

# Monitoring and Prosocial Contribution

Matthew Draper

Department of Political Science

University of California San Diego

## Abstract

This paper examines the variation in an individual's propensity to contribute to their society by engaging in group-oriented actions. I argue that the individual utility of contribution is a function of the cost of contribution, the reputational impact of (non)compliance with prevailing norms, and any individual satisfaction produced by the act of contribution itself. I further argue that social sanctions are themselves a function of the magnitude of group-oriented social norms and the extent of monitoring mechanisms to relay individual compliance decisions to the group. I construct and test a formal model incorporating these parameters. Agents in the model exhibit behavior that resembles the concepts of rapid norm shifts (Bicchieri 2006) and "norm cascades" (Keck and Sikkink 1998). Counterintuitive conclusions are developed, particularly the dependence of group-oriented norms on the presence of altruistic actors. Finally, I use the model developed in Kuran 1991 to introduce the concept of "contribution thresholds," and I explore deep structural resemblances between the concepts of revolution and prosocial contribution.<sup>1</sup>

**Word Count:** 7,277    **Keywords:** social norms, monitoring, prosocial, norm change, altruism

---

<sup>1</sup>The author benefitted greatly from conversations with David Lake, Kevin Rossillon and Geoff Hoffman.

# 1 Introduction

Different political systems inspire different kinds of participation. Some states seem better than others at persuading citizens to make socially useful contributions to their societies. What exactly induces people to make these kinds of prosocial contributions? Are political institutions playing a role in this heterogeneity? If so, might it be possible to design these institutions to either foster or frustrate group-oriented contributions?

People frequently face a choice between acting in the pursuit of individual gain or in the furtherance of their group's interest. We can label the second choice a "group-oriented" or prosocial action. I define a group-oriented action as an action by an individual with ramifications for the group. These actions are far broader than explicitly political actions, and group-oriented contributions might include volunteering time, donating money or coordinating support. What unifies these actions conceptually is the aim of fostering collective success and their nugatory benefit to the individual (Downs 1957). The only reason to undertake a group-oriented action is an interest in the welfare of the group as a whole. Of course, even generalized civic goodwill can be self-interested, given a long enough time horizon. More generally, perceptions of cooperation as zero-sum or positive-sum can dramatically affect willingness to engage in group-oriented activity.

Different states in different periods of history have had widely differing norms of participation and consent, and these patterns seem culturally sticky, outlasting particular regimes. We might be tempted to conclude that institutions simply enact prevailing local norms, and that states are powerless to shape the participatory environment within which they exist. However, recent examples of deliberately-chosen norm change are encouraging. Bogotá, a city that had become a byword for violence and insecurity, managed to implement a so-called *culutra ciudadana* emphasizing public participation in group-oriented activity and public shaming of violators. The result was a dramatic increase in social cohesion, reported trust, civic pride and even voluntary tax contribution (Riaño

2011). Bogotá achieved these effects through the harmonization of moral, social and legal norms, appealing to shared civic culture and publicly shaming those who did not voluntarily make prosocial contributions.

This paper will examine the variation in an individual’s propensity to contribute to their society by engaging in group-oriented actions. I argue that the individual utility of prosocial contribution is a function of the resources expended, the social benefits or sanctions ensuing from the decision, and the intrinsic individual satisfaction produced by contribution. I further argue that social sanctions are themselves a function of the magnitude of group-oriented social norms and the extent of monitoring mechanisms to relay individual compliance decisions to the group. I then construct a formal model incorporating these parameters. Agents in the model exhibit behavior that resembles the concepts of rapid norm shifts (Bicchieri 2006) and “norm cascades” (Keck and Sikkink 1998). Counterintuitive conclusions are developed, particularly the dependence of group-oriented norms on the presence of altruistic actors. Finally, I use the model developed in Kuran 1991 to explore the concept of “contribution thresholds,” as well as several startling parallels between the concepts of revolution and prosocial contribution.

## 2 Conceptual Preliminaries

We can say that a social norm exists when the socially-defined right of control over an action is held not by the actor but by others in the group (Coleman 1990 ch.10).<sup>2</sup> Social norms spread epidemiologically within “reference groups”, and culture is made up of those norms capable of replication without substantial transformation (Sperber 1996 ch. 4). For a norm to persist within a reference group, a sufficient proportion of individuals must share normative beliefs, empirical expectations

---

<sup>2</sup>Robert Axelrod defines a norm as a situation where parties are vengeful, in the sense of punishing defection, but timid, in the sense of being unwilling to risk exposure as a defector (Axelrod 1986).

and conditional preferences for conformity (Bicchieri 2006). Norms can operate unidirectionally or polydirectionally. Weak social interactions, such as the use of social proof or social heuristics to make decisions, are unidirectional or one-way empirical, in the sense that the actor bases her decisions on the actions of others but does not believe that this will inform others' attitudes toward her. By contrast, strongly social interactions like social norms (including coordination and cooperation norms) are polydirectional, in the sense that one follows a rule because one believes that others do (“weakly social”) *and* because one believes that others believe one should follow it. Finally, institutional and legal norms are formal, legitimate, explicit, and enforceable by coercion (Mackie et al. 2011). The impact of each of these levels on group-oriented contribution seems to be mediated by a distinct mechanism. This paper will focus on the impact of polydirectional social norms on prosocial action.

The predominant epidemiological approach to social norms gives some clues to how institutions may be shaping norm accretion and diffusion. Two types of social norms motivate behavior: injunctive norms (what most others approve of) and descriptive norms (what most others do) (Cialdini and Kallgren 1990). Injunctive norms motivate behavior by conditioning it with rewards and punishments (Cialdini and Trost 1998). Descriptive and injunctive norms can act antagonistically to one another, with impacts varying depending on whether the norm is salient (or primed) in a particular context. This priming can be accomplished by activating related, “nearby” norms, though the effect fades with distance. Action patterns can lead to norm accretion even in the absence of explicit communication (Cialdini 2007). Interestingly, minor norm violation (a single piece of litter) can motivate norm restoration by increasing the norm’s salience, but major norm violation (much litter) erodes compliance despite salience.<sup>3</sup> Situational signals can activate one norm over another

---

<sup>3</sup>Similarly, public service messaging can unwittingly normalize either desirable or undesirable conduct. Messaging can normalize either desirable or undesirable conduct. Messaging that employs descriptive normative information (“this norm is often violated”) tends to normalize undesirable conduct, while messaging using injunctive normative

in particular contexts (Cialdini and Trost 1998).

Epidemiological diffusion can be consciously abetted. So-called “norm entrepreneurs” arise quasi-randomly and agitate for norm change. Norm entrepreneurs change norms by strategically manipulating the logic of appropriateness to encompass new behavior (Finnemore and Sikkink 1998). Once a new norm emerges, increased salience can initiate a norm cascade leading the system to rapidly adopt the new norm. Over time, this new norm becomes internalized and is no longer acted on from pure calculation, coming to be seen as proper conduct (Finnemore and Sikkink 1998). At a broad level, theorists distinguish between non-internalized (“material”) norms that are complied with on the basis of their consequences, and internalized (“ideational”) norms that are complied with on the basis of appropriateness (Fearon and Wendt 2002).

## **2.1 Prosocial Activity and the Altruism Question**

We can define a prosocial action as action taken by an individual that benefits society as a whole. These include “helping, sharing, donating, co-operating and volunteering,” (Brief and Motowidlo 1986) as well as compliance with legitimate rules against self-dealing. Prosocial norms can be activated (and subjects can be induced to act against their interests) by mere promises and formal covenants, implying the existence of deeper evolutionary mechanisms for altruistic cooperation (Bicchieri 2006). While it may seem natural to think of increased prosocial cooperation as a good thing, such a conclusion is dependent on our definition of group welfare. Although the network structures of small towns are quite effective at motivating prosocial behavior, many people find such norm environments oppressive. More generally, there may be limits to the circumstances under which we want to cooperate with others (Axelrod 1980, pp.124-5). This paper leaves such concerns aside, tantalizing as they are, to focus on the causal mechanisms underlying the individual information (“norm compliance is really important”) is most likely to suppress it (Cialdini et al. 2006).

choice to contribute prosocially.

How might we explain variation in an individual's propensity to contribute to their society? It seems unlikely that this decision is made from first principles each time. Reliance on heuristics, habits and norms seems far more likely to motivate variation in behavior. Social expectations are constructed by beliefs about what "relevant others" will do, and also by our beliefs about what relevant others approve of (Cialdini 1998). A major factor proposed to explain this individual variation in contribution propensity is altruism. Altruism consists in one person's utility being positively affected by another's welfare (Axelrod 1980 p. 135). In experiments with human subjects, individuals who were relatively more aware of the consequences of their behavior for others and who had difficulty denying responsibility for their behavior tend to behave more altruistically (Cialdini and Trost 1998). The existence of genuine altruism is the subject of incandescent debate in the psychology literature, and it seems that altruistic acts are frequently undertaken for the motive of creating an obligation or for performative reasons. This perspective receives support from studies of within-group behavior. Our interactions within the reference group are characterized by communitarian, relatively more altruistic principles, while our interactions outside the reference group are based on exchange principles and distributive justice (Cialdini and Trost 1998).

Indeed, altruism can be fatal to the enforcement mechanisms required to sustain cooperation. Within a network structure, free riders are typically punished spontaneously, in an uncoordinated fashion. The more an individual negatively deviates from the average contribution of group members, the heavier the punishment becomes. This punishment does not benefit the punisher directly, and in fact punishment is usually costly to the punisher. Willingness to punish constitutes a credible threat for potential free riders, and causes a large increase in cooperation levels (Fehr and Gächter 2000). In an evolutionary study of norms, Axelrod (1986) found that cooperation could only be sustained by the development of "metanorms" requiring third-party punishment of those

who refuse to punish defectors. Without the metanorm, declines in boldness reduced the salience of norm violation, and this reduced salience removed the direct incentive for cooperators to punish defectors (because they constituted a relatively small proportion of the population). Once vengeance became rare, boldness returned (with a vengeance?) and the norm completely collapsed. Without the metanorm, Axelrod found that this collapse was a stable outcome (Axelrod 1986). Most conceptions of altruists take them to oppose punishment of free-riders (e.g. Snidal 1991). I am not so sure. It seems possible to combine generalized goodwill with a desire not to be taken advantage of. In the model I develop *infra*, I assume that agents with high altruism scores are nevertheless willing to exercise social sanction.

## 2.2 Networks, Markets and Hierarchies

The central role of information sharing and reputation management suggests that norm monitoring and enforcement are maintained by a network structure. Network forms of economic and social organization are typified by reciprocal patterns of communication and exchange. These network forms contrast with market and hierarchical governance structures, providing a third major form of social organization (Powell 1990). Different organizational forms will typically coexist under normal conditions (Jung and Lake 2011). Networks are more social and less guided by formal authority, featuring significant interdependence among participants. Agents are motivated by norms of reciprocity and reputational concerns. In essence, “the parties to a network agree to forego the right to pursue their own interests *at the expense of others*” (Powell 1990, p.302, italics mine). The network form of organization is characterized by “voluntary, reciprocal, and horizontal patterns of communication and exchange” (Keck and Sikkink 1998, p.8). Networks might best be characterized as mechanisms for information acquisition as a form of indirect reciprocity (Jung and Lake 2011).

Networks are particularly good at information transmission, and research indicates that infor-

mation passed through networks is “thicker” than information obtained via the market and “freer” than information communicated in a hierarchy. (Powell 1990, citing Kaneko and Imai 1987). By establishing patterns of repeat interaction, networks restrict entry by outsiders. These networks will arise most easily in homogeneous settings with widespread trust where rapidly sharing and incorporating new information is essential, and they are typically dependent on the political and legal infrastructure provided by a hierarchy such as a state (Powell 1990). Networks incentivize learning as well as information dissemination, causing agents to rapidly react to new information and ideas. Agent-based modeling indicates that as the size of the network grows larger relative to the size of the population, the individual utility derived from the network declines because it is unable to provide new information to participants. Similarly, as the population grows larger agents are less likely to select the network form (Jung and Lake 2011). This may be because, for networks of a given size, larger populations reduce the utility of intra-network information transmission. As a population becomes “nastier,” relatively nicer agents are the first to forsake the network in favor of a hierarchy, and nastier types join the hierarchy last.<sup>4</sup> Hierarchy, and particularly autocracy, may emerge due to the cooperation it facilitates even when levels of distrust within the population are very high (Jung and Lake 2011).<sup>5</sup> Selective affinity (the ability to select transaction partners) produces robust networks that persist indefinitely, explaining the persistence of networks in homogeneous settings. Observed networks that endure for long periods are more likely to be founded on gains that arise from partner selection than from information. However, selective affinity can also cause agents to perceive the world as more threatening than it really is, driving them to take

---

<sup>4</sup>“... we see individuals insulating themselves from opportunism by turning to the centralized, legal enforcement mechanisms of the state” (Jung and Lake 2011, p.985). The authors cite increased litigiousness in American society as possible evidence of this broad transition from network to hierarchical organization.

<sup>5</sup>Social structure is necessary to elicit cooperation, and status hierarchies have been found to do this effectively (Axelrod 1980 p.145)

refuge in hierarchy. Recent scholarship has explained findings of diminished social capital on the basis of abandonment of the network form in favor of hierarchy (Jung and Lake 2011).

Norms exert the greatest influence when conditions are uncertain, when the source of the norm is similar to us, or when we are particularly concerned about establishing or maintaining a relationship with the source (Cialdini and Trost 1998). This finding indicates that network structures are the most likely channel for the mediation of social norms, and it gives us a reason to reconsider the role of political institutions in norm mediation. State institutions are necessarily hierarchical, and the information transmitted within a hierarchy is not of the same caliber as information relayed through a network (Powell 1990). However, participation in a network is always costly, as participation involves time spent providing information, developing social capital, and sending costly signals of commitment. “The alternatives to social networks are not only an anomic market of declining cooperation, but also a civil society in which mutually beneficial cooperation is enforced by the threat of centralized punishment” (Jung and Lake 2011). We might expect networks to operate synergistically with hierarchies to organize areas of interaction that are not suited to command-and-control methods. While it is certainly possible to compel prosocial contribution (in the form of taxes, for instance), this approach is crude and does not deploy the power of group-oriented social norms and their associated monitoring mechanisms to achieve fine-grained compliance. Because the scope of norm-based networks appears to be more finely attuned than the broad power of the state, abandonment of network forms of social organization in favor of the safety provided by hierarchy ought to generate a corresponding deadweight loss to social utility.

### **2.3 The Role of Institutions**

As we have just seen, the network structure is particularly vulnerable to heterogeneous populations. However, institutions can preserve the effectiveness of a monitoring system even in the context of

large populations (Milgrom et al. 1990). The system of specialist judges (“law merchants”) relied on to enforce commercial law prior to the rise of the state was precisely this type of institution, and it induced merchants to take personally costly actions that redounded to the benefit of the group. Specifically, judges induced merchants to behave honestly, sanction violators, inform themselves about the behavior of others, testify against violators, and pay damages assessed against them. By coordinating sanctioning activity, law merchants guaranteed a return on sanctioning behavior, in the form of the gains from exchange resulting from the persistence of cooperation (Milgrom et al. 1990). The authors instance other examples of institutions arising to enhance the effectiveness of reputation mechanisms through information transmission, citing the Hebrew *Mishipora* and the English “hue and cry” (used to identify cheaters), as well as the stocks and pillories of colonial-era New England. Medieval transnational commercial networks deployed precisely this type of reciprocity and information sharing to allocate resources in the absence of hierarchical, authoritative institutions (Greif 1989, 2006).

Once established, institutions like the law merchant and the Champagne fairs benefit from substantial network effects. Informal sanctions by community members can induce desirable behavior even in infrequent interaction, as long as information is reliably transmitted locally within the reference group (Kandori 1992). However, the permanence of these institutions is by no means guaranteed. We can view factors like population size and technological development as “quasi-parameters” which can be altered without disrupting equilibrium. Institutions can endogenously produce marginal shifts in the values of quasi-parameters, and these shifts can aggregate into sudden change that ultimately disrupts the institutional equilibrium (Greif and Laitin 2004).

## 2.4 The Problem of Relative Gains

The relative gains problem occurs when agents become concerned that their counterparties might achieve relatively greater gains from collaboration and, “thus strengthened, become more domineering friends in the present or possibly more formidable foes in the future” (Grieco et al. 1993, Waltz 1976). One party may use its disproportionate gain to implement a policy to damage the other. Such fears can preclude the realization even of particularly large gains as long as each party fears the uses the other will find for its augmented capabilities (Waltz 1979). Perceptions of cooperation as zero-sum or positive-sum can dramatically affect willingness to engage in prosocial activity because of the specter of relative gains (Powell 1991). In fact, the problem of relative gains has been shown to be a special case of the tradeoff between long- and short-term absolute gains (Snidal 1991, Powell 1991).

Concerns about relative gains will inhibit cooperation when the utility of force is high, but not when the utility of force is low. An augmented threat must be a plausible use of the relative gains for such gains to swamp cooperation (Powell 1991). Hierarchy thus seems likely to diminish relative gains concerns and promote cooperation. In addition, the relative gains problem tends to diminish as the number of relevant actors increases (Snidal 1991). If the relative gains problem arises, it will be necessary to incorporate the positive utility of other agents as a negative value in each agent’s utility function. However, cooperation under conditions of a state monopoly of violence should render relative gains concerns moot, so these considerations seem unlikely to apply to individual prosocial contribution. One concern remains. When individuals organize themselves into groups, the independence of these actors can be lost as a new level of organization emerges, which behaves as an actor in its own right (Axelrod 1997). This aggregation process can reintroduce the relative gains debate at a higher level of complexity.<sup>6</sup>

---

<sup>6</sup>Indeed, this prosperity-sapping struggle among a small number of interest groups is precisely what Madison, in

### 3 Theory

I propose that variations in individual propensity to make prosocial contributions can be explained by three independent variables: group-oriented social norms, monitoring mechanisms and altruism propensity. Each of these factors is individually insufficient to elicit group-oriented behavior, but they combine to create the perception that individual interests can be advanced most effectively through contributions to group objectives. I further speculate that effective monitoring is inversely related to 1) group size, 2) group heterogeneity, and 3) velocity of circulation. If this is correct, we can expect small, homogeneous societies whose cultures feature group-oriented social norms to exhibit more cooperation than large, heterogeneous societies that lack these norms. We would also expect little cooperation small homogeneous societies without group-oriented social norms and in large heterogeneous societies that did have such norms.

There seems to be no reason to follow a social norm if individual actions are not monitored by relevant others. In circumstances where no monitoring is possible, social norms (as distinct from, say, moral norms) lose their efficacy. A group-oriented social norm is therefore a necessary but not a sufficient condition for group-oriented behavior. In a situation of perfect monitoring, we might expect agents to expend significant effort on anticipatory compliance, whereas in a situation of no monitoring the incentive to expend costly effort complying with the expectations of others would evaporate. In a situation of incomplete or partial monitoring, we could expect patchy or situational compliance.

Individuals will contribute if the utility of contribution,  $u(c)$ , outweighs the utility of noncontribution. The utility of noncontribution is the opportunity cost of the resources (time, money) contributed, or  $c$ . The utility of contribution is a function of group-oriented social norms and monitoring mechanisms. Very roughly, I suggest the following structure. I understand monitoring

---

Federalist 51, argued that his new constitution would prevent.

Table 1: Contribution to group objectives requires both norms and monitoring

Reference Group Has:	Monitoring Mechanisms	No Monitoring Mechanisms
Group-Oriented Social Norms	Contribution probable	Contribution improbable
No Group-Oriented Social Norms	Contribution improbable	Contribution improbable

mechanisms to be institutions for the verification and dissemination of information regarding norm compliance. Even in situations of low information transmission, specialist institutions can induce individual agents to comply with norms at significant personal costs (Milgrom et al. 1990).<sup>7</sup> In addition, the emergence of rudimentary monitoring mechanisms may exhibit network effects and create an incentive in being monitored. As monitoring mechanisms increase in effectiveness the salience of social norms in agent decision-making increases. Monitoring mechanisms thus seem to be mediating the individual relevance of group-oriented norms.

### 3.1 Similarities With Revolution

A ready analogy for the process I am describing is the latent “revolutionary threshold” within a population, as described in Kuran 1991. Kuran seeks to understand the conditions under which individuals will display antagonism toward the regime under which they live. He describes a population with heterogeneous levels of exogenous anger towards the regime, and introduces a threshold at which individuals will engage in revolutionary activity. Kuran demonstrates that infinitesimal changes to the revolutionary threshold of a single individual can trigger a cascade of protest<sup>8</sup>(Kuran 1991). In Kuran’s work, we see broadly the same interaction of norms and monitoring with an endogenous emotion that I am describing in the case of prosocial contribution.

---

<sup>7</sup>However, the utility of such networks is attenuated both by time and by large populations (Jung and Lake 2011).

<sup>8</sup>“What endows intrinsically insignificant events with potentially explosive power in the context of political change is that public preferences are interdependent” (Kuran 1991 p.39).

Consider a society whose members are indexed by  $i$ . We can distinguish between an individual's private preference (fixed and exogenous) and her public preference (a variable under her control). Divergence in these preference is known as "preference falsification" (Kuran 1991). The individual decision to act will be based on a tradeoff between the payoffs for norm compliance and the costs of that compliance. In the present model, Kuran's psychological cost of preference falsification is analogous to the psychological benefit of compliance ( $x$  - see below). We can easily imagine a model analogous to Kuran's where the role of the revolutionary threshold is played by a contribution threshold indexing the number of contributors and the cost of contribution. These thresholds function as quasi-parameters, invisibly shifting without immediately disrupting the equilibrium (Greif and Laitin 2004). Kuran supports this interpretation, writing that "[a] massive change in private preferences may leave the incumbent equilibrium undisturbed, only to be followed by a tiny change that destroys the status quo, setting off a bandwagon that will culminate in a very different equilibrium" (Kuran 1991 p.46).

Historically, the conditions eliciting prosocial contribution - ethnically homogeneous groups with rapid information transmission - seem to mirror those that foster insurgency (Larson and Lewis 2018). The coordination of antisocial grievances appears to be the mirror image of the coordination of prosocial contribution. In his study of resistance movements, Roger Peterson introduces a similar concept to Kuran's revolutionary threshold which he characterizes as a "spectrum" of roles which individuals occupy as they are radicalized towards revolution (Peterson 2001). The analogy is limited - states are hierarchies, not networks - but the concept of latent action tendency proves extremely useful in elucidating individual reasons for prosocial contribution. In Section 4 *infra*, I model prosocial contribution using Kuran's approach.

## 4 Model

### 4.1 Individual Utility Functions With Exogenous Norms

I argue that the individual utility of contribution is a function of the cost of contribution, the reputational impact of (non)compliance with prevailing norms, and any individual satisfaction produced by the act of contribution itself. This utility function is given by the following equation:

$$u_i(c) = (1 - c) + (2c - 1)(\gamma * \mu) + xc \tag{1}$$

where  $c = \{0, 1\}$ ,  $\gamma = \{0..1\}$ ,  $\mu = \{0..1\}$ , and  $x = \{0..1\}$ <sup>9</sup>. The terms in the equation may be restated in the following way:

**Definition 1.** Utility of Contribution = Resource Endowment + Group Satisfaction + Self Satisfaction

The utility of prosocial contribution is given by the preceding three terms. Individuals are assumed to have an endowment of 1 resource, which they may retain ( $c = 0$ ) or expend on prosocial contribution, which carries a cost of 1 ( $c = 1$ ). This resource endowment is represented by  $(1 - c)$ . Next, the group satisfaction term interacts the strength of group-oriented social norms ( $\gamma$ ) with the likelihood that individual actions will be monitored by the reference group ( $\mu$ ). This interaction is multiplicative, because a norm's existence is only salient in individual behavior to the extent that noncompliance will be observed by relevant others. This effect is captured by  $(\gamma * \mu)$ . We wish the magnitude of this interaction to redound to the credit of a norm-complier and to the discredit of a norm-evader, which is accomplished by the term  $(2c - 1)$ . Finally, individuals' enjoyment of prosocial activity varies, and this variation is represented by the altruism term ( $x$ ). While prosocial activity carries a cost of 1, altruists may receive some portion of that cost back in the

---

<sup>9</sup>The upper bound of  $x$  is an asymptote, as perfect altruism seems impossible (though I am only asserting this here).

form of benefits. Note that these gains are only realized when the choice is made to engage in prosocial activity. This feature is incorporated by the term  $(xc)$ .

After simplifying terms, we observe that when  $c = 1$  (the contribution case), individual utility is given by  $x + (\gamma * \mu)$ . When  $c = 0$  (the noncontribution case), utility is given by  $1 - (\gamma * \mu)$ . These functions capture the essential features of the contribution decision. In the case of contribution, an individual receives any idiosyncratic benefit from the act of contribution itself, plus the benefit of being a known contributor, but only to the extent that the contribution will be monitored and transmitted to the reference group. In the case of noncontribution, an individual retains the initial resource endowment (time, money, etc.), but bears the costs of being known as a noncontributor to the extent that the contribution will be monitored and transmitted.

## 4.2 Three-Player Game With Endogenous Norms

Merely defining the utility function has not allowed us to make any predictions regarding individual behavior. The presence of two unknowns in the equation ( $\gamma$  and  $\mu$ ) would force us down a comparative statics path. However, if we endogenize group-oriented social norms, we may be able to gain some purchase on the decisions induced by varying levels of contribution monitoring. While individuals may ordinarily feel like price-takers with respect to the social norms prevalent within their reference group (Mackie 2011), such norms are genuinely endogenous products of repeated interaction. In the long run, norms are maintained and weakened by individual acts of compliance and enforcement.

Imagine a community made up of three individuals  $(p1, p2, p3)$  with heterogeneous levels of altruism  $(x)$ . Let us assume that group-oriented social norms  $(\gamma)$  are a function of these individuals' contribution decisions. In cases where no-one contributes, no norm exists, and  $\gamma = 0$ . In cases where all contribute, the norm is perfectly strong, and  $\gamma = 1$ . Where two out of three contribute,

Table 2: A three-player game with endogenous group-oriented social norms.

$p_1 \implies c$ ( $c = 1$ )			$p_1 \implies \neg c$ ( $c = 0$ )			
	$p_2 \setminus p_3$	$c$	$\neg c$	$p_2 \setminus p_3$	$c$	$\neg c$
$c$	$\mu + 0.1$	$0.8\mu + 0.1$		$c$	$1 - 0.8\mu$	$1 - 0.2\mu$
	$\mu + 0.5$	$0.8\mu + 0.5$			$0.8\mu + 0.5$	$0.2\mu + 0.5$
	$\mu + 0.9$	$1 - 0.8\mu$			$0.8\mu + 0.9$	$1 - 0.2\mu$
$\neg c$	$0.8\mu + 0.1$	$0.2\mu + 0.1$	$\neg c$	$1 - 0.2\mu$	$1$	
	$1 - 0.8\mu$	$1 - 0.2\mu$		$1 - 0.2\mu$	$1$	
	$0.8\mu + 0.9$	$1 - 0.2\mu$		$0.2\mu + 0.9$	$1$	

we can assume the existence of a strong norm, arbitrarily assigning a value of  $\gamma = 0.8$ . Where only one contributes, we can assume only a weak norm, arbitrarily assigning a value of  $\gamma = 0.2$ .

While these particular parameter values are arbitrary, the network effects underlying norm operation imply a concave function with a maximal rate of change at 50% of the population (Bicchieri 2006). Norms, it seems, tend to not matter until enough relevant others suddenly expect compliance with them, at which point salience rapidly increases, benefiting from network effects. My goal is to effectively model the individual reaction to the choice environment, and by endogenizing the group-oriented social norm in this way, the norm remains outside the control or manipulation of individual agents because the norm environment each player faces is always dictated by the behavior of the other two players. Altruism levels in this population are evenly distributed about  $x = 0.5$ , and I have arbitrarily assigned p1 an altruism level of  $x = 0.1$  (low altruism), p2 a level of  $x = 0.5$  (moderate altruism), and p3 a level of  $x = 0.9$  (high altruism).

Table 3: Monitoring thresholds for prosocial contribution

	$p1$	$p2$	$p3$
0 contribute	$\mu > 4.5$	$\mu > 2.5$	$\mu > .5$
1 contributes	$\mu > .9$	$\mu > .5$	$\mu > .1$
2 contribute	$\mu > .5$	$\mu > .277$	$\mu > .055$

### 4.3 Discussion

We see at once that there are no values of  $\mu$  on our scale (0..1) that can induce  $p1$  and  $p2$  to contribute in the absence of group-oriented social norms, and that  $p1$  (the low-altruism player) will require enormous monitoring efforts ( $\mu = .9$ ) to induce contribution if norms are only modestly high (only one other player contributes). Increased altruism dramatically increases the likelihood of contribution even at low monitoring levels. The high-altruism player ( $p3$ ) will require virtually no monitoring ( $\mu = .055$ ) to induce a contribution in a high-norm world, and even in a setting of modest norms, the monitoring required is scarcely higher ( $\mu = .1$ ).

What will be the likely path of play? At values of  $\mu < .5$ , contribution is not rational for any player. At  $\mu = .5$ ,  $p3$  will be indifferent between solo contribution and noncontribution. If  $p3$  nevertheless contributes, we discover that  $p2$  is also indifferent between contribution under weak norms and noncontribution. However, with the further assumption that  $p2$  contributes, we find that  $p1$  favors contribution in a strong-norm environment. At values of  $\mu > .5$ , the system shifts into a cooperative state. Contribution becomes a dominant strategy for  $p3$ , and once we eliminate the possibility of  $p3$  not contributing, contribution becomes a dominant strategy for  $p2$ . As we have seen,  $p1$  is already eager to contribute when both others are.

The model indicates that norm formation may be parasitic on altruism. Without some kernel of the population who feel that the contribution in question is worthwhile for its own sake, a

norm regarding the desirability of that contribution seems unlikely to arise. With the present model parameters, it is literally impossible to motivate pro-social contribution without involving the high-altruism player ( $p3$ ). This presents us with an important scope condition on the type of contributions likely to motivate norm formation - they must be enjoyable for their own sake to some non-trivial portion of the population. In addition, monitoring levels must be non-trivial for contribution utility to rise above 1 for any player. At  $\mu < .1$ , the highest utility for any player is found in that player's zero-norm non-contribution cell. At  $.1 < \mu < .5$ , utilities higher than 1 are available, but only off the path of play.

We may also notice that net social utility actually declines as contribution begins. At values of  $\mu < .5$ ,  $\sum_{i=3}^1 u_i(c) = 3$  Where  $.5 < \mu < .75$ ,  $u_i(c) < 3$ . It is not until we reach values of  $\mu$  greater than .75 that the players receive more collective utility than they derived from noncontribution. When  $p3$  moves to a higher utility at  $\mu = .5$ , this move actually reduces the immediate welfare of  $p2$  and  $p3$ . The imposition of irksome, utility-draining norms might appear to be a significant downside of living among altruists. However, this model does not capture the extensive positive externalities generated by prosocial contribution. In terms of welfare, we can assume that once the system shifts into a contributory state, some fillip to the utility function is being provided as a result of widespread contribution. However, I have elected not to model this effect because I believe it to be inoperative in individual choice.<sup>10</sup>

#### 4.4 Contribution Thresholds

I argued above that the coordination of antisocial grievances appears to be the mirror image of the coordination of prosocial contribution. Consider a society whose members are indexed by  $i$ .

---

<sup>10</sup>While society does indeed benefit greatly from prosocial contribution, individuals cannot rationally expect to benefit personally as a result of their own inchoate contribution, except for the altruistic self-satisfaction which I have already modeled by  $x$ .

Each individual member must choose whether or not to make a public prosocial contribution, and depending on the configuration of norms, this decision will redound to their credit or discredit. This external observation, even if imperfect, creates a distance between an individual's private preference, which I am modeling as altruism ( $x$ ), and her public preference, which incorporates the feelings of others with respect to norm compliance.

Private Preference:

$$u_i(c) = (1 - c) + xc \tag{2}$$

Public Preference:

$$u_i(c) = (1 - c) + (2c - 1)(\gamma * \mu) + xc \tag{3}$$

Let  $\gamma$  represent the existence of a prosocial norm. Initially, it is near zero, implying that individuals will not contribute unless the altruistic benefit ( $x$ ) to be gained outweighs the cost of doing so ( $c$ ). As more people participate in prosocial activity, however, the group-oriented social norm ( $\gamma$ ) grows stronger, from a strength of .1 if one person participates to a strength of .9 if 9 people participate. If monitoring mechanisms ( $\mu$ ) increase in strength over time (analogous to Kuran's increasing public opposition to the regime), then at a certain point the individual calculation will flip and  $u_i(c)$  will be greater than  $u_i(-c)$ . This will occur when  $x + (\gamma * \mu) > 1 - (\gamma * \mu)$ . By analogy to Kuran, we can call this switching point a "contribution threshold".

Now imagine a ten-person society featuring the following threshold sequence:

$$A = (0, .2, .2, .3, .4, .5, .6, .7, .8, .9, 1) \tag{4}$$

Person 1 ( $x = 0$ ) will never undertake a prosocial action in the absence of norms and monitoring, and Person 10 ( $x = 1$ ) undertakes frequent prosocial action even without norms or monitoring. The remaining people's preferences are sensitive to the group satisfaction term  $((2c - 1)(\gamma * \mu))$ . If Person 1 contributes,  $\gamma$  takes on a value of .1, which is insufficient to induce Person 2 to contribute

without very high levels of monitoring. However, consider this new threshold sequence:

$$A = (0, .1, .2, .3, .4, .5, .6, .7, .8, .9, 1) \tag{5}$$

In this sequence, we will observe a cascade of contribution because each player's contribution threshold is just reached by the actions of players with lower thresholds. We saw a similar dynamic in the 3-player model. While this analogy is imperfect, the “now out of never” dynamics that Kuran theorized seem to be applicable in the context of prosocial contribution.

#### 4.5 Policy Implications

Axelrod famously advised policymakers seeking to promote cooperation to enlarge the shadow of the future, either by making interactions more durable or more frequent (Axelrod 1980 pp.129-130). He also argued that policymakers should encourage altruism. After reviewing the model above, we can see the reasons for this recommendation. If norm formation is genuinely parasitic on the existence of some non-trivial population of altruists, then the patient cultivation of that population would seem to be a clear objective for policymakers. An additional piece of low-hanging fruit might be to make the fruits of prosocial contribution more salient in individual utility calculations. Another implication of the model is that at low levels of monitoring, increases in resources expended on monitoring may fail to induce any changes. Policymakers should not lose heart, as the model indicates that change, though very difficult to initiate, takes effect very quickly once the contribution threshold of the most altruistic members is reached. Despite the initial inertia, as monitoring levels are increased the eventual ensuing norm cascade will induce compliance across the population, provided that there exists a population of altruists who find the activity in question worthwhile for its own sake. This last point also implies that policymakers should only attempt to elicit varieties of contribution that at least some subset of the population finds intrinsically worthwhile.

## 5 Conclusion

This paper examined the variation in an individual's propensity to contribute to their society by engaging in group-oriented actions. I argued that the individual utility of contribution is a function of the cost of contribution, the reputational impact of (non)compliance with prevailing norms, and any individual satisfaction produced by the act of contribution itself. In addition, I argued that social sanctions are themselves a function of the magnitude of group-oriented social norms and the extent of monitoring mechanisms. I then constructed and tested a formal model incorporating these parameters. Agents in the model exhibited behavior that resembles behavioral predictions made by the concepts of rapid norm shifts (Bicchieri 2006) and "norm cascades" (Keck and Sikkink 1998). Counterintuitive conclusions became apparent, particularly the dependence of group-oriented norms on the presence of altruistic actors. Finally, I used the model developed in Kuran 1991 to introduce the concept of "contribution thresholds," and I explored deep structural resemblances between the concepts of revolution and prosocial contribution.

The work presented here, and the present model in particular, are merely a first pass at this topic. I am interested in extending the model by increasing the number of players, iterating play, and adding a memory component, among other features. I wish to find a way of modeling the distinction between descriptive and injunctive norms - early attempts to incorporate them here increased complexity without yielding insight. I would also like to explore the consequences of fuzzy or incomplete information, and the role of uncertainty in motivating network exit. I have identified three case studies of this mechanism in action, and I plan to develop and deepen those case studies using process tracing techniques. I aim to develop a theoretical normative treatment of the concept of prosocial contribution. I would like to explore the differences between altruism-based cooperation and norm-coerced cooperation - the present model treats them as equivalent. Finally, I plan to conduct empirical tests of several of the modeling assumptions, particularly the endogenous

altruism levels ( $x$ ). Unraveling the mechanisms behind prosocial contribution seems important and worth pursuing. Much attention is justly given to the processes by which things can go wrong. It seems only fair to spare some consideration for how things might go right.

## 6 References

- Axelrod, Robert. 1986. "An Evolutionary Approach to Norms." *The American Political Science Review*, Vol. 80, No. 4 (December), pp. 1095-1111. doi: 10.1017/S0003055400185016.
- Axelrod, Robert. 1980. *The Evolution of Cooperation*. New York: Basic Books.
- Axelrod, Robert. 1997. *The Complexity of Cooperation*. Princeton: Princeton University Press.
- Baldwin, David (ed.). 1993. *Neorealism and Neoliberalism: The Contemporary Debate*. New York: Columbia University Press.
- Batson, Daniel. 2008. "Prosocial Motivation: Is it Ever Truly Altruistic?" in *Advances in Experimental Social Psychology*, Vol. 20, pp. 65-122. doi: 10.1016/S0065-2601(08)60412-8.
- Bicchieri, Christina. 2006. *The Grammar of Society*. Cambridge: Cambridge University Press.
- Brief, Arthur and Stephan Motowidlo. 1986. "Prosocial Organizational Behaviors." *The Academy of Management Review* Vol. 11, No. 4 (October), pp. 710-725. doi: 10.5465/amr.1986.4283909.
- Cialdini, Robert. 2007. "Descriptive Social Norms as Underappreciated Sources of Social Control." *Psychometrika*, Vol. 72 No. 2 (June), pp. 263-268. doi: 10.1007/s11336-006-1560-6.
- Cialdini, Robert, Linda Demaine, Brad Sagarin, Daniel Barrett, Kelton Rhoads, Patricia Winter. 2006. "Managing Social Norms for Persuasive Impact." *Journal of Social Influence*, 1 (1), pp. 3-15. doi: 10.1080/15534510500181459.
- Cialdini, Robert, Raymond Reno and Carl Kallgren. 1990. "A Focus Theory of Normative Conduct." *Journal of Personality and Social Psychology*, Vol. 58, No. 6, pp. 1015-1026. doi: 10.1037/0022-3514.58.6.1015.
- Cialdini, Robert and Melanie Trost. 1998. "Social Influence: Social Norms, Conformity and Compliance" in Gilbert, Daniel, Susan Fiske and Gardner Lindzey (eds.). 1998. *The Handbook of*

*Social Psychology Volume II*. New York: Oxford University Press, pp. 151-162.

Coleman, James. 1990. *Foundations of Social Theory*. New York: Harper and Row.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper and Row.

Fearon, James and Alexander Wendt. 2002. "Rationalism v. Constructivism: A Skeptical View" in Walter Carlsnaes, Thomas Risse, and Beth Simmons (eds.), *Handbook of International Relations, 1st edition*, Thousand Oaks, CA: Sage Publications, 2002, pp.52-72.

Fehr, Ernst and Simon Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments." *The American Economic Review*, Vol. 90, No. 4 (September), pp. 980-994. doi: 10.1257/aer.90.4.980.

Finnemore, Martha and Kathryn Sikkink. 1998. "International Norm Dynamics and Political Change." *International Organization*, Vol. 52, No. 4, *International Organization at Fifty: Exploration and Contestation in the Study of World Politics (Autumn)*, pp. 887-917. doi: 002081898550789.

Forman-Barzilai, Fonna. 2009. *Adam Smith and the Circles of Sympathy*. Cambridge: Cambridge University Press.

Greif, Avner. 2006. *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*. New York: Cambridge University Press.

Greif, Avner. 1989. "Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders." *The Journal of Economic History*, Vol. 49, No. 4 (December), pp. 857-882. doi: 10.1017/S0022050700009475.

Greif, Avner and David D. Laitin. 2004. "A Theory of Endogenous Institutional Change." *American Political Science Review*, Vol. 98, No. 4 (November) pp. 633-652. doi: S0003055404041395.

Greif, Avner, Paul Milgrom and Barry R. Weingast. 1994. "Coordination, Commitment, and Enforcement: The Case of the Merchant Guild." *The Journal of Political Economy*, Vol. 102, No.

4 (August), pp. 745-776. doi: 10.1086/261953.

Grieco, Joseph, Robert Powell and Duncan Snidal. 1993. "The Relative Gains Problem for International Cooperation." *The American Political Science Review*, Vol. 87, No. 3 (September), pp. [727]-743. doi: 10.2307/2938747.

Jung, Danielle F. and David A. Lake. 2011. "Markets, Hierarchies, and Networks: An Agent-Based Organizational Ecology." *American Journal of Political Science*, Vol. 55, No. 4 (October), pp. 972-990. doi: 10.1111/j.1540-5907.2011.00536.x.

Kandori, Michihiro. 1992. "Social Norms and Community Enforcement." *The Review of Economic Studies*, Vol. 59, No. 1 (January), pp. 63-80. doi: 10.2307/2297925.

Keck, Margaret and Kathryn Sikkink. 1998. *Activists Beyond Borders: Advocacy Networks in International Politics*. Ithaca: Cornell University Press.

Kuran, Timur. 1991. "Now out of Never: The Element of Surprise in the East European Revolution of 1989." *World Politics*, Vol. 44 No. 1 (October), pp. 7-48. doi: 10.2307/2010422.

Larson, Jennifer and Janet Lewis. 2018. "Rumors, Kinship Networks and Rebel Group Formation." *International Organization*, Vol. 72, No. 4 (Autumn), pp. 871-903. doi: 10.1017/S0020818318000243.

Mackie, Gerald, Francesca Moneti, Holly Shakya and Elaine Denny. 2015. "What Are Social Norms? How Are They Measured?" UNICEF / University of California Center on Global Justice.

Milgrom, Paul, Douglass North and Barry Weingast. 1990. "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges and the Champagne Fairs." *Economics and Politics*, Vol. 2 No. 1 (March), pp. 1-23. doi: 10.1111/j.1468-0343.1990.tb00020.x.

Peterson, Roger. 2001. *Resistance and Rebellion: Lessons from Eastern Europe*. Cambridge: Cambridge University Press.

- Powell, Robert. 1991. "Absolute and Relative Gains in International Relations Theory." *The American Political Science Review*, Vol. 85, No. 4 (December) pp. 1303-1320.
- Powell, Walter W. 1990. "Neither Market Nor Hierarchy: Network Forms of Organization." *Research in Organizational Behavior*, Vol. 12, pp. 295-336.
- Riaño, Yvonne. 2011. "Addressing Urban Fear and Violence in Bogotá through the Culture of Citizenship: Scope and Challenges of a Unique Approach." in *EthniCities: Metropolitan Cultures and Ethnic Identities in the Americas*. Butler, Gurr and Kaltmeier (eds.). Tempe: Bilingual Review Press, p.209-226.
- Sperber, Dan. 1996. *Explaining Culture: A Naturalistic Approach*. Cambridge: Blackwell.
- Snidal, Duncan. 1991. "Relative Gains and the Pattern of International Cooperation." *The American Political Science Review*, Vol. 85, No. 3 (September) pp. 701-726. doi: 10.2307/1963847.
- Waltz, Kenneth. 1979. *Theory of International Politics*. Reading: Addison-Wesley.