of regulation must satisfy $b > t$; i.e., $y^t(1; \theta) > y^o$ if and only if $b > t$.

The citizens adopt regulation under majority rule only if

$$\frac{y^t(1; \theta) - y^o}{L} > \frac{1}{2},$$

so regulation is never adopted if a majority of citizens would self-regulate; i.e., when the scope of self-regulation is greater than $\frac{1}{2}$. Self-regulation then crowds out public regulation.²¹

The political support $y^t(1; \theta) - y^o$ for regulation is independent of θ and hence is the same for unconditional and reciprocal altruism in a homogeneous society. The support for regulation on demand is decreasing in the net benefits and in the strength of moral preferences, so self-regulation crowds out public regulation. If $y^t(1; \theta) \in (y^o, L)$ and $t < b$, the support for regulation is increasing in the strength (higher x , lower η) of moral preferences and in the quality ($2b - c$) of the public good. Regulation is then a normal good unless every citizen with a match $y > y^o$ benefits from regulation. As with uniform regulation the demand for regulation is greater in a heterogeneous society with reciprocal than with unconditional altruism. If $y^t(1; \theta) = L$, the support for regulation is decreasing in x and increasing in η since the scope of self-regulation is greater, so the stronger are moral preferences, the lower is the political support for regulation. Similarly, the support for regulation is decreasing in b and increasing in c when $y^t(1; \theta) = L$.

The results of this section are summarized in the following proposition.

Proposition 3: Ex ante uniform public regulation is a normal good in that the demand for regulation is increasing in the net benefits from the public good and in the strength of moral preferences, but the number of matches for which citizens gain from regulation is decreasing in both. Regulation on demand is ex post individually rational, and when regulation benefits every match $y^t(1; \theta) = L$, its political support is decreasing in both the quality of the public good and the strength of moral preferences. Regulation is then crowded out by self-regulation. If $y^t(1; \theta) < L$ and regulation is not costly ($b > t$), the support for regulation is increasing in the quality of the public good and the strength of moral preferences. The support for regulation is the same with unconditional and reciprocal altruism in a homogeneous society, but in a heterogeneous society the demand for regulation is greater with reciprocal than with unconditional altruism.

V. Privately-Organized Self-Regulation

²¹ Public regulation could also be supported by the citizenry if it lowered the cost c of providing the public good.

A. A Social Label Organization

This section considers private alternatives to public regulation where citizens utilize a voluntary, self-regulation organization to affect their behavior. In the context of the model an organization can affect the scope of self-regulation in three ways. First, it could allow citizens to reveal their type, allowing them to coordinate their behavior. Second, the organization could certify that a citizen contributed to the public good in a prior period and make that information available to other citizens in the current period. Third, the organization could provide enforcement that induces contributions by raising the cost of not contributing. Sections V.C and V.D consider enforcement provided by non-profit organizations and by for-profit firms, respectively, and compares the strength of their enforcement. In Section VI enforcement is provided by NGOs funded by voluntary contributions from citizens. The analysis in this section does not explain the formation of an organization but instead explains whether an organization can exist in the sense that citizens avail themselves of its services.

A social label organization allows its members to trade only with other members. For example, fast food chains can purchase only from suppliers that practice humane treatment of food animals, and farmers employing those practices can supply only those chains. Similarly, retailers can buy only from overseas suppliers that meet certain standards for working conditions in their factories, and suppliers meeting those standards can concentrate their sales on retailers that only sell products produced under those standards. As considered in Section IV with public regulation the equilibria are the same with both unconditional and reciprocal altruism, but the equilibria with a social label organization depend importantly on the nature of preferences. A social label organization cannot exist with unconditional altruism, whereas it can exist with reciprocal altruism.

Consider two types of citizens with $\eta_1 < \eta_2$. Citizens cannot credibly reveal their types to others, but they can join a social label organization that attracts particular types of members. The organization is open to all citizens, and the social label received by members is publicly observable. Both types of citizen are assumed to be distributed uniformly on the circle, and assume that a match selects a distance y and places the citizen before citizens of both types. Citizens who join the organization can trade among themselves, and those who do not join trade with other citizens who also did not join the club. The screening instrument is the membership fee f for the organization.

To determine if a social label organization can attract type 1s but not type 2s, consider the case of unconditional altruism. In the absence of an organization, contributions are maximal given the limited moral preferences of citizens. Nevertheless, citizens with stronger moral preferences

can gain because they could be assured that their trading partner would provide the public good for matches with distance up to y_1^o rather than y_2^o with probability $1 - \beta$. Their expected gain is $\Delta EU_1^u = EU_1^* - EU_1^\beta$, where the superscript u denotes unconditional altruism, EU_1^* is given in (3) with η_1 replacing η and y_1^o replacing y^o , and EU_1^β is given in (6). This can be evaluated as

$$\Delta EU_1^u = \frac{(1 - \beta)b(e^{-\alpha y_2^o} - e^{-\alpha y_1^o})}{1 - e^{-\alpha L}},$$

which results from avoiding the loss due to free riding by the type 2s.

The social label organization will exist if there is a fee $f \leq \Delta EU_1^u$ that no type 2 would be willing to pay. If a type 2 does not join the organization, he trades only with type 2s and his utility is given in (3) with η_2 and y_2^o replacing η and y^o , respectively. If he joins the organization, he trades only with type 1s and hence gains by free riding on the local public good provided for matches $y \in (y_2^o, y_1^o]$. The gain ΔEU_2^u for a type 2 is then

$$\Delta EU_2^u = b \left(\frac{e^{-\alpha y_2^o} - e^{-\alpha y_1^o}}{1 - e^{-\alpha L}} \right). \quad (17)$$

The gain to a type 2 citizen is greater than the gain to a type 1 citizen, so there is no fee that can separate the types. When altruism is unconditional, adverse selection thus prevents a social label organization that includes type 1s but not type 2s.

If citizens have reciprocal altruism, a social label organization can provide separation. With reciprocal altruism the type 1s gain not only from the public good provided by the other type 1s but also from the reciprocation of their altruism. The type 2s have no such gain, since they do not contribute for $y > y_2^o$. To provide the toughest test for an organization, assume that in the absence of an organization citizens play the best of the self-regulation equilibria characterized in Section III. That is, in the absence of an organization and when (12) is satisfied the type 1s play C for $y \in [0, y_1^r(\beta; \theta)]$, whereas the type 2s play C for $y \in [0, y_2^o]$. If only type 1s join the organization, they trade with each other for matches $y \leq y_1^o$. Their expected gain ΔEU_1^r , where r denotes reciprocal, then is

$$\begin{aligned} \Delta EU_1^r = & \frac{1}{1 - e^{-\alpha L}} \left[(2b - c) \left(e^{-\alpha y_1^r(\beta; \theta)} - e^{-\alpha y_1^o} \right) + (1 - \beta)b \left(e^{-\alpha y_2^o} - e^{-\alpha y_1^r(\beta; \theta)} \right) \right. \\ & + \frac{\alpha x}{\alpha + \eta_1} \left((1 - \theta)(1 - \beta) \left(e^{-(\alpha + \eta_1)y_2^o} - e^{-(\alpha + \eta_1)y_1^o} \right) \right. \\ & \left. \left. + (\theta + \beta(1 - \theta)) \left(e^{-(\alpha + \eta_1)y_1^r(\beta; \theta)} - e^{-(\alpha + \eta_1)y_1^o} \right) \right) \right]. \end{aligned}$$

The expected gain ΔEU_2^r to a type 2 if he joined the organization rather than trade only with type 2s is given by (17).

If there exists a membership fee f satisfying

$$\begin{aligned} \Delta EU_2^u < f \leq \Delta EU_1^r = \Delta EU_2^u + \frac{1}{1 - e^{-\alpha L}} \left[(b - c) \left(e^{-\alpha y_1^r(\beta; \theta)} - e^{-\alpha y_1^o} \right) \right. \\ \left. + \frac{\alpha x}{\alpha + \eta_1} \left((1 - \theta)(1 - \beta) \left(e^{-(\alpha + \eta_1) y_2^o} - e^{-(\alpha + \eta_1) y_1^o} \right) + (\theta + \beta(1 - \theta)) \left(e^{-(\alpha + \eta_1) y_1^r(\beta; \theta)} - e^{-(\alpha + \eta_1) y_1^o} \right) \right) \right], \end{aligned} \quad (18)$$

self-selection results in an organization with only type 1s as members. The right side of (18) can be rewritten as

$$\begin{aligned} \Delta EU_2^u + \frac{1}{1 - e^{-\alpha L}} \left[\int_{y_1^r(\beta; \theta)}^{y_1^o} (b - c + x e^{-\eta_1 y}) \left(\frac{\alpha e^{-\alpha y}}{1 - e^{-\alpha L}} \right) dy \right. \\ \left. + \frac{\alpha x}{\alpha + \eta_1} (1 - \theta)(1 - \beta) \left(e^{-(\alpha + \eta_1) y_2^o} - e^{-(\alpha + \eta_1) y_1^r(\beta; \theta)} \right) \right]. \end{aligned}$$

Since $b - c + x e^{-\eta_1 y} > 0$ for $y \in (y_1^r(\beta; \theta), y_1^o)$, there exist a membership fee that results in separation. Separation results because the type 1s receive the benefits from the public good by their type 1 partner and utility from their own provision being reciprocated, whereas type 2s do not gain from reciprocal altruism because they do not contribute for matches $y \in (y_2^o, y_1^o]$.

Reciprocal rather than unconditional altruism thus can give rise to organized self-regulation. Moreover, with reciprocal altruism the social label organization expands the scope of self-regulation. That is, the organization allows type 1s to provide the public good for matches up to y_1^o , whereas in the absence of the organization they provide the public good only up to $y_1^r(\beta; \theta)$. A social label organization expands the scope of self-regulation only by eliminating the effect of free-riding on the willingness of those with stronger moral preferences to provide the public good. A social label organization thus allows citizens to achieve the same equilibrium as with unconditional altruism but not to expand the scope of self-regulation beyond the scope with unconditional altruism. These results are summarized in the following proposition.

Proposition 4: A social label organization can exist in a heterogeneous society when preferences reflect reciprocal but not unconditional altruism. The organization expands the scope of self-regulation and allows citizens to achieve the same equilibrium as with unconditional altruism.

B. A Certification Organization

A social label organization expands the scope of self-regulation when altruism is reciprocal by separating the types of citizens, which eliminates free riding by those with weaker moral preferences. In contrast, a certification organization expands the scope of self-regulation when altruism is reciprocal by inducing some citizens with weaker moral preferences to contribute in one period

so that in the next period they can free ride on citizens with stronger moral preferences. Free riding is then reduced in the first-period which induces citizens with stronger moral preferences to expand the set of matches for which they provide the public good. A certification organization thus can expand the scope of self-regulation, but it cannot achieve the same equilibrium as when altruism is unconditional. With unconditional altruism citizens have a dominant strategy in both periods, so their behavior is unaffected by free riding.

This section considers a two-period extension of the model with two types of citizens and reciprocal altruism and shows that pooling results in the first period and expands the scope of self-regulation. For an extended horizon to affect the scope of self-regulation, information must be provided to future match partners about a citizen's play in the first period. The information system that accomplishes this is not modeled here. One simple type of information system is for a citizen in the first period match to give a certificate to her partner if and only if he played C in a match of distance y . For example, when a citizen buys a carbon offset, she receives a receipt that can be shown to her matched partner next period or posted on a secure Internet site that can be checked by future trading partners. This system, however, has opportunities for fraud, counterfeiting the certificate, or corruption, paying the first-period partner to give a certificate when N is played. A more elaborate information system, as in the model of the law merchant by Milgrom, North, and Weingast (1990), however, could resolve the issue of the credibility of the certificate. Alternatively, an independent NGO or an organization formed by citizens could grant the certificate. Hence, a citizen who plays C in the first period will be assumed to receive certification that identifies her action along with the socioeconomic distance of the match. With such an information system in place, type 2 citizens can have an incentive to contribute in the first period for some matches in which they would not contribute in a single-period model. This increases the scope of self-regulation directly through more contributions from the type 2s and indirectly by inducing the type 1s to contribute for more matches.

For type 2 citizens to contribute in the first period they must be able to free-ride on the contributions of the type 1s in the second period. This requires that (12) be satisfied, which is assumed here, and in addition (11) is assumed not to be satisfied to simplify the exposition. The analysis proceeds by conjecturing an equilibrium with pooling on an interval $(y_2^o, y^p]$ in the first period and determines the set of matches such that no citizen prefers to deviate. The intuition is developed here, and a proof of Proposition 5 below is presented in the Appendix.

For citizens who pool in the first period for a set of match distances, all their potential period-

two partners have the same beliefs about their type at the beginning of period two as at the beginning of period one. The period-two equilibrium for a match between such citizens then is the same as the single-period equilibrium characterized in Section III. Again the Pareto dominant equilibrium is considered, and the expected period-two utility for a type 2 with pooling is $EU_2^r(\beta)$ given in (14). If a type 2 chooses N in the first period for matches in the pooling interval $(y_2^o, y^p]$, his type is revealed, and in the second period no partner will contribute for matches $y > y_2^o$, as shown in the Appendix. The expected period-two utility then is EU_2^o given in (3) with η_2 and y_2^o replacing η and y^o , respectively. The utility difference ΔEU_2 for a type 2 citizen from playing C in period one versus playing N is then

$$\Delta EU_2 = b - c + (\delta + \theta(1 - \delta))xe^{-\eta_2 y} + \tau(EU_2^r(\beta) - EU_2^o), \quad (19)$$

where $\tau \in (0, 1]$ is the discount factor.

The term $b - c + xe^{-\eta_2 y}$ in (19) (for $\delta = 1$) is negative for $y \in (y_2^o, y_1^o]$, so to free ride in period two the type 2 citizen incurs a loss in period one. Note from (19) that the gain $EU_2^r(\beta) - EU_2^o$ from free-riding in the second period is independent of the match distance in the first period, whereas the first period loss is increasing in y . Consequently, a type 2 has an incentive to contribute for some y close to y_2^o . The strongest incentive for a type 2 to play N in the first period is for a match $y = y^p$, which has the largest period-one loss. For that match the type 2 will not deviate if

$$b - c + xe^{-\eta_2 y^p} + \tau(EU_2^r(\beta) - EU_2^o) \geq 0, \quad (20)$$

where $EU_2^r(\beta) - EU_2^o = b \left(\frac{e^{-\alpha y_2^o} - e^{-\alpha y_1^r(\beta; \theta)}}{1 - e^{-\alpha L}} \right)$. By definition of y_2^o , $b - c = -xe^{-\eta_2 y_2^o}$, and substituting this into (20) yields

$$x \left(e^{-\eta_2 y^p} - e^{-\eta_2 y_2^o} \right) \leq \tau b \left(\frac{e^{-\alpha y_2^o} - e^{-\alpha y_1^r(\beta; \theta)}}{1 - e^{-\alpha L}} \right). \quad (21)$$

The right side is positive and independent of y^p , whereas the left side is positive and decreasing in y^p . Consequently, for a y^p sufficiently close to y_2^o the inequality is satisfied. Let \bar{y}_2^p be defined by (21) as an equality. Then, for match $y \in [0, \bar{y}_2^p]$ the type 2s contribute in the first period and thus pool with the type 1s over a larger set of matches. The boundary \bar{y}_2^p is strictly increasing in τ , so the more important is the free-riding in the second period the larger is the first-period pooling interval.

As shown in the Appendix the type 1s gain in period two from having their type revealed in period one for matches $y \in (\bar{y}_2^p, y_1^r(\beta; \theta)]$ where they contribute and the type 2s do not contribute. This gain results because a revealed type 1 could be matched with a revealed type 1 in period

one, in which case they both contribute for $y \in [0, y_1^o]$. This gain provides an incentive for type 1s to expand their scope of self-regulation in period one beyond $y_1^r(\beta; \theta)$ to $\bar{y}_1^p(\beta; \theta)$, as defined and shown in the Appendix, so as to separate from the type 2s. A certification organization thus expands the scope of self-regulation for both types 1 and 2 in period one.

The equilibrium with a certification organization is characterized in the Appendix in conjunction with the proof of the following proposition.

Proposition 5: With reciprocal altruism and a certification organization, if (11) is not satisfied and (12) is satisfied, (A) an equilibrium exists in which all citizens contribute in the first period for matches $y \in [0, \bar{y}_2^p]$, $\bar{y}_2^p > y_2^o$. This induces type 1s to contribute for matches $y \in [0, \bar{y}_1^p(\beta; \theta)]$ in period one, where $\bar{y}_1^p(\beta; \theta) > y_1^r(\beta; \theta)$. (B) There is no equilibrium in which the types separate for all $y \in (y_2^o, y_1^o]$. (C) A certification organization has no effect on behavior if preferences reflect unconditional altruism.

A certification organization increases the scope of self-regulation in the first-period by expanding the self-regulation by the type 2s. This results because the 2s are in a dilemma. If they do not contribute for matches $y \in (y_2^o, \bar{y}_2^p]$ they reveal their type and will not be able to free ride on their partner in the second period for matches $y \in [y_2^o, y_1^r(\beta; \theta)]$. So the threat of being excluded by the type 1s in the second period expands the scope of self-regulation. Moreover, the cost of contributing in the first period is offset in part by the reciprocal altruism. The type 1s then are induced to expand their scope of self-regulation in the first period. The result requires that there is an opportunity for free-riding in period two, which requires that there are sufficient type 1s for (12) to be satisfied.

In period two the type 2s contribute only for $y \in [0, y_2^o]$, and the type 1s contribute for $y \in [0, y_2^r(\beta; \theta)]$ if either their or their partner's type was not revealed in period one. If a type 1 was revealed in period one by contributing for matches $(\bar{y}_2^p, \bar{y}_1^p(\beta; \theta)]$, she contributes for $y \in [0, y_1^o]$ if she is matched with another type 1 whose type was also revealed. If matched with a type 2 whose type was revealed, she contributes for $y \in [0, y_2^o]$. Letting $q = \frac{e^{-\bar{y}_2^p} - e^{-\alpha \bar{y}_1^p(\beta; \theta)}}{1 - e^{-\alpha L}}$ denote the probability of a first-period match $y \in (\bar{y}_2^p, \bar{y}_1^p]$, the expected scope of self-regulation for the type 1s is given in (A3) in the Appendix. The difference between that expected scope and the single-period scope of self-regulation $y_1^r(\beta; \theta)$ is

$$\beta q^2 (y_1^o - y_1^r(\beta; \theta)) + (1 - \beta) q^2 (y_2^o - y_1^r(\beta; \theta)). \quad (22)$$

A certification organization thus expands the scope of self-regulation in both periods if

$$\beta(y_1^o - y_1^r(\beta; \theta)) > (1 - \beta)(y_1^r(\beta; \theta) - y_2^o).$$

C. Enforcement Organizations

1. Non-Profit Enforcement

A social label organization has the sole function of screening the types of citizens, and it can expand the scope of self-regulation when citizens have reciprocal altruism. A certification organization expands the scope of self-regulation by providing information to future trading partners about a citizen's play in the first period. An organization could also have an enforcement capability. Citizens may be thought of as pledging to contribute in the face of incentives to free-ride, and enforcement may be thought of as raising the cost of breaking that pledge. The firms participating in the Fair Labor Association (FLA) have an incentive to shirk on meeting FLA standards for working conditions, so independent inspections are used and enforcement takes the form of internal reporting within the FLA and with board approval the release of the inspection reports to the public. The penalty or harm imposed for a broken pledge is not exogenous but instead is determined by the participants in the organization. Participation is assumed to be voluntary, so citizens subject themselves voluntarily to enforcement. Enforcement provided by a non-profit organization is considered in this section, and enforcement by a for-profit firm is considered in the next section.

Enforcement is assumed to take the form of punishment or harm h in the event that a citizen violates her pledge to play C .²² The harm could be reputation damage from public exposure in the case of the members of the FLA. Enforcement is assumed to be available everywhere on the circle, and the strength of enforcement is taken to be h .²³ Participation in the organization requires a payment, which for a non-profit organization equals its cost $f(h)$ per citizen of enforcement. That cost is assumed to be strictly increasing and convex in h . The participation decision is assumed to be made after the match has occurred but before the play of the public goods game.

Consider the case in which there is one type of citizen in society and preferences reflect

²² In this sense enforcement is analogous to a contract with a penalty for breach. The situations in which self-regulation occurs, however, are generally those in which a contract would be costly to enforce in a court. Moreover, the participants in an organization such as the FLA would be reluctant to turn jurisdiction over to a court.

²³ In a repeated game with random matching Kandori (1992) showed that cooperation is sustained if there is enough local punishment. See also Ellison (1994). Here punishment is administered by the organization, where participation in the organization is voluntary.

reciprocal altruism.²⁴ Citizens with matches $y \in [0, y^o]$ have a best response of contributing, so the only citizens with a demand for enforcement are those with more distant matches. To provide the toughest test for non-profit enforcement, assume that in the absence of the organization the Pareto dominant equilibrium is played. A citizen who pledges to contribute and demands enforcement will contribute if

$$(1 + \delta)b - c + (\delta + \theta(1 - \delta))xe^{-\eta y} - f(h) \geq \delta b - h,$$

so enforcement is demanded only if $f(h) < h$, and attention is restricted to that case. Citizens with matches $y \in (y^o, \hat{y}(1; \theta)]$ benefit from enforcement when $f(h) < h$, where

$$\hat{y}(\delta; \theta) = \begin{cases} 0 & \text{if } (\delta + \theta(1 - \theta))x \leq c + f(h) - b - h \\ \frac{1}{\eta} \ln\left(\frac{(\delta + \theta(1 - \delta))x}{c + f(h) - b - h}\right) & \text{if } (\delta + \theta(1 - \theta))x > c + f(h) - b - h. \end{cases} \quad (23)$$

The organization expands the scope of self-regulation for $h > f(h)$ and $\hat{y}(\delta; \theta)$ is decreasing in $f(h)$, so the scope of self-regulation is limited by the cost of organization and enforcement.²⁵

The non-profit organization can choose the strength h of its enforcement, and it is assumed to maximize the aggregate utility of those using its services. Their expected gain EU^n in the Pareto dominant equilibrium ($\delta = 1$) is

$$\begin{aligned} EU^n &= \int_{y^o}^{\hat{y}(1; \theta)} (2b - c + xe^{-\eta y} - f(h)) \frac{\alpha e^{-\alpha y}}{(1 - e^{-\alpha L})} dy \\ &= \frac{1}{1 - e^{-\alpha L}} \left[(2b - c - f(h)) (e^{-\alpha y^o} - e^{-\alpha \hat{y}(1; \theta)}) + \left(\frac{x\alpha}{\alpha + \eta} \right) (e^{-(\alpha + \eta)y^o} - e^{-(\alpha + \eta)\hat{y}(1; \theta)}) \right]. \end{aligned}$$

The expected utility EU^n is strictly concave, and the optimal strength h^* of enforcement then satisfies the first-order condition for $\hat{y}(1; \theta) < L$

$$(b - h^*)\alpha e^{-\alpha \hat{y}(1; \theta)} \left(\frac{1 - f'(h^*)}{\eta(c + f(h^*) - b - h^*)} \right) - f'(h^*) (e^{-\alpha y^o} - e^{-\alpha \hat{y}(1; \theta)}) = 0. \quad (24)$$

A sufficient condition for the second-order condition to be satisfied is $\alpha \geq \eta$. The second term in (24) is the marginal cost of enforcement, and the first term is the marginal gain from enforcement.

²⁴ The results in this section also hold for unconditional altruism by letting $\theta = 1$.

²⁵ If $f(h)$ is paid ex ante as in the case of insurance, it is sunk and does not affect the ex post enforcement choice. For the Pareto dominant equilibrium, enforcement then takes place on an interval $(y^o, \tilde{y}(1; \theta)]$, where $\tilde{y}(1; \theta) = L$ if $c - b - h \leq 0$ and otherwise

$$\tilde{y}(1; \theta) \equiv \begin{cases} 0 & \text{if } x \leq c - b - h \\ \frac{1}{\eta} \ln\left(\frac{x}{c - b - h}\right) & \text{if } x > c - b - h. \end{cases}$$

Since $\tilde{y}(1; \theta) > \hat{y}(1; \theta)$, when $f(h)$ is sunk, the scope of self-regulation is greater than when $f(h)$ is paid ex post.

The marginal effect of enforcement on the scope of self-regulation is proportional to $1 - f'(h^*)$, and $b - h^*$ is the utility of the citizen in the most distant match $y = \hat{y}(1; \theta)$ for which enforcement occurs.

The optimal enforcement by the non-profit organization when $\hat{y}(1; \theta) < L$ is summarized in the following proposition, which is proven in conjunction with the proof of Proposition 7 below.

Proposition 6: The optimal strength h^* of enforcement for a non-profit organization has the following properties:

- (i) $f'(h^*) < 1$; (ii) $f(h^*) < h^* < b$; (iii) $b - h^* > (\leq) 2b - c - f(h^*) \iff \hat{y}(1; \theta) < (=) L$;
- (iv) h^* is independent of x and θ ; (v) $\hat{y}(1; \theta) > y^o$, so the scope of self-regulation is increased by non-profit enforcement; (vi) A sufficient condition for h^* to be decreasing in b is $f'(h^*) \geq \frac{1}{2}$, and a sufficient condition for h^* to be increasing in c is $\alpha \geq \eta$.

Enforcement extends to the point at which its marginal cost is less than its marginal effect on self-regulation; i.e., $f'(h^*) < 1$, and the scope of self-regulation is increased by enforcement since $h^* > f(h^*)$. The enforcement policy is independent of θ , since contributions occur for matches $y \in [0, \hat{y}(1; \theta)]$. Enforcement is also independent of x , since that parameter affects y^o and $\hat{y}(1; \theta)$ in the same proportion. If the cost of enforcement is sufficiently low, the optimal enforcement could result in contributions by all citizens. The non-profit organization then chooses the strength h^* of enforcement to minimize $f(h)$ and satisfy

$$c + f(h^*) - b - h^* = x.$$

If h^* is decreasing in b and increasing in c in Proposition 6 (vi), self-regulation crowds out non-profit enforcement. That is, an increase in the quality of the public good increases the scope of self-regulation which decreases the strength of non-profit enforcement. A higher quality public good has a direct effect of increasing $\hat{y}(1; \theta)$ and an indirect effect of decreasing $\hat{y}(1; \theta)$ by decreasing h^* . The effect on the demand for non-profit enforcement thus depends on the specific values of the parameters.

2. For-Profit Enforcement

A society could rely on the for-profit sector rather than non-profit organizations for enforcement. For-profit enforcement is common in providing security, and industry associations such as the FLA require that independent organizations conduct the required inspections. This section explores whether enforcement of the form in the preceding section will be supplied in the market-

place by a profit-maximizing firm and how that enforcement compares with the enforcement by a non-profit organization. Enforcement on demand is again considered.

The firm chooses the strength h of enforcement and the price p for enforcement. Assume that the cost of enforcement is $f(h)$. The citizens who demand enforcement are those with matches $y \in (y^o, \hat{y}^\pi(1; \theta)]$, where $\hat{y}^\pi(1; \theta)$ is defined as in (23) with p replacing $f(h)$, provided that $p < b + h - c + xe^{-\alpha \hat{y}^\pi(1; \theta)}$. The firm thus has demand for its services only if it increases the scope of self-regulation. The revenue R of the firm is²⁶

$$R = p \left(\frac{e^{-\alpha y^o} - e^{-\alpha \hat{y}^\pi(1; \theta)}}{1 - e^{-\alpha L}} \right),$$

and if the cost $f(h)$ is incurred only for those citizens who demand enforcement, the profit Π of the firm is given by

$$\Pi = (p - f(h)) \left(\frac{e^{-\alpha y^o} - e^{-\alpha \hat{y}^\pi(1; \theta)}}{1 - e^{-\alpha L}} \right).$$

In contrast to non-profit enforcement, the firm does not take into account the citizens' utility from altruism other than through its effect on demand.

The optimal price is characterized first and then the optimal enforcement is characterized. The first-order condition for the optimal price $\hat{p}(h)$ is

$$e^{-\alpha y^o} - e^{-\alpha \hat{y}^\pi(1; \theta)} + (\hat{p}(h) - f(h)) \alpha e^{-\alpha \hat{y}^\pi(1; \theta)} \frac{d\hat{y}^\pi(1; \theta)}{dp} = 0, \quad (25)$$

where

$$\frac{d\hat{y}^\pi(1; \theta)}{dp} = -\frac{1}{\eta(c + \hat{p}(h) - b - h)} < 0.$$

A sufficient but not necessary condition for the second-order condition to be satisfied is $\alpha \geq \eta$. The condition in (25) implies that $\hat{p}(h) > f(h)$, so the price is greater than cost of enforcement. A sufficient condition for the price to be strictly increasing in h is $\alpha \geq \eta$. The profit-maximizing enforcement \hat{h} satisfies the first-order condition

$$(\hat{p}(\hat{h}) - f(\hat{h})) \alpha e^{-\alpha \hat{y}^\pi(1; \theta)} \left(\frac{1}{\eta(c + \hat{p}(\hat{h}) - b - \hat{h})} \right) - f'(\hat{h}) \left(e^{-\alpha y^o} - e^{-\alpha \hat{y}^\pi(1; \theta)} \right) = 0. \quad (26)$$

The properties of the equilibrium are summarized in the following proposition and related to those with non-profit enforcement, and the proofs are in the Appendix.

Proposition 7: (A) The optimal enforcement policy of a profit-maximizing firm satisfies:

²⁶ This assumes that the firm cannot observe the match distance and price discriminate based on y .

(i) $f'(\hat{h}) = 1$, (ii) $\hat{p}'(\hat{h}) = 1$; (iii) $\hat{p}(\hat{h}) > f(\hat{h})$; (iv) $\hat{y}^\pi(1; \theta) > y^o$; (v) $\frac{d\hat{p}(\hat{h})}{d\hat{h}} - \frac{d\hat{h}}{d\theta} > 0$; $\frac{d\hat{p}(\hat{h})}{dc} - \frac{d\hat{h}}{dc} < 0$; (vi) $\frac{d\hat{y}^\pi(1; \theta)}{db} < 0$; $\frac{d\hat{y}^\pi(1; \theta)}{dc} > 0$; (vii) $\hat{p}(h)$ and \hat{h} are independent of x and θ .

(B) The enforcement policies of the non-profit organization and the profit-maximizing firm have the following relations:

(i) $\hat{h} > h^*$; (ii) $f(\hat{h}) > f'(h^*)$; (iii) $\hat{p}(\hat{h}) > f(h^*)$.

At the optimal for-profit enforcement strength \hat{h} , the marginal price equals the marginal cost, since p and h are perfect substitutes in the boundary $\hat{y}^\pi(1; \theta)$ of enforcement, so only the difference between $\hat{p} = \hat{p}(\hat{h})$ and \hat{h} affects demand. An increase in the net benefits from the public good increases the difference between \hat{p} and \hat{h} , which decreases the demand for enforcement. That is, the firm responds to a higher quality public good by increasing price by more than it increases the strength of enforcement.

Enforcement by a for-profit firm is more aggressive than that by a non-profit organization. This results because the firm has a first-order incentive to increase its price, but that reduces demand. Demand can be increased by more stringent enforcement, which is carried to the point at which the marginal cost equals the marginal revenue product of enforcement. Since the firm is profitable, entry would be expected which would drive down the price, which would lead the firm to provide less aggressive enforcement. The demand for for-profit enforcement is thus crowded out by self-regulation.

VI. Social Pressure

An alternative to public regulation and organized self-regulation by a non-profit organization or for-profit firm is reliance on social pressure to strengthen the incentives to self-regulate. Activists and NGOs direct social pressure to economic actors for, in the context of this model, their failure to provide public goods, or mitigate harmful externalities, associated with their economic activity. Environmental NGOs in particular pressure firms to reduce the harmful environmental impacts of their activities. The basic instrument of activist NGOs is “naming and shaming.” That is, identifying a citizen who has failed to provide the public good and informing other citizens of that failure, resulting in shame. This social pressure typically is funded by voluntary donations by citizens, and those donations face their own free-rider problem. Nevertheless, many environmental NGOs are well-funded by membership dues and donations, which are encouraged by tax-deductibility.

This section introduces an activist funded by voluntary donations by citizens. Those donations support the naming and shaming of citizens who fail to provide the public good associated with

their trades. Social pressure and naming and shaming thus mitigate the public goods problem by imposing harm in the form of shame. The harm to the citizen could vary depending on the nature of the public goods problem. For example, an environmental issue could result in more harm than a working conditions issue in overseas factories. Epstein and Schnietz (2002) found that the protests at the failed 1999 Seattle WTO talks resulted in a statistically significant decrease in the market values of firms targeted as environmentally abusive and but had no significant effect on the market value of firms targeted as having abusive labor condition in overseas factories. Similarly, the harm from not addressing an environmental issue could vary by industry. In contrast to the enforcement models in the previous section in which participation by a citizen was voluntary, all citizens are subject to the possibility of naming and shaming by the activist. Citizen-funded self-regulation thus serves as a probabilistic commitment mechanism to mitigate the free-rider problem.

Assume that each citizen can donate an amount a to the activist, and the total donations A received are used to detect the play in a match with probability $q = Q(A)$, where $Q(0) = 0$ and $Q'(A) > 0$. When play is detected, the activist can make public the play of N , resulting in harm h to the citizen. The harm could depend on the action of other citizens. For example, a firm may not incur significant harm if it is revealed to have played N when all other firms in its industry also played N . In contrast, if the other firms played C , the firm could incur significant harm. In the latter case the harm results from the public shame of having been disclosed as having played N when others played C . Shame is considered here.

The contributions to the activist are made ex ante before the trade and the public goods game take place. With reciprocal altruism and the threat of incurring shame, a citizen contributes if²⁷

$$(1 + \delta)b - c + (\delta + \theta(1 - \delta))xe^{-\eta y} \geq \delta(b - qh). \quad (27)$$

If $qh > c - b$, the expected social pressure is sufficiently great that citizens contribute to the public good for all L . The focus is on the case in which the detection probability and social pressure are not that strong. The equilibrium pure strategy of a citizen in the Pareto dominant equilibrium then is

$$S^* = \begin{cases} C & \text{if } y \leq y^q(\delta; \theta) \\ N & \text{if } y > y^q(\delta; \theta), \end{cases}$$

where

$$y^q(\delta; \theta) \equiv \begin{cases} 0 & \text{if } (\delta + \theta(1 - \delta))x \leq c - b - \delta qh \\ \frac{1}{\eta} \ln\left(\frac{(\delta + \theta(1 - \delta))x}{c - b - \delta qh}\right) & \text{if } (\delta + \theta(1 - \delta))x > c - b - \delta qh. \end{cases}$$

²⁷ If the activist makes public the play (N, N) , the right side of (27) is replaced by $\delta b - qh$. The equilibrium is the same as with shame when the partner contributes with probability one.

The boundary $y^q(\delta; \theta)$ is strictly increasing and strictly convex in q with $y^0 = y^\theta$. Consequently, greater donations to the activist expand the scope of self-regulation.

With social pressure all citizens contribute for matches $y \in [0, y^q(1; \theta)]$ in the Pareto dominant equilibrium. A citizen recognizes that her donation to the activist induces other citizens as well as herself to contribute for a larger set of matches. As discussed in Section II a citizen's altruism pertains to the increase in utility her actions provide to others. In addition to the benefits provided to her partner by contributing to the public good, her donation increases social pressure which induces contributions for a larger set of matches. The gain she provides to the two citizens induced to contribute in a match through her donation is $4b - 2c + 2xe^{-\eta y}$. This gain is provided for matches $y \in (y^{q'}(\delta; \theta), y^q(\delta; \theta)]$, where q' is the detection probability without her contribution and q is the probability with her contribution. As an approximation view each citizen as an atom, and assume that there are $2M$ citizens. Thus, $q' = Q(\sum_{j \neq i} a_j)$ and $q = Q(\sum_j a_j)$, where a_j is the contribution of citizen j . The expected utility EU_i^A of a citizen i in the Pareto dominant equilibrium is then²⁸

$$EU_i^A = \int_0^{y^{q'(1; \theta)}} (2b - c + xe^{-\eta y}) \left(\frac{\alpha e^{-\alpha y}}{1 - e^{-\alpha L}} \right) dy + \int_{y^{q'(1; \theta)}}^{y^q(1; \theta)} M (4b - 2c + 2xe^{-\eta y}) \left(\frac{\alpha e^{-\alpha y}}{1 - e^{-\alpha L}} \right) dy - a_i. \quad (28)$$

Donations thus are motivated in (28) by the gains to citizens in the additional matches in which they contribute in the public goods game.

The optimal donation by a citizen satisfies the first-order condition which in a symmetric equilibrium is

$$2M(b - Q(A^*)h) \left(\frac{\alpha e^{-\alpha y^{q(1; \theta)}}}{1 - e^{-\alpha L}} \right) \left(\frac{Q'(A^*)h}{\eta(c - b - Q(A^*)h)} \right) - 1 = 0, \quad (29)$$

where $A^* = \sum_i a_i^*$ the optimal contribution. A sufficient condition for the second-order condition to be satisfied is $Q(A)$ concave and $\alpha \geq \eta$.²⁹ The effect of the donation on the scope of self-regulation is valued at the marginal incentive $(b - Q(A^*)h)$ to contribute to the public good multiplied by the marginal probability $\frac{\alpha e^{-\alpha y^{q(1; \theta)}}}{1 - e^{-\alpha L}}$ of a match in the expanded set of matches on which contributions

²⁸ This specification is consistent with $x = b$.

²⁹ The second-order condition is

$$\frac{d^2 EU_i^A}{da_i^2} = 2M(b - Q(A^*)h) \left(\frac{\alpha e^{-\alpha y^{q(1; \theta)}}}{1 - e^{-\alpha L}} \right) \left[(-\alpha + \eta) \left(\frac{Q'(A^*)h}{\eta(c - b - Q(A^*)h)} \right)^2 + \frac{hQ''(A^*)}{\eta(c - b - Q(A^*)h)^2} \right] < 0.$$

take place.³⁰ The form of the first-order condition in (29) is directly analogous to that in (24) for non-profit enforcement and in (26) for for-profit enforcement. All three first-order conditions equate the marginal cost to the organization or donor to the marginal benefit which is composed of the benefit to the citizen at the boundary of contributions, the marginal probability of being at the boundary scaled by the interval on which enforcement is applied, and the marginal effect of enforcement on the boundary of contributions.

Since each citizen takes the utility of all citizens into account at the margin, the social pressure is collectively optimal. Although social pressure expands the scope of self-regulation, moral preferences are insufficient to resolve the free-rider problem unless the detection probability is sufficiently high. The results of this section are summarized as:

Proposition 8: Social pressure funded by voluntary donations increases the scope of self-regulation. With reciprocal or unconditional altruism donations to the activist are collectively optimal for the citizens given the technology $Q(A)$, but social pressure is second-best unless $Q(A^*)h \geq c - b$. The equilibrium contribution a^* is strictly decreasing in x , and $\frac{da^*}{dc} > (=) (<) 0$ as $\alpha > (=) (<) \eta$ and $\frac{da^*}{db} > 0$ if $\alpha \leq \eta$.

Stronger moral preferences (higher x) decrease social pressure, i.e., crowds out social activism. In contrast a higher quality public good increases social pressure if $\alpha \leq \eta$.

VIII. Conclusions

Self-regulation can result from a variety of motivations, including self-interest, forestalling public or private politics, and moral. Two forms of moral preferences that seem natural are limited morality and reciprocal altruism. Limited morality can overcome the incentive to free ride for interactions among citizens who are close on some dimension but not for those who are distant. The scope of self-regulation is increasing in both the strength of those moral preferences and the quality of the public good.

The private provision of public goods with moral preferences faces two types of free-rider problems. The first occurs when both players have incentives not to provide the public good. The second occurs in a heterogeneous society in which citizens with stronger moral preferences provide the public good and those with weaker moral preferences free ride. The second free-rider problem reduces the incentives of those with stronger moral preferences to provide the public good.

³⁰ If the detection probability is sufficiently high that $b - Q(A^*)h \leq 0$, all citizens contribute to the public good. Then, (29) is not satisfied, and citizens do not donate to the activist at the margin. Citizens then have a participation game.

With heterogeneous moral preferences reciprocal altruism results in a smaller scope of self-regulation than with unconditional altruism. This is due to the second free rider problem – citizens with weaker moral preferences free ride on the private provision of the public good by those with stronger moral preferences. This then limits the set of matches for which those with the stronger moral preferences provide the public good. Organizations can help mitigate this second free-rider problem and increase the private provision of the public good.

When moral preferences reflect reciprocal altruism, social label and certification organizations increase the scope of self-regulation. A social label organization allows those with similar preferences to interact among themselves. This reduces the free riding by those with weaker moral preferences, which then elicits more private provision from those with stronger moral preferences. A social label organization expands the scope of self-regulation by allowing separation of the types of citizens, and a certification organization expands self-regulation by inducing pooling by those citizens with weaker moral preferences. Pooling results because the opportunity to free ride in the second period outweighs the loss in the first period, which for some matches is small because of altruism. The pooling by those with weaker moral preferences creates an opportunity for those with stronger moral preferences to separate from those with weaker moral preferences by providing the public good for additional matches. Neither a social label organization nor a certification organization, however, can expand the scope of self-regulation beyond that which would result with unconditional altruism.

A private enforcement organization can expand the scope of self-regulation beyond that with unconditional altruism by imposing harm on a citizen who fails to provide the public good. Enforcement can be provided by both non-profit and for-profit organizations. A profit-maximizing firm provides stronger enforcement than does a non-profit organization, but it charges a price higher than the (average cost) fee required by the non-profit organization. The scope of self-regulation is decreasing in the quality of the public good because the profit-maximizing firm increases its price by more than it increases the strength of its enforcement.

An alternative to public regulation and voluntary, private organization is to rely on social pressure by an activist NGO funded by voluntarily donations by citizens. Naming and shaming can impose harm on citizens by publicly disclosing their failure to provide the public good. Citizens face a free-rider problem on their donations to the NGO, but their altruism overcomes this problem. The scope of self-regulation is increased by the social pressure, but unless the detection probability is high, the free rider problem is mitigated but not eliminated.

Public regulation is an alternative to voluntary measures, but it can be crowded out by self-regulation. From a normative perspective self-regulation reduces the number of matches for which public regulation is beneficial. From a positive perspective the political support for regulation is reduced by self-regulation, and if the scope of self-regulation is greater than one-half, regulation cannot be adopted. Since private organizations expand the scope of self-regulation, they also reduce the support for public regulation and can also reduce the contributions to fund social pressure. Self-regulation can also crowd out private organizations, but those organizations do not require majority approval, since they can be formed by a minority. The demand for public regulation is greater with reciprocal than with unconditional altruism, but so is the incentive to form a private self-regulation organization.

Self-regulation can be motivated by moral considerations as well as by a variety of self-interested considerations. A challenge for empirical analysis is to devise methods to identify the motivation that underlies the private provision of public goods.

Appendix

Period-Two Equilibrium with a Certification Organization

As in Section V.B assume that (11) is not satisfied and (12) is satisfied, so in a single-period model a type 2 has an incentive to free ride for $y \in (y_2^o, y_1^r(\beta; \theta)]$. In the first period separation occurs for matches $y \in [y_1^r(\beta; \theta), \bar{y}_1^p(\beta; \theta)]$, where $\bar{y}_1^p(\beta; \theta)$ is defined below. The equilibrium in the second period depends on whether the types of citizens have been revealed or not. If they have not been revealed, the citizen's beliefs about the type of her partner are the same as her prior beliefs. The second-period equilibrium is then as characterized for the single-period model in Section III.

Consider a period-two match of a citizen whose type has been revealed and a citizen whose type has not been revealed. A revealed type 2 has a best response of contributing for $y \in [0, y_2^o]$, since both a type 1 and a type 2 partner have a best response of contributing for $y \in [0, y_2^o]$. A revealed type 1 has a best response of contributing for $y \in [0, y_1^r(\beta; \theta)]$, since a type 1 partner's best response is to contribute for $y \in [0, y_1^r(\beta; \theta)]$ and a type 2 partner's best response is to contribute for $y \in [0, y_2^o]$. A type 1 citizen whose type has not been revealed thus has a best response of contributing for $y \in [0, y_1^r(\beta; \theta)]$ with a revealed type 1 and for matches $y \in [0, y_2^o]$ with a revealed type 2. A type 2 citizen whose type has not been revealed contributes only for all matches $y \in [0, y_2^o]$.

Next consider a match of a revealed type i and revealed type j , $i = 1, 2, j = 1, 2$. Two type 1s have best responses to contribute for $y \in [0, y_1^o]$, and two type 2's have best responses of contributing for $y \in [0, y_2^o]$. A type 1 and a type 2 have best responses of contributing for $y \in [0, y_2^o]$, since (11) is not satisfied.

In period two a revealed type 2 thus is in equilibria in which his partner contributes only for $y \in [0, y_2^o]$. His expected utility EU_2^o is thus given by (3) with η_2 and y_2^o replacing η and y^o , respectively.

A revealed type 1 is in equilibria in which both players contribute for $y \in [0, y_2^o]$, and with probability βq she is in equilibria in which she and her partner contribute for $y \in [0, y_1^o]$. With probability $(1 - \beta)q$ she is in equilibria in which she and her type 2 partner contribute for $y \in [0, y_2^o]$, where

$$q = \frac{e^{-\alpha y_1^r(\beta; \theta)} - e^{-\alpha \bar{y}_1^p(\beta; \theta)}}{1 - e^{-\alpha L}}$$

is the probability that a citizen's type is revealed in the first period.

The period-two expected utility $EU_1^p(\beta)$ of a revealed type 1 is thus

$$\begin{aligned}
EU_1^p(\beta) &= \left[(1-q) \int_0^{y_2^o} (2b-c+xe^{-\eta_1 y}) + (1-q) \int_{y_2^o}^{y_1^r(\beta;\theta)} (b(1+\beta)-c+(\beta+\theta(1-\beta))xe^{-\eta_1 y}) \right. \\
&\quad \left. + \beta q \int_0^{y_1^o} (2b-c+xe^{-\eta_1 y}) + (1-\beta)q \int_0^{y_2^o} (2b-c+xe^{-\eta_1 y}) \right] \left(\frac{\alpha e^{-\alpha y}}{1-e^{-\alpha L}} \right) dy \\
&= EU_1^r(\beta) - q(1-\beta) \int_{y_2^o}^{y_1^r(\beta;\theta)} (b-c+\theta xe^{-\eta_1 y}) \left(\frac{\alpha e^{-\alpha y}}{1-e^{-\alpha L}} \right) dy \\
&\quad + q\beta \int_{y_1^r(\beta;\theta)}^{y_1^o} (2b-c+xe^{-\eta_1 y}) \left(\frac{\alpha e^{-\alpha y}}{1-e^{-\alpha L}} \right) dy.
\end{aligned} \tag{A1}$$

Since $2b > c$ and $b-c+\theta xe^{-\eta_1 y} < 0$ for $y \in (y_2^o, y_1^r(\beta;\theta)]$ when (11) is not satisfied, $EU_1^p(\beta) > EU_1^r(\beta)$.

The gain in period two identified in (A1) provides an incentive for a type 1 to expand the set for which she contributes. A type 1 then plays C in period one for matches such that

$$b-c+(\beta+\theta(1-\beta))xe^{-\eta_1 y} + \tau(EU_1^p(\beta) - EU_1^r(\beta)) \geq 0. \tag{A2}$$

Defining $\bar{y}_1^p(\beta;\theta)$ by the equality in (A2) (where q is a function of $\bar{y}_1^p(\beta;\theta)$), a type 1 contributes for $y \in [0, \max\{\bar{y}_1^p(\beta;\theta), y_1^o\}]$ in period one. This does not affect the strategy of a type 2 in period one. The expected utility for a type 2 whose type is not revealed in period one is $EU_2^r(\beta)$.

The scope of self-regulation in period two: A type 2 contributes for $y \in [0, y_2^o]$ regardless of whether his type was revealed by his play in period one. A type 1 contributes for $y \in [0, y_1^r(\beta;\theta)]$ if her type is not revealed or her partner's type is not revealed. If both are revealed, a type 1 contributes for $y \in [0, y_1^o]$ when matched with a type 1. The probability that the types of both partners were revealed in period one is q^2 , so the expected scope of self-regulation by type 1s is

$$(1-q^2)y_1^r(\beta;\theta) + \beta q^2 y_1^o + (1-\beta)q^2 y_2^o, \tag{A3}$$

which is used to obtain (22).

Proof of Proposition 5: (A) Conjecture an equilibrium in which all citizens contribute for matches $y \leq \bar{y}_2^p$. The play in the first period thus provides no information about the type of a citizen with a match in that interval, so the beliefs of all citizens are the same as their prior beliefs. The equilibrium in the second period is then that characterized in Section III in the Pareto dominant equilibrium. The expected period-two utilities are then $EU_1^r(\beta)$ and $EU_2^r(\beta)$ given in (13) and (14), respectively.

A type 2 has a best-response strategy of playing C for $y \in [0, y_2^o]$ in period one, so consider a $y \in (y_2^o, \bar{y}_2^p]$. If a type 2 deviates by playing N , she avoids the loss $b - c + xe^{-\eta_2 y}$ in period one and is revealed as a type 2 under the standard off-the-equilibrium-path belief refinements. Consider the Pareto dominant period-two equilibrium in which the revealed type 2 plays C for $y \in [0, y_2^o]$. As shown above in the proof of the period-two equilibrium any period-two partner has a best response of contributing for $y \leq y_2^o$ and not contributing for $y > y_2^o$. As argued in the text for $y \in (y_2^o, \bar{y}_2^p)$, the gain in period one from playing N is exceeded by the loss in period two. A type 2 thus has no incentive to deviate by playing N on $y \in [0, \bar{y}_2^p]$.

Similarly, a type 2 has no incentive to deviate by playing C for $y > \bar{y}_2^p$. If he deviates and is believed to be a type 1, in the second period his expected utility will be $EU_2^*(\beta)$ rather than EU_2^o if he had played N . By definition of \bar{y}_2^p a type 2 has no incentive to deviate by playing N .

Consider a type 1 in period one. If the type 1 deviates and plays N for a match $y \in [0, \bar{y}_2^p]$ her period one utility is lower than if she plays C , since $b - c + xe^{-\eta_1 y} > 0$. Also, citizens believe that she is a type 2 under the standard off-the-equilibrium-path belief refinements, and her period two partners will contribute only for matches $y \in [0, y_2^o]$. A type 1 thus has no incentive to deviate by playing N for $y \in [0, \bar{y}_2^p]$.

Consider a deviation by a type 1 of playing C for a period one match $y > \bar{y}_1^p(\beta; \theta)$. This results in a lower utility in period 1 and reveals the citizen as a type 1. As shown in the proof of the period-two equilibrium, a type 1 cannot gain for $y > \bar{y}_1^p(\beta; \theta)$.

(B) To show that there is no separating equilibrium, consider the incentives of a type 2 to contribute for some $y > y_2^o$. A type 2 gains $\tau b \left(\frac{e^{-\alpha y_2^o} - e^{-\alpha y_1^*(\beta; \theta)}}{1 - e^{-\alpha L}} \right)$ from free riding on the 1s in period two and loses $c - b - xe^{-\eta_2 y}$ in the first period. As argued in the context of (21), for some $y > y_2^o$ the loss in period one is exceeded by the gain in period 2, so a type 2 will play C . Q.E.D.

(C) With unconditional altruism a citizen has a dominant strategy in both periods.

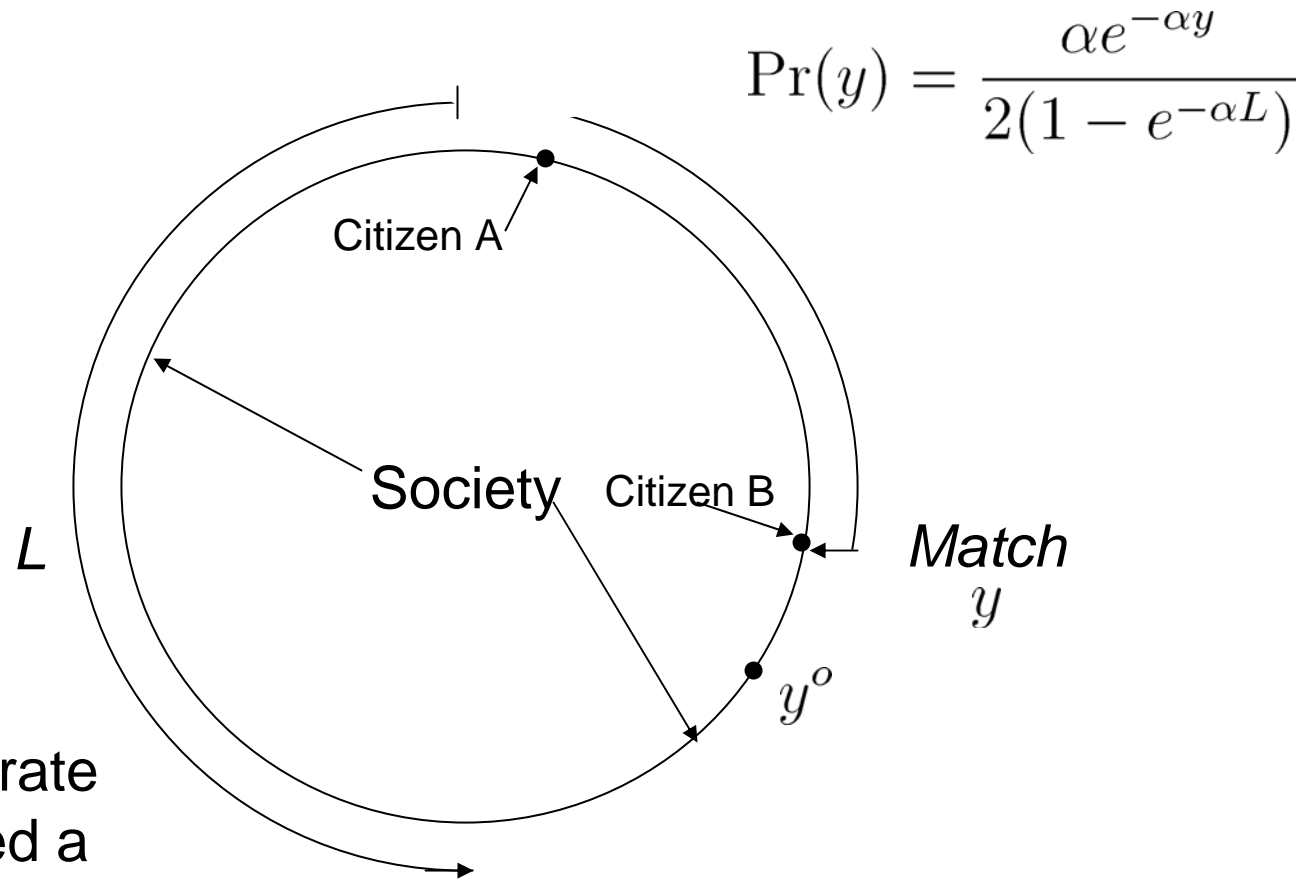
Proof of Proposition 7: A(i) follows from substituting (25) into (26). Differentiating $\hat{p}(h)$ in (25) and using A(i) yields A(ii). A(iii) is implied by (26) and the requirement that the firm is profitable. A(iv) is implied by (26) given A(iii). A(v) follows from totally differentiating (25) and (26). A(vi) follows because $\hat{y}^\pi(1; \theta)$ is decreasing in $\hat{p}(h) - h$. A(vii) results because (25) and (26) are independent of x and independent of θ when $\delta = 1$.

Proof of Proposition 6: The derivative $\frac{dEU^n}{dh}$ in (24) evaluated at $h = \hat{h}$ is negative, which given the strict concavity of EU^n implies that $h^* < \hat{h}$. Then, since $f(h)$ is strictly increasing,

$f(h^*) < f(\hat{h}) = 1$. Then, $b > h^*$ is implied by (24). Property (iii) in Proposition 6 then follows directly from $\hat{y}(1; \theta) < L$. That h^* is independent of x and θ follows directly from (24), which is independent of x and of θ when $\delta = 1$. Property (vi) follows directly from differentiating (24) and simplifying.

Parts B(i) and (ii) of Proposition 7 have been established above, and (iii) is immediate from (26) and B(ii). Q.E.D.

Figure 1
Society and Matching (continuum of citizens located on a circle)



L is how disparate or fractionalized a society is.

Figure 2
Representation of Moral
(Altruistic) Preferences

	Generalized	Limited
Unconditional	x	$xe^{-\eta y}$
Conditional	θx	$\theta xe^{-\eta y}$

$$\theta \in [0, 1)$$

Figure 3 Limited Morality, Unconditional Altruism

Citizen B

		Contribute	Not contribute
Citizen A	Contribute	$2b - c + xe^{-\eta y}$	$b - c + xe^{-\eta y}$
	Not contribute	b	0
		$2b - c + xe^{-\eta y}$	b
		b	0

Strategic
neutrality

Assumptions: $b - c + xe^{-\eta L} < 0$

$b - c + x > 0$

References

- Andreoni, James. 1988. "Privately provided public goods in a large economy: the limits of altruism." *Journal of Public Economics*. 35: 57-73.
- Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm Glow Giving." *Economic Journal*. 100: 464-477.
- Andreoni, James and John Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*. 70: 737-753.
- Banfield, E.C. 1958. *The Moral Basis of a Backward Society*. New York: The Free Press.
- Baron, David P. 2007a. "Corporate Social Responsibility and Social Entrepreneurship." *Journal of Economics and Management Strategy*. 16: 683-717.
- Baron, David P. 2007b. "Managerial Contracting and Corporate Social Responsibility." *Journal of Public Economics*. (forthcoming)
- Baron, David P. 2007c. "A Positive Theory of Moral Management, Social Pressure, and Corporate Social Performance." *Journal of Economics and Management Strategy*. (forthcoming).
- Baron, David P. 2007d. "Credence Standards and Social Pressure." Working paper, Stanford University.
- Baron, David P. and Daniel Diermeier. 2007. "Strategic Activism and Nonmarket Strategy." *Journal of Economics and Management Strategy*. 16: 599-634.
- Bohnet, Iria and Bruno S. Frey. 1999. "Social Distance and Other-Regarding Behavior in Dictator Games: Comment." *American Economic Review*. 89: 335-339.
- Calveras, Aleix, Juan-Jose Ganuza, and Gerard Llobet. 2007. "Regulation, Corporate Social Responsibility and Activism." *Journal of Economics and Management Strategy*. 16: 719-740.
- Dixit, Avinash. 2003. "Trade Expansion and Contract Enforcement." *Journal of Political Economy*. 111: 1293-1317.
- Dixit, Avinash. 2004. *Lawlessness and Economics*. Princeton University Press: Princeton, NJ
- Ellison, Glenn. 1993. "Learning, Local Interaction, and Coordination." *Econometrica*. 61: 1047-1072.
- Ellison, Glenn. 1994. "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching." *Review of Economic Studies*. 61: 567-588.
- Epstein, Marc J. and Katherine Schnietz. 2002. "Measuring the Cost of Environmental and

- Labor Protest to Globalization: An Event Study of the Failed 1999 Seattle WTO Talks.” *International Trade Journal*. 16: 129-160.
- Eshel, Ilan, Larry Smauelson, and Avner Shaked. 1998. “Altruists, Egoists, and Hooligans in a Local Interaction Model.” *American Economic Review*. 88: 157-179.
- Graff Zivin, Joshua and Arthur Small. 2005. “A Modigliani-Miller Theory of Altruistic Corporate Social Responsibility.” *Topics in Economic Analysis & Policy*. 5: Article 10.
- Kandori, Michihiro. 1992. “Social Norms and Community Enforcement.” *Review of Economic Studies*. 59: 63-80.
- Keser, Claudia and Frans van Winden. 2000. “Conditional Cooperation and Voluntary Contributions to Public Goods.” *Scandinavian Journal of Economics*. 102: 23-39.
- King, Andrew A. and Michael J. Lenox. 2000. “Industry Self-Regulation with Sanctions: The Chemical Industry’s Responsible Care Program.” *Academy of Management Journal*. 43: 698-716.
- King, Andrew A. and Michael J. Lenox. 2002. “Exploring the Locus of Profitable Pollution Reduction.” *Management Science*. 48: 289-299.
- Kotchen, Matthew J. 2006a. “Voluntary Provision of Public Goods for Bads: A Theory of Environmental Offsets.” Working paper, University of California–Santa Barbara.
- Kotchen, Matthew J. 2006b. “Green Markets and Private Provision of Public Goods.” *Journal of Political Economy*. 114: 816-834.
- La Ferrara, Eliana. 2003. “Kin Groups and Reciprocity: A Model of Credit Transactions in Ghana.” *American Economic Review*. 93: 1730-1759.
- Levine, David K. 1998. “Modeling Altruism and Spitefulness in Experiments.” *Review of Economic Dynamics*. 1: 593-622.
- Levy, Gilat and Ronny Razin. 2007. “A Theory of Religious Organizations.” Working paper, London School of Economics, London, UK.
- List, John A. 2007. “On the Interpretation of Giving in Dictator Games.” *Journal of Political Economy*. 115: 482-493.
- Lyon, Thomas P. and John W. Maxwell. 2004. *Corporate Environmentalism and Public Policy*. Cambridge, UK: Cambridge University Press.
- Maxwell, John W., Thomas P. Lyon, and Steven C. Hackett. 2000. “Self-Regulation and Social Welfare: The Political Economy of Corporate Environmentalism.” *Journal of Law and Economics*. 43: 583-618.

- Milgrom, Paul, Douglas North, and Barry Weingast. 1990. "The Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs," *Economics and Politics*, 2 (March):1-20.
- Prakash, Aseem and Matthew Potoski. 2006. *The Voluntary Environmentalists*. Cambridge, UK: Cambridge University Press.
- Prakash, Aseem and Matthew Potoski, eds. 2007. *Voluntary Programs: A Club Theory Perspective*. MIT Press, Cambridge, MA (forthcoming).
- Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*. 83: 1281-1302.
- Rabin, Matthew. 1998. "Psychology and Economics." *Journal of Economic Literature*. 36: 11-46.
- Siegel, Donald S. and Donald F. Vitaliano. 2007. "An Empirical Analysis of the Strategic Use of Corporate Social Responsibility." *Journal of Economics and Management Strategy*. 16: 773-792.
- Tabellini, Guido. 2007. "The Scope of Cooperation: norms and incentives." Working paper, Bocconi University.