Collective Action, Rival Incentives, and the Emergence of Antisocial Norms

James A. Kitts
University of Washington

Centralized sanctions (selective incentives) and informal norms have been advanced as distinct solutions to collective action problems. This article investigates their interaction, modeling the emergence of norms in the presence of incentives to contribute to collective goods. Computational experiments show how collective action depends on a three-way interaction among the value of incentives, the rivalness of incentives (ranging from independence to zero-sum competition), and group cohesiveness (effectiveness of peer influence). This investigation shows a broad range of conditions in which social norms promote the collective good and thus peer influence complements a centralized regime of selective incentives. It also shows conditions in which the two systems clash because incentives lead to antisocial norms that discourage contributions to collective goods. In these conditions, social scientists must reconsider the widely predicted relationships of collective action to selective incentives, group cohesiveness, and second-order free riding.

The crux of the theoretical puzzle of collective action is the free-rider problem. Where a group of self-interested actors may produce a shared good that will be available to everyone in the group, rational individuals will seek to free ride on others' efforts, and thus the group should fail to produce the good. More generally, any group that experiences externalities (such as collective rewards or punishments) as a consequence of individuals' behavior is vulnerable to selfish opportunism and, thus, has a "regulatory interest" (Heckathorn 1988) in controlling that behavior. The fact that many people do devote substantial effort and other resources to social movements, charitable organizations, religious and fraternal orders, and other voluntary associations, with little or no direct compensation for doing so, suggests that at least some real-world groups have found solutions to this fundamental problem of collective action.

Rational choice scholars have proposed two families of solutions to account for observed collective action in interest groups. In a top-down version, a central regime assigns sanctions (Coleman 1990; Hechter 1987) to encourage individuals to serve the collective interest. I customarily refer to this approach as *formal control* and refer to the sanctions as "selective incentives" (Olson 1965). In a bottom-up version, *informal norms* emerge from and percolate through social interaction (Homans 1961), as peers pressure each other to forego their selfish interests for the collective good. Extensive research on formal control and informal norms has examined the efficacy of enforcement while leaving the content of those norms exogenous.

This article engages the prevailing scholarship on social norms in sociology, economics, political science, and law, which builds on this

Direct correspondence to James A. Kitts, Department of Sociology, University of Washington, 202 Savery Hall, Box 353340, Seattle, WA 98195 (kitts@u.washington.edu). Earlier versions were presented at the 1999 meeting of the Eastern Sociological Society and the 2003 meeting of the American Sociological Association. Research supported by a National Science Foundation grant, SES0433086. For valuable feedback and suggestions, the author thanks William Litsch, Christian Steglich, Xuesong Geng, Michael Macy, Michael Hechter, Karl-Dieter Opp, Victor Nee, Joseph Whitmeyer, Yen-Sheng Chiang, Edgar Kiser, participants in the University of Washington seminar in Institutional Analysis, and four *ASR* reviewers.

rational choice framework (Coleman 1990; Ellickson 2001:36). The scholars in this area argue that norms emerge in groups to solve collective action problems. Norms are generally defined as *prosocial.* That is, norms mandate individually costly behavior that is beneficial to others (Hechter 1987; Kanazawa 1997) or prohibit individually gratifying behavior that is harmful to others.[1]

This assumption that emergent norms are prosocial has generated strong predictions for collective action outcomes. First, scholars predict that groups in which members are better able to regulate each other's actions ("cohesive" groups) should be more successful at promoting collective goods (Heckathorn 1988; Homans 1961; Horne 2001a). Second, research on the "second-order free-rider problem" (Oliver 1980; Yamagishi 1986) investigates a parallel collective action problem embodied in norm enforcement itself. Each actor may prefer to free ride in enforcement efforts and thus allow others to bear the burden of pressuring peers. The prediction that second-order free riding will be an obstacle to collective action again follows from the assumption of prosocial norms. Studies of informal control, including cohesiveness and second-order free riding, have provided valuable insights into collective action outcomes given prosocial norms, but do not explain the origins of the norms themselves.

Scholars who consider the emergence of informal norms (Axelrod 1986; Bendor and Swistak 2001) generally focus on the case in which no formal rule or centralized enforcement regime exists.[2] In this article, I investigate the interplay of both regimes, modeling the emergence of norms in an institutional context that includes a centralized selective incentive. Analysis of this model proves that selective incentives to work for the collective good may paradoxically lead rational group members to enforce *antisocial* norms that discourage contributions to collective goods. As a consequence of this novel conclusion, three widespread beliefs in the literatures on collective action and social norms become problematic under the logic of the rational choice framework that serves as their foundation. Specifically, three interventions widely believed to promote collective action—providing selective incentives to contribute to the collective good, increasing the effectiveness of peer influence, and preventing second-order free riding—all may *diminish* contributions to the collective good. Beyond showing that this paradoxical result is logically possible, my set of computational experiments specifies the conditions under which we should expect it to obtain according to the model. Most surprisingly, they show that these scope conditions for the emergence of antisocial norms correspond to conditions under which scholars generally predict the emergence of functional norms.

The stark difference in conclusions demonstrated in this investigation is attributable to a single innovation. I begin by identifying "rivalness" as a very general but rarely recognized property of selective incentives to participation in collective action. By *rivalness,* I mean a characteristic form of interdependence among recipients of an incentive such that each recipient diminishes the value of incentives to other recipients. I argue that this negative interdependence is typical of many goods described as incentives to participation in collective action groups, especially rewards of *kudos* (e.g., status, prestige, or esteem).

After clarifying my assumptions and usage of key terms, I present an elementary mathematical model of a task group, in which actors may "work" or "shirk" for the collective good as well as enforce norms by choosing to "promote" or "oppose" work among peers. Analysis of this model allows us to derive general propositions about actors' inclinations to contribute to the collective good (given social norms) and to enforce norms (given the choice to work or shirk). To make predictions for a group in which both processes operate simultaneously, I perform simulations, mapping the model's behavior as I manipulate parameters of the control regimes. These "computational experiments" (Hanneman, Collins, and Mordt 1995; Macy and Willer 2002) generate novel and nonobvi-

[1] These have been called "essential norms" (Coleman 1990:249) or "prisoner's dilemma norms" (Heckathorn 1988; Ullmann-Margalit 1977) to distinguish them from conventions that do not resolve a conflict of interests between individuals and others. This is now common usage for norms in rational choice theory.

[2] Such work arguably investigates the more fundamental process because a system of selective incentives is itself a collective good that must ultimately be provided by individuals.

ous hypotheses for empirical investigation, and also identify scope conditions for these hypotheses. In conclusion, I recapitulate and distinguish my mathematically proven propositions, my hypotheses derived from the computational experiments, and some suggestive conjectures based on my exploration of the model.

Informal Norms: Peer Pressure Emerging from Individuals' Regulatory Interests

There is no universally accepted definition of social norms, and whatever definition is used inevitably constrains the theoretical questions that can be asked. In my usage, norms are regulatory forces exerted by group members prohibiting or mandating behavior within the group. Norms consist of voluntary efforts by group members to regulate their peers' behavior. This definition allows that conflicting norms may exist simultaneously. We can measure their strength by the level of pressure exerted to influence behavior toward particular ends.[3] For convenience and clarity, I decompose norms into *valence* (the direction of pressure) and *strength* (the force exerted toward that end in the group).

Notably, I regard the content of the norm (the direction of social pressure) as a phenomenon to be explained and do not assume that norms are always functional for the groups in which they emerge. This differs from conventional usage, in which scholars define social norms as prosocial.[4] For example, scholars often define norms as "(second-order) public goods that are instrumental in providing (first-order) public goods" (Opp 2001:236) by enforcing costly behavior that benefits the group. Such a definition presupposes the link between social norms and prosocial behavior, with reference to the group in which the norm emerges. Similarly, recent reviews have regarded attitudinal consensus and even widespread adherence as defining properties of norms (Horne 2001b:5). I note that presupposing consensus and compliance undermines interesting questions about normative conflict and failure.

I grapple with the standard assumption that social norms reflect the collective interests of the groups in which the norms emerge. Hechter and Opp (2001:xvi) note, for example, that "the view that norms are created to prevent negative externalities, or to promote positive ones, is virtually canonical in the rational choice literature." Accordingly, "once researchers identify the externalities experienced by the group, norm content ought to be predictable" (Horne 2001b:10). A tacit assumption—that the *group* experiences externalities and that norms emerge at the *group* level to regulate behavior in favor of the *group*'s collective interests—conflicts with the methodological individualist foundation of rational choice theory.[5] Ellickson (2001) exemplifies the unegoistic nature of the assumption that norms are collectively functional:

> Each member of a social group, when acting in the role of a member of the audience, has a utilitarian bias—that is, a selfless preference for norm changes that satisfy the criterion of Kaldor-Hicks Efficiency. . . . As long as members of the group would gain in the aggregate, audience members would not object to a norm on the ground that it would disadvantage them individually. (p. 39)

Although most authors are not so explicit, the prevailing rational choice accounts propose that norms emerge because they are needed to reg-

[3] My defining norms as patterns of regulatory behavior (goal-oriented social pressure or peer sanctions) among members does not include rules that exist as abstract discursive objects without any social consequences. In my framework, norms that would never be enforced simply do not exist. This is appropriate given my interest in regulatory behavior, but would not be appropriate if I were interested in studying the dynamics of discursive rules (March, Schulz, and Zhao 2000)

[4] Rational choice scholars (e.g., Hechter and Borland 2001) do not extend this prosocial claim to *disjoint* norms (Coleman 1990:247), which are rules imposed by some exogenous external authority. I will not keep repeating that by norms I mean voluntary regulatory efforts by group members, reflecting their own regulatory interests. In Coleman's terminology, this entire article conventionally focuses on *conjoint* norms.

[5] It may be possible for a set of rational actors to behave as a corporate actor in developing and enforcing norms, but this implies a model of behavior that is entirely different from the atomized choices in collective action theory. If groups are able to act in their collective interest for second-order behavior (norm enforcement), we might ask why they cannot simply cooperate in the first order (production of collective goods) and thus save themselves the effort of enforcing norms.

ulate the selfish behavior of individuals. This is a collective–functionalist solution to an individualist problem. Scholars argue that this suspension of methodological individualism is appropriate in certain kinds of groups, such as groups with dense and multifaceted interaction (Ostrom 1990; Taylor 1982). For example, Ellickson (1991:167) observes informal norms in a variety of "close-knit" communities and posits that norms in such groups will be "welfare-maximizing" for the group. Although he uses a game theoretic model to motivate and frame his theory, Ellickson cannot derive his hypothesis of welfare-maximizing norms from the rational choice model, and instead induces it from observation of empirical norms, such as those among whalers and ranchers. Readers are left with an empirical claim that social norms tend to be functional in certain kinds of groups, but without a logically consistent explanation of how they got that way. Scholars have simply asserted that prosocial norms "somehow emerge" (Eggertsson 2001:81), while acknowledging that idiosyncratic dysfunctional norms also exist because of "historical accidents or for other reasons" (Voss 2001:117).

There are three reasons why we may question the assumption that norms are collectively functional. First, we may simply observe that many norms in the real world do not seem to be functional for the groups that enforce them. Indeed, rules prohibiting harmless pleasures or mandating gratuitous sacrifices have been endemic to cultures through all known history. I do not emphasize this reason because the vague scope conditions on previous theory allow authors to find functional consequences at some level of analysis or time scale for virtually any norm. Given the difficulty of defining a sampling frame for the population of norms, not to mention determining the relevant beneficiaries, empirical research cannot offer much leverage on the question of the relative preponderance of functional or dysfunctional norms. I take a stronger position: *Even if prosocial norms are indeed ubiquitous, a theory of norms should be able to explain how they got that way.* And if some norms do turn out to be dysfunctional, the theory should explain the difference.

Second, we may question the assumption of prosocial norms on theoretical grounds. Contemporary theory emphasizes the stability of welfare-maximizing norms, but offers no dynamic account for how groups invent and maintain these norms amid a variety of obstacles. Because of change in regulatory interests over time (e.g., shifting values, drifting environments, or changing composition of the group), current norms may be vestigial leftovers (Sherif 1966) that no longer serve current members. Even without such "normative inertia" (Ellickson 2001:56–57), new norms may reflect generative processes that distort the preferences of members. For example, perceptive (Bicchierri and Fukui 1999) or communicative (Kitts 2003; Kuran 1995) biases may misrepresent normative preferences in the group, leading members to advocate or enforce a norm against their own collective interest. To focus on a more fundamental problem, I rule out such sources of inertia or inefficiency and model systems with little or no friction, in which scholars conventionally agree that norms should be prosocial.

The third reason why we may question this assumption is methodological. The conclusion that norms should be collectively functional has been either induced from empirical observations of seemingly functional social systems or derived from rational choice assumptions using loose intuition. Efforts to examine this link formally have yielded much weaker conclusions. For example, work in evolutionary game theory (Allison 1992; Axelrod 1984, 1986; Bendor and Swistak 1997; Boyd et al. 2003) has suggested ways that prosocial norms may be favored by selection. However, the theorists in this field show that this by no means guarantees that emergent norms will promote Pareto-optimal outcomes (Bendor and Swistak 2001), in which no actor could benefit from change without a concomitant loss by another actor. In fact, formal models show that emergent norms and decentralized enforcement regimes can mandate collectively harmful behavior (Boyd and Richerson 1992, 2001; Hirshleifer and Rasmussen 1989) as well as cooperation.

Evolutionary game theory demonstrates that emergent informal norms need not always be prosocial, and empirical observation informs us that norms are not always prosocial. On the other hand, decades of research in several disciplines have used the assumption that norms reflect the group's collective interests. Rather than answer the question *whether* norms are prosocial, our most constructive efforts should

aim to show *when* this should be true. That is, we should account for the content of norms. I propose a small step toward this end.

I begin with the rational choice foundation that motivates the general theories of norms, and maintain this same methodological individualist perspective in accounting for both norms and the behaviors they regulate. Norms emerge not from a collective need, but from the decentralized interaction of egoists according to their own regulatory interests. Although it seems intuitive that the regulatory interests aggregated across a group of actors should correspond directly to the interests of the aggregate, I prove that they differ systematically under well-defined conditions. Even if a group exists in which members have homogeneous preferences, and even in the absence of error or transaction costs, individuals in the group may rationally advocate norms that would diminish the welfare of the entire group. I show that these antisocial norms may emerge because of an interaction of informal norms with centralized formal control.

Formal Control: Centralized Sanctions as "Selective Incentives"

In my usage, formal control is a prescriptive or proscriptive rule enforced by a central authority using sanctions. This enforcement agent corresponds to Coleman's "despot" (1990:337), which could be a benevolent leader, corporate actor, or third party. The key feature of formal control is the *centralized* distribution of incentives, consistent with notions of "organization enforcement" (Ellickson 1991:131), "centralized institutions" (Ingram and Clay 2000:534), and "centralized structures" of control (Bendor and Mookherjee 1987:136),[6] whereas informal norms consist of members' decentralized regulatory behavior.

Conventionally, this article considers only the important case in which formal rules promote the collective good. Of course, a powerful central authority may impose harmful rules on a group, but this is orthogonal to the questions investigated here. I am interested in the emergence of informal norms among group members in the presence of a selective incentive to cooperate (Olson 1965). At no point do I aim to explain the behavior of the central enforcement agent.

I do not give an anecdotal description of the selective incentive, except to note that collective action theory focuses overwhelmingly on voluntary associations and other groups that cannot compensate members for their efforts. Theorists such as Chong (1991) instead describe intangible sentiments or symbols (e.g., kudos awards of social esteem, reputation, or status) as incentives for members to participate in collective action. An award of kudos often is inexpensive to the enforcement agent, and yet may serve as a powerful motivator for members of certain kinds of groups. I do not aim to explain the value of the selective incentive, but instead take the incentive's value to recipients as an independent variable and examine its side effect on informal norms.

Interplay of Formal Rules and Informal Norms

The emergence of collectively beneficial norms seems an intuitive outcome for groups of rational actors. It also seems that collective action should improve when a prosocial selective incentive is combined with effective peer influence. Despite this intuition, I show there is no guarantee that norms will optimize or even promote the collective interest, and there is a particular problem when these two regimes are combined. Specifically, formal control may pervert the regulatory interests of members, leading them to develop norms that are harmful to the entire group.

Some previous work has examined the interplay of formal control and informal norms, and also has offered predictions for the appearance of norms that challenge or subvert formal rules ("oppositional norms"). Most of this work has adhered to the conventional assumption that norms reflect the collective interest, and has offered predictions for organizational performance given variation in formal rules. For example, Nee (1998) proposes a condition for "close

[6] Scholars make substantive distinctions between *public* and *private* institutions (e.g., Ingram and Clay 2000), or between control by organizations and control by the state (e.g., Ellickson 1991). My model is devoid of substantive content, and requires only that the enforcement be centralized (i.e., independent of the individual group members). I thus use the generic term "selective incentives" from Olson.

coupling" of formal and informal control, with consequences for group productivity:

> To the extent that the formal rules are consonant with the preferences and interests of organizational actors, informal processes of social control largely subsume the cost of monitoring and enforcement . . . often leading to high economic and organizational performance. (p. 88)

The assumption that informal norms match the collective interest further predicts oppositional norms that hurt organizational performance whenever formal rules are dissonant with the collective interests of those who are asked to obey. Indeed, authors (Homans 1950; Shibutani 1978) cite classic cases of informal norms undermining formal control in such circumstances. If organization members will naturally influence each other to promote their common interests, peer influence will boost organizational performance whenever the formal rules and informal norms are congruent and will undermine performance when they are incongruent. This reflects the assumption that norms are functional for the group that enforces them: norms will either promote or oppose productivity depending on whether productivity is in the group's interests.

Heckathorn (1988) also considers the interplay of informal norms and centralized formal enforcement, including the possibility that group members "revolt" against the regime of informal control. Like Nee (1998), Heckathorn (1988) predicts a tendency for emergent norms to promote the collective interest. He predicts oppositional norms as well, but only when such norms correct "overproduction" and thus engender an optimal level of production for members. Thus, neither of these scholars observes or aims to explain antisocial norms that actually diminish collective welfare.

Rivalness and Value of Incentives: Interdependence Determines Regulatory Interests

The problem arises when the excludable goods used as selective incentives have the property of *rivalness* (Taylor 1987).[7] Where incentives are rival, such as when rewards come from a finite pool of resources, each worker's reward received diminishes the expected reward for other workers. I regard rivalness as an abstract property of incentive systems, a form of interdependence in actors' rewards, not as an intrinsic property of the goods themselves.

Consider a few familiar empirical illustrations of rivalness. If a professor assigns grades on the basis of an objective test with fixed evaluation criteria, she is offering students a nonrival incentive to work in the class: effort is rewarded with a grade, but the grade is independent of peers' performance. By contrast, if the professor assigns grades "on the curve" within a class, she is using a rival incentive: each student's achievement rewards come at some cost to other students. Similarly, a university honors award is rival. If the number of students who receive honors is fixed (e.g., top 10% of students), then each hardworking student diminishes the prospect of other students attaining honors. If the criterion is fixed (e.g., grade point average greater than 3.5), then there is no limit to the number of students who may obtain honors, but its value diminishes as more students attain it. In either case, the expected value of the reward to a hard worker diminishes as more peers earn the reward.[8]

It is more difficult to suggest examples of purely nonrival incentives, because most incentives exhibit at least some rivalness. Even the seemingly nonrival case of independent student grades may imply some rivalness at a broader level, such as where grades are aggregated to a grade point average for each student, with the students then assigned to a broader class rank in their senior year. Furthermore, some empir-

[7] Although the terms "rivalness" and its opposite "jointness of supply" generally refer to public goods, this article applies the same distinction to private goods. Macy (1993) and Oliver (1980) have similarly applied jointness to sanctions, examining its effect on second-order costs to enforcement agents. In contrast, I vary the value and rivalness of a centralized formal incentive, in which rivalness manipulates not the marginal cost to the central agent (who is exogenous), but the expected value of rewards to the recipients.

[8] Consider the designation of *cum laude* at Harvard University, which awarded honors to 91% of its undergraduates in 2001. In the words of former Dean, Henry Rosovsky, "Honors at Harvard has lost all meaning" (Healy 2001:A1).

ical instances of nonrival incentives seem motivated by sensitivity to issues discussed in this analysis. University departments, in an effort to discourage competition and enhance cooperation among junior colleagues, strive to award tenure in a nonrival fashion by evaluating tenure cases independently rather than comparatively within departments.

Although rivalness is an important part of competition, the latter term is vague and conflates this abstract form of interdependence with the value of the incentive and other trappings of rivalry. For example, increasing the value of a prize may heighten competition, but it does not make the prize more rival. Similarly, emotional states such as enmity or jealousy may heighten observed "competitiveness" in a contest, but they do not make the incentive more rival. I thus use the term "rivalness" to clearly represent this form of negative interdependence in an incentive system.

I argue that many goods described as selective incentives in collective action literatures are explicitly rival, and thus challenge the conventional assumption that individuals' choices to participate in collective action are independent. Of course, rewards of material benefits (e.g., a cash award for a "volunteer of the year") often imply negative interdependence, but this is also true of most intangible kudos rewards (e.g., esteem, status, or prestige), which are more often described as selective incentives. For example, in-group status and prestige are inherently comparative (Harary 1959).[9] Even without an explicit status ranking, reputations or esteem rewards diminish in value as they are spread over actors, implying negative interdependence among recipients.

Classic research in group dynamics (Deutsch 1949) offers a framework for understanding the effect of this interdependence on interpersonal influence. A nonrival incentive allows *promotive* interdependence: each member's choice to work for a collective good benefits all members while also earning herself a selective incentive. Thus, rational members will have personal inclinations to work for the incentive as well as regulatory interests in enforcing work among peers. In contrast, a rival incentive engenders *contrient* interdependence: each actor's work robs from other workers some share of the selective incentive. Whenever this loss to competition for the incentive exceeds the individual benefit from peers' contributions, group members who work will have a regulatory interest in opposing work among peers. Further empirical work (Blau 1963; Crombag 1966; Raven and Eachus 1963) found that group performance, motivation, and mutual evaluations were more positive when interdependence was promotive rather than contrient.

The fact that competition arising from rival performance rewards generates perverse pressures (e.g., ostracism of hard-working students) is intuitive and well known. Indeed, schoolteachers, athletic coaches, and managers of work teams are familiar with the destructive side effects of internal competition on cooperation. However, this point has unappreciated consequences for theories of norms and theories of collective action more generally. Recall that scholarship on norms has built on an assumption that norms are always prosocial, whereas collective action theory has always assumed that selective incentives to participate should promote the collective good. I show that the regimes of formal sanctioning and peer influence clash under the highly general condition of rivalness, making us challenge some of the most widely accepted derivations in both literatures. This intuitive and well-documented point about rivalness thus offers an important innovation to the general theories of norms and collective action.

Group Cohesiveness: Effectiveness of Social Influence

In choosing whether to comply with a prosocial formal rule, actors consider the direct costs and benefits of working as well as selective incentives. However, informal peer pressure can sometimes restrict their discretion and thus override their personal inclinations

[9] In a complementary model, Loch, Huberman, and Stout (2000:41) show that competition for status can undermine collective action. However, they focus on members' investment of resources in unproductive "politicking" activities instead of opposing or undermining peers' productivity. They do not consider regulatory interests, social influence, or the emergence of norms.

(Heckathorn 1988).[10] Although literatures in group dynamics have addressed factors that may affect the strength of peer influence, these specific mechanisms are exogenous to the theory presented in this article. Instead, I simply assume a continuum representing the effectiveness of influence without specifying the particular causes of variation in effectiveness. Following classic research on group dynamics, I call this dimension "cohesiveness" (Cartwright 1968; Lott and Lott 1965). This article investigates a broad range of conditions, from scenarios in which peer pressure is ineffective to scenarios in which such pressure is strongly determinative.

Experimental and ethnographic research has found that cohesiveness increases productivity (Blau 1963; Festinger, Schachter, and Back 1950), ostensibly because members are more responsive to influence by their peers (Back 1951; Schachter 1951), and are thus less inclined to free ride. However, other work (Berkowitz 1955; Schachter et al. 1951) has shown that effective social influence may diminish productivity if the valence of informal pressure turns against work:

> The greater the cohesiveness the greater the power of the group to influence its members . . . whether cohesiveness will increase or decrease productivity, however, is determined largely by the direction of group induction. (Schachter et al. 1951:230)

I take as exogenous that some groups exert more effective pressure on members' behavior than others. What I aim to explain is the "direction of group induction," by providing a rigorous theory of normative valence. Accounting for the content of norms in this way may improve our understanding of collective action and organizational performance. My model shows how collective action depends on a nonobvious three-way interaction between the *value* of selective incentives (ranging from worthless to very valuable), the *rivalness* of incentives (ranging from independence to zero-sum competition), and group *cohesiveness* (ranging from ineffective peer influence to extremely powerful peer influence).

ASSUMPTIONS

I begin with a basic collective goods problem, in which a set of individuals value a jointly produced good, but the good is not "excludable" from those who fail to contribute. Furthermore, "jointness of supply" implies that one member's consumption of the collective good does not reduce the amount available for others to consume. This article uses a task group as an anecdotal setting, including a set of members who may choose to *work* toward the collective good or *shirk* while enjoying the fruits of others' contributions.[11] This formalization involves several simplifying assumptions, some of which will be relaxed in sensitivity analyses and in later research. In this analysis, I assume the following:

1. All actors are risk-neutral myopic egoists who select a preferred course of action based on the information immediately available, but they neither learn from the past nor contemplate the long-run future. In particular, they cannot fathom an infinite regress of reciprocal influence between actors.
2. Actors have uniform interests, productive capacity, and power to influence peers.
3. The level of collective good produced is a strictly increasing function of the number of members who contribute toward its production. All group members receive an equal share of this collective good, regardless of their work choice.
4. A selective incentive is assigned to reward workers. However, to the extent that an incentive is rival, the expected value of the reward received by each worker is a decreasing function of the number of peers who also work.

[10] Some authors also may use the term "selective incentives" to refer to pressure by peers. To prevent confusion, I consistently use "selective incentive" to refer to the sanctions imposed by the central enforcement agent. I use the term "norms" to refer to the valence and pressure emerging among group members.

[11] Throughout this article, I use *group* as shorthand for a "set of individuals who value a collective good" and I use *work* as shorthand for "make a costly contribution to a collective good." Both terms are standard usages from Olson's (1965) model of collective action in interest groups. Although the theory certainly applies to collective action problems in the workplace (just as Olson used the example of labor union organizing), it does not imply any substantive relevance to employment, wages, or management–employee relations.

5. Work is costly and not directly cost effective for any actor. Thus, in the absence of formal control (selective incentives) or informal norms (peer pressure), no member will contribute to the collective good.
6. Actors cannot interact selectively or form special relations with certain peers. Each must choose to promote or oppose work uniformly across all peers.

MODEL

This section further specifies my assumptions in a very basic mathematical model of actors' inclinations to work and their regulatory interests in peers' work. I present the derived propositions in natural language here and give proofs in the Appendix.

The model considers two decisions: a first-order choice to contribute to the collective good *(work or shirk)* and a second-order choice to influence others' first-order choices *(promote or oppose peers' contributions)*. I begin with the production function, G, expressed as the net benefit received by each member i from all N members' work choices:

Production Function:

$$G(w_i,n) = n\,g + (g - c)\,w_i, \qquad (1)$$

where w_i denotes actor i's work choice ($w_i = 1$ for "work"; $w_i = 0$ for "shirk"), n is the total number of i's peers who are working ($0 \le n \le N-1$), g is a parameter representing i's benefit created by each member's work, and c is a parameter representing the cost of working. Clearly, actor i receives a benefit g for each of n peers who works, and one more unit g if she chooses to work herself. The cost of working for the collective good, c, is also a constant decrement for each worker, but this cost is borne privately by i, whereas the benefit is enjoyed by all. The linear production function implies that a member's direct cost and benefit of working remain the same, $(g - c)\,w_i$, regardless of peers' participation. Given assumption 5—that work is not personally profitable ($c > g$)—it is easy to see that the work choice will give a negative payoff for workers and a null payoff for shirkers. This yields the familiar N-person prisoner's dilemma: regardless of other members' first-order choices, *shirk* strictly dominates *work* in the baseline model.

Now I will consider the implications of formal control, in which an incentive is awarded only to those who contribute to the collective good. The possibility for selective incentives to promote work among egoists is well understood. Unlike previous research in collective action theory, I allow that selective incentives may be *rival*, and thus participation choices are interdependent: Each actor's choice to work for the collective good diminishes the expected share of selective incentives remaining for other workers. I have assumed that an incentive is awarded to workers, and that all workers receive an equal share, as represented by function R for actor i:

Reward Function:

$$R(w_i, n) = \mu\left(1 - \lambda\frac{n}{n+1}\right)w_i \qquad (2)$$

where λ is a parameter representing the rivalness of the incentive ($0 \le \lambda \le 1$), μ is the total value of the incentive ($\mu \ge 0$), and n and w_i are as defined in Equation 1. In the purely nonrival scenario ($\lambda = 0$), all workers receive the full value of the incentive (μ) regardless of others' choices. In the perfectly rival condition ($\lambda = 1$), worker i must share the selective incentive with n working peers, yielding a share of $\mu/(n+1)$. Intermediate values of λ allow for a range of partially rival incentives, in which higher values of λ indicate that each worker's share shrinks more steeply as n increases.

An actor i's utility is the sum of the production function G and reward function R, given her own work choice (w_i) and the number of her peers who work (n):

Utility Function:

$$U(w_i, n) = gn + w_i\left(g - c + \mu\left(1 - \lambda\frac{n}{n+1}\right)\right) \qquad (3)$$

where the g and c parameters are as defined in Equation 1, and μ and λ are as defined in Equation 2. We may interpret actor i's utility as the total production by peers ($g\,n$) plus the personal cost and benefit of working $(g - c)w_i$ and i's share of the selective incentive,

$$\mu\left(1 - \lambda\frac{n}{n+1}\right)w_i.$$

Each actor considers the payoff for working, or the change in utility associated with the work

choice, which I call the actor's inclination to work (*IW*):

Inclination to Work:

$$IW(n) = \frac{\partial U}{\partial w_i} = (g - c) + \mu\left(1 - \lambda\frac{n}{n+1}\right) \quad (4)$$

The inclination to work remains the same whether I treat the work choice as binary $w_i \in \{0,1\}$ or as a proportion of effort in the interval [0,1]. Because *i*'s utility depends linearly on w_i, *i* will never prefer an interior value of w_i (partial effort) to either extreme value. Simulations and discussion conventionally treat contribution choices as binary, but all propositions also hold if "work" is defined as any positive effort.

A positive inclination to work ($IW > 0$) means that the actor will profit from choosing to work, whereas $IW < 0$ implies a net loss for choosing to work. Analysis of Equation 4, detailed in the Appendix, allows us to derive three propositions.

Proposition 1: Actors' inclination to work increases with the value of the selective incentive.

Proposition 2: If, and only if, the incentive is valuable ($\mu > 0$) and peers are working ($n > 0$), actors' inclination to work decreases as the rivalness of the selective incentive increases.

Proposition 1 is intuitive, but Proposition 2 seems to conflict with the common intuition that people work harder in "competitive" groups. In fact, that usage of competition conflates my distinct dimensions of value and rivalness of incentives. I allow that when a good is more rival (i.e., an actor receives less of it or receives it with less certainty when peers also consume it), it should be less effective as an incentive. Proposition 3 shows the effect of peers' work on any actor's inclination to work:

Proposition 3: If, and only if, the selective incentive is valuable ($\mu > 0$) and rival ($\lambda > 0$), actors' inclination to work decreases as a greater number of peers work.

Given that *g*, *c*, and μ are homogeneous across the population, a larger number of peers working always implies a smaller share of a rival incentive for any worker and thus a lower inclination to work. If $\lambda = 0$ or no peers are working ($n = 0$), then there is no expected loss to peers and Equation 4 reduces to $IW = g - c + \mu$.

Now let us consider regulatory interests. All members receive a personal benefit *g* from each peer's contribution to the collective good, implying a positive baseline regulatory interest. However, when a rival ($\lambda > 0$) incentive is valuable enough to induce an actor to work, it may also create a perverse regulatory interest in opposing peers' participation. This is because increasing the number of peers who work (*n*) will also increase the loss of the incentive to peers, as specified in Equation 2. Any actor who is working must consider both the benefit and loss attributable to peers' work. The partial derivative of *U* with respect to *n* represents the change in utility attributable to a marginal change in the number of peers working. This yields the *regulatory interest function:*

Regulatory Interest Function:

$$RI(w_i, n) = \frac{\partial U}{\partial n} = g - w_i\frac{\lambda\mu}{(n+1)^2} \quad (5)$$

Treating regulatory interest as the marginal value (for a given actor *i*) of peers' work allows us to derive general propositions about the dependence of regulatory interests on the two parameters of the selective incentive.

Proposition 4: If, and only if, the selective incentive is rival ($\lambda > 0$), workers' regulatory interests in production decrease as the incentive grows more valuable.

Proposition 5: If, and only if, the selective incentive is valuable ($\mu > 0$), workers' regulatory interests in production decrease as the incentive grows more rival.

We have seen that the regulatory interests of shirkers, who do not receive the incentive, are uniformly positive under the simple model. We also have seen that the value and rivalness of the selective incentive are jointly relevant to members who expect to receive the incentive. For weak or nonrival incentives, they prefer that peers work toward the collective good, but for strong and rival incentives they prefer that their peers shirk. As proven in the Appendix, the qualitative conclusions represented by these propositions hold true regardless of the parameters of the production function, *g* or *c*, for any

finite group size (N) or current level of collective action (n).

TWO COMPUTATIONAL EXPERIMENTS

The propositions show how members' inclinations to work and their regulatory interests regarding peers' behavior vary with these parameters of the collective action problem when both choice processes operate independently. It would be simple to find the equilibrium number of members choosing to work in the absence of social pressure or to find the aggregated social pressure if given the number of members who work. However, it is not simple, particularly for the actors themselves, to generate predictions when both processes operate simultaneously. Following the assumption of bounded rationality, I posit that actors avoid this analytical puzzle by making locally optimal choices with constrained information and cognitive capacity.

In a pair of computer simulations, actors will make a first-order choice (*work* or *shirk*) followed by a second-order choice (*promote* or *oppose*), selecting the locally preferred option in each case. Either promoting or opposing entails an enforcement cost. Actors also may choose to *abstain* from enforcement (i.e., neither *promote* nor *oppose*) to avoid paying that enforcement cost. Thus the model allows for second-order free riding.

Although Propositions 4 and 5 are valid regardless of the expected effectiveness of social influence, an actor's costly choice to pressure peers requires her to assess the quantitative impact of her enforcement efforts. I do not assume that actors have common knowledge of the model that governs their peers' behavior; nor do I assume that actors are able to compute the portion of peers' behavior attributable to their own regulatory efforts. Such an inferential task is complicated by a stubborn regress: changes in peers' work choices affect their own regulatory interests, leading to further changes in norms before the actor has observed the direct effect of her own pressure on peers' subsequent work choices. For these reasons, individual actors have no means for accurately calculating the scope of their own influence. Allowing for individual-level idiosyncrasies, I assign a uniformly distributed random variable, θ_i, representing actor i's subjective *scope* of influence. This value is the maximum number of peers i expects to be able to influence, varying from the extreme belief that i's social pressure will have no effect ($\theta_i = 0$) to the extreme belief that i can convince all peers to change ($\theta_i = N - 1$).

For the enforcement choice, actor i first observes the number of peers who are currently working (n). I define n^+ as the number of peers that actor i expects to work if she promotes work. Note that $n^+ = \min(n + \theta_i, N - 1)$. That is, i expects her influence to yield a work level equal to n plus her scope (θ_i), but not to exceed the full set of peers ($N - 1$). The expected payoff for promoting is the change in utility for increasing the number of working peers from n to n^+. This payoff also includes a constant cost of enforcement (e):

Payoff for Promoting

$$P_{Promote} = U(w_i, n^+) - U(w_i, n) - e$$

which simplifies to (6)

$$= (n^+ - n)g - \left(\frac{n^+}{n^+ + 1} - \frac{n}{n + 1}\right)\lambda\mu w_i - e$$

An identical process operates for opposition. I use n^- to represent i's expected number of peers working because of his downward pressure, $n^- = \max(n - \theta_i, 0)$. That is, i expects his downward pressure to yield a work level of n minus his scope (θ_i), but not to fall below zero. This produces an analogous payoff for opposing[12]:

Payoff for Opposing:

$$P_{Oppose} = U(w_i, n^-) - U(w_i, n) - e$$

which simplifies to (7)

$$= (n^- - n)g - \left(\frac{n^-}{n^- + 1} - \frac{n}{n + 1}\right)\lambda\mu w_i - e$$

The payoffs for *promote* and *oppose* are both defined in contrast to *abstain*, which yields a

[12] I can also express $P_{Promote}$ or P_{Oppose} as a definite integral, $\int_a^b g - w_i \frac{\lambda\mu}{(n+1)^2}\, dn$, where a is the current number working, and b is the number i expects to work as a result of his upward or downward pressure.

standard expected payoff of zero. Figure 1 illustrates the enforcement choice for two workers, *A* and *B*, with different scope values (θ_i) and different numbers of peers currently working (n), but with all other parameters held constant.[13]

Actor *A* has a scope of 2 and is making the decision when only one of her peers is currently working. Believing she can force a maximum of two peers to change their behavior, *A* expects $n^- = 0$ if she opposes, $n = 1$ if she abstains, and $n^+ = 3$ if she promotes. Clearly, opposing work among peers brings the highest gain, and convincing two peers to work actually would result in a net loss because of competition over the incentive. Actor *B* has a scope of 1, and six of his peers are currently working. In this illustration, it can be seen that *B* would expect a gain by promoting and a loss by opposing.

In the norm enforcement decision, the actor first selects the normative *valence* (promote or oppose) that brings her the highest payoff, then faces a distinct choice of whether to enforce or abstain from enforcing that norm (i.e., free ride on others' enforcement efforts). If enforcing the preferred norm promises to bring an expected benefit that exceeds the cost, she will enforce. If neither *promote* nor *oppose* promises to yield a net profit (or if they yield identical payoffs) given n and θ_i, then actor i will *abstain*.

Recall that actors are homogeneous in their power to influence peers in this model. The total force acting upon all N actors to work or shirk is then a sum of individual agents' choices to *promote* ($v_i = 1$), *oppose* ($v_i = -1$), and *abstain* ($v_i = 0$):

$$V = \sum_{i=1}^{N} v_i \qquad (8)$$

In computing this group valence (V), it is not obvious how to deal with the effect on actor i of his own pressure. Are actors immune to their own regulatory efforts (Heckathorn 1993; Whitmeyer 2002), so that each actor faces a different normative environment (determined by the enforcement efforts of his $N-1$ peers)? Alternatively, do actors feel a need for consistency between their actions and their words or a punishment for hypocrisy (Festinger 1957; Kim and Bearman 1997), leading them to be influenced disproportionately by their own regulatory efforts? I see little reason to prefer either of these conflicting arguments generally, so I assume most parsimoniously (Heckathorn 1990) that actors' own pressure counts the same as others' pressure in figuring the normative valence that influences their future choices.

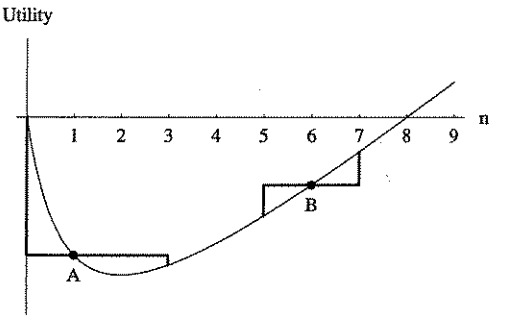

Figure 1. Two Actors' Expected Gains and Losses from Enforcement

Computational experiments will let the first- and second-order processes operate in tandem, allowing that agents' control efforts may influence members' work choices, which may further influence control efforts. I represent the strength of social influence as another parameter, α, "group cohesiveness." This parameter ($0 \leq \alpha < 1$) determines the extent that each actor's work choice (w_i) is influenced by the collective valence (V) as in Equation 8, versus the member's own inclination to work as in Equation 4.[14]

$$w_i = \begin{cases} 1 \text{ if } \alpha V + (1-\alpha)IW > 0 \\ 0 \text{ if } \alpha V + (1-\alpha)IW \leq 0 \end{cases} \qquad (9)$$

[13] I use the particular values $N = 10$, $\mu = 9$, $\lambda = 1$, and $g = 1$ for Figure 1. Although this qualitative shape is typical of a collective goods problem with a rival and moderately valuable incentive, my goal in presenting Figure 1 is simply to illustrate the implementation of influence scope.

[14] The simulations will consider a hypothetical group of ten members ($N = 10$), so the range of V is always [–10, +10], implying that peer pressure can be as important as the full value of the collective good. This is equivalent to modeling V as the *mean* over all v_i, but rescaling v to [–10, +10]. Of course, the numerical scales of valence and cohesiveness are arbitrary. The values used here allow us to investigate a broad range of influence effectiveness. If V were defined

Actors' choices to promote or oppose peers' work may depend on their own choices to work, and their choices to work may depend on the pressure they receive. To derive predictions from this model, we must resolve this feedback. A conventional solution is to compute work (w_i) and enforcement (v_i) choices iteratively, using a sequential decision model (Heckathorn 1990; Oliver, Marwell, and Teixeira 1985). The simulation loops through the list of actors in random order. An actor weighs her inclination to work (reflecting the share of the incentive currently available) against the norm valence, as in Equation 9. Given her updated work choice, she then decides whether to promote or oppose. Both choices may in turn affect peers' inclinations and regulatory interests. The simulation protocol and other supplementary materials are available from the author online (http://faculty.washington.edu/kitts/).

Experiment 1

The dependent variable of interest in Experiment 1 is the stable proportion of actors choosing to work ("participation"), as I investigate a space of three parameters of interest (μ, λ, and α). All simulations assume the same baseline values for the collective action problem ($g = 1$, $c = 5$, $N = 10$).[15] I map the proportion of members working as I finely manipulate the value of the selective incentives (μ) and cohesiveness (α), and coarsely manipulate rivalness by comparing the response surfaces for the nonrival ($\lambda = 0$) and maximum rivalness ($\lambda = 1$) conditions.

I examine a wide range of μ, from worthless selective incentives ($\mu = 0$) to valuable selective incentives that never fall below the cost of working ($\mu = 50$), and a range of cohesiveness, from ineffective peer influence ($\alpha = 0$) to influence that is almost strictly determinative ($\alpha = 0.98$). I exclude the uninteresting special case in which all actors have exactly $\alpha = 1.0$, because this would suggest that everyone in the group ignores the costs and benefits of work. I thus map the level of collective action for all important combinations of formal and informal control. The simulations begin with the conventional initial condition of free riding at both levels: $w_i = 0$ and $v_i = 0$ for all actors $i \in \{1, 2, \ldots N\}$. Once actor i has made a work choice $w_i \in \{0,1\}$ in the first iteration, he then adopts an enforcement choice $v_i \in \{-1,0,+1\}$. The next actor in the sequence then makes her choices.

Although authors typically interpret iterations as the passage of time, readers need only think of them as a numeric search for a stable distribution of behaviors in the model, given parameter values. Iteration halts if no choices have changed since the previous round, as no further change can occur. With a nonrival incentive of at least moderate value, the model converges to stable levels of collective action during the first few iterations (i.e., after each actor has made three or four choices). Such rapid convergence is typical for models of this type (Whitmeyer 2002). However, regions of the parameter space in which the solution includes peer pressure will generally never reach actor-level stability, but instead exhibit stochastic closed orbits around the given mean values as individual actors exchange social pressure. The results are indistinguishable whether I let the system iterate for only 5 rounds or for 10,000 rounds, but the latter allows slightly more "converged" solutions, so I report the results of the more extensive search.[16]

A $2 \times 50 \times 51$ factorial design implements two levels of rivalness $\lambda \in \{0,1\}$, , 50 levels of cohesiveness $\alpha \in \{0, 0.02, 0.04, \ldots 0.98\}$, and 51 levels of incentive value $\mu \in \{0.1.2, \ldots 50\}$, giving a total of 5,100 unique parameter combinations. Each of these unique conditions is replicated 250 times, and means for response

in the range [–1, +1], higher values of cohesiveness (α) would be required to observe the same effects.

[15] Here I introduce a numeric value for production cost c, which had not been relevant for regulatory interest. Although this exact value is arbitrary, it allows that a member's cost of working always greatly exceeds his direct benefit to working, as assumed, and that the value of selective incentives will exceed the cost of working in a broad range of conditions. This particular value matches a common assumption in public goods experiments: each member's contribution yields an aggregated benefit for the group ($Ng = 10$) that is twice as large as her cost ($c = 5$).

[16] Of the simulations in the nonrival condition, approximately 90.4% converged to a stable fixed point within five rounds; 0.5% converged at a later point; and 9.1% never converged to actor-level stability (i.e., followed a stochastic closed orbit around the given mean productivity) in 10,000 rounds.

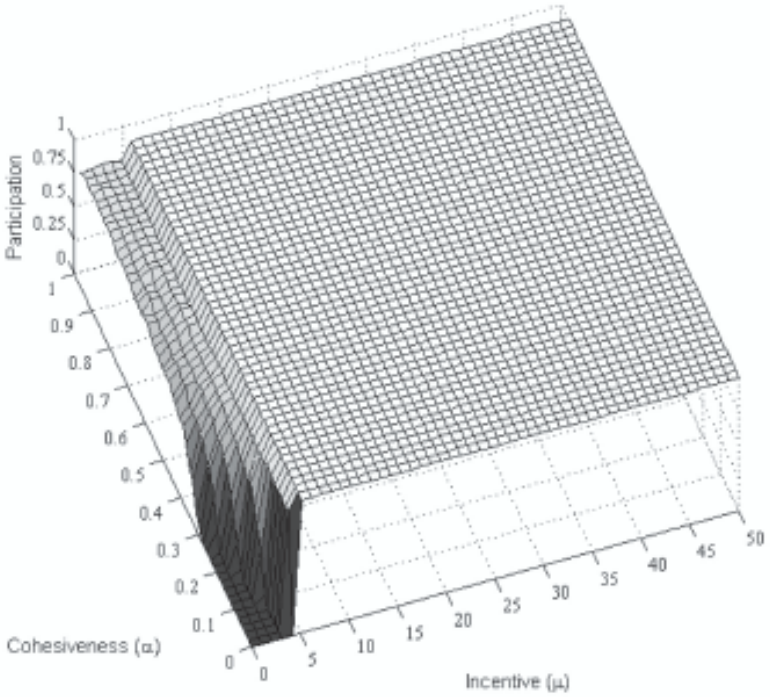

Figure 2. Proportion Working Over Range of Value (μ) of Selective Incentives and Cohesiveness (α) in Group; *Nonrival* Incentive

Note: $g = 1$, $c = 5$, $e = 2$, $N = 10$, $\lambda = 0$.

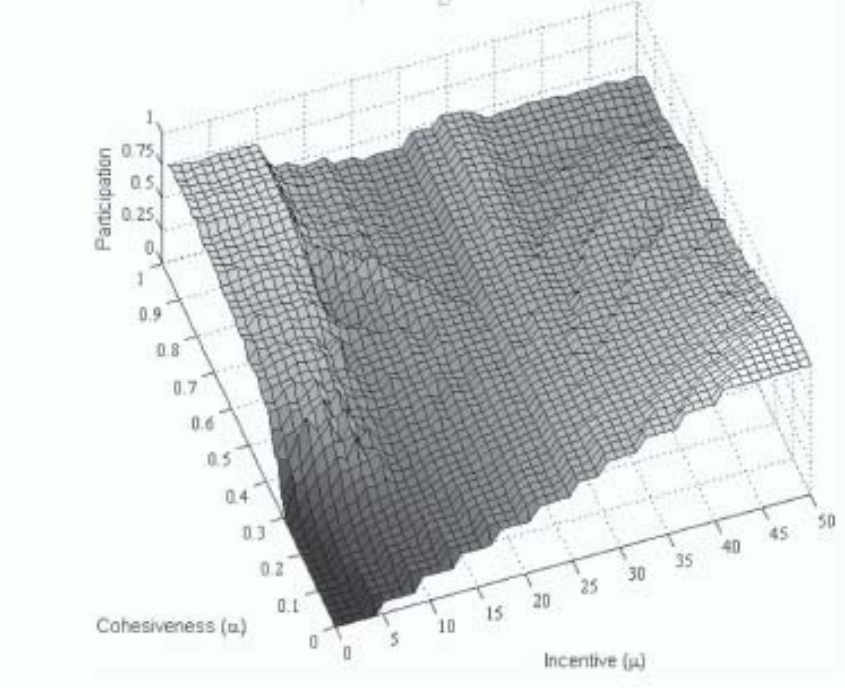

Figure 3. Proportion Working Over Range of Value (μ) of Selective Incentives and Cohesiveness (α); *Rival* Incentive

Note: $g = 1$, $c = 5$, $e = 2$, $N = 10$, $\lambda = 1$.

variables are computed from the final iteration in each simulation. Experiment 1 thus includes 1,275,000 independent observations on the model's stable behavior. The large number of replications allows me to describe the mean of response variables across the ranges of α and μ parameters (at minimum and maximum λ) without using inferential statistics. We can be practically certain that averaging over an infinite number of replications would yield a qualitatively identical response surface.[17] Replications guard only against sampling error, of course. Sensitivity analyses later investigate other forms of robustness.

Figure 2 depicts collective action outcomes for the nonrival ($\lambda = 0$) case.

As expected, no actors will work in the corner where social pressure is toothless (α near zero) and the selective incentive is too weak to justify work ($\mu \leq 4$), although universal work would maximize collective welfare. In the nonrival condition, a rising value of the selective incentive obviously fosters work.[18] When incentives are too small to compensate workers, members are inclined to shirk. However, members still have regulatory interests in promoting work among peers, and this leads them to enforce prosocial norms whenever they deem doing so as cost effective. Informal control "kicks in" as cohesiveness rises along the left edge of Figure 2. Participation in collective action rises with cohesiveness until it reaches 0.8 (8 of 10 members working), where second-order free riding prevents further progress: No member can profit in trying to force the remaining members to work, because of the enforcement cost ($e = 2$).

Figure 3 shows that results are very different when the incentive is rival ($\lambda = 1$):[19]

It can be seen that participation in collective action is lower over virtually the entire space of μ and α in Figure 3 (for maximum rivalness), as compared with Figure 2 (for zero rivalness). This overall drop in participation with rivalness obtains even for the slice without informal influence ($\alpha = 0$), because of the simple effect that rivalness has on inclinations to work (as in Proposition 2). However, the collapse of participation in the hybrid-control corner (high-μ/high-α) is attributable to enforcement of antisocial norms. Examining the peak levels of participation, see that the effects of cohesiveness (peer influence) and selective incentives are both highly contingent. Across all levels of rivalness, participation is maximal either in groups of atomized actors with a valuable incentive (low-α, high-μ) or in highly cohesive groups with a weak incentive (high-α, low-μ).

It is instructive to consider norm enforcement patterns (both prosocial and antisocial) that underlie the collective action response surfaces shown in this analysis. I describe overall patterns here and provide surface plots of norm enforcement in the author's supplementary materials (http://faculty.washington.edu/kitts/). Where participation in collective action had been climbing with cohesiveness (on the left edge of Figure 3), there is a corresponding drop in promotion. Here, effective prosocial norms undermine the demand for prosocial norms by increasing the portion of the group with perverse regulatory interests. When cohesiveness is low, however, rising incentive value encourages work and discourages promotion for the same reason.

We might expect that enforcement of prosocial norms should generally fall as a rival incentive grows more valuable (an intuition based on Proposition 4), but this is not the case when informal pressure is powerful (high-α). We can understand this counterintuitive result by considering the incidence of antisocial norms (opposition to collective action). In fact, opposition jumps when the incentive exceeds the value necessary for at least two

[17] I also implemented a $2 \times 100 \times 251$ factorial design [$\lambda \in \{0,1\}$, $\alpha \in \{0,0.01,0.02 \ldots 0.99\}$, and $\mu \in \{0,0.2,0.4 \ldots 50.0\}$] using 1,000 replications at each point, but only five iterations in time instead of 10,000. The results from these 50,200,000 independent simulations were identical to those shown here.

[18] Most of the range of incentive value here is above the level needed to support full productivity, given that $\lambda = 0$, so most of the surface in Figure 2 is flat. I nevertheless show the entire response surface to facilitate comparison with surfaces in the rival incentive condition.

[19] For the rival condition in Experiment 1, approximately 0.2% of simulations converged to a stable fixed point within five rounds; 1.6% converged at a later point; and 98.1% never converged to actor-level stability (but followed a stochastic closed orbit around the mean shown) in 10,000 rounds.

actors to work $\mu > 8$), but then levels out as μ rises further. Where peer influence is effective (high-α), opposition diminishes the number of workers, keeping demand for antisocial norms stable as μ rises. This yields a balance between prosocial and antisocial norms in the high-α/high-μ region. An intriguing result obtains for extremely valuable incentives and low cohesiveness. Although competition is fiercest here by any measure, and thus regulatory interests of workers turn against the collective good, the level of antisocial norms surprisingly drops in this corner. This occurs because the incentive makes working attractive, whereas low cohesiveness makes antisocial norms too weak to prevent work from reaching a high level. Once most members are working, they would not expect to recover enforcement costs for opposing peers' work, so they abstain.[20] Second-order free riding thus diminishes antisocial norms in this corner.

I am now ready to derive hypotheses for the relationship between the three investigated parameters and participation in collective action. I refrain from making general arguments based on the local patterns and texture of the response surface, and focus on the strongest and most robust observations of the model's qualitative behavior seen in Figures 2 and 3.

Hypothesis 1: Contingent effect of cohesiveness: The effect of cohesiveness (strength of peer pressure) on collective action will depend on the value and rivalness of selective incentives. A rise in cohesiveness will increase participation in collective action only where the incentive is weak or nonrival. Where the incentive is both strong and rival, participation will fall as cohesiveness increases, because of peer enforcement of antisocial norms.

Hypothesis 2: Contingent effect of selective incentive: The effect of a centralized selective incentive on collective action will depend on the rivalness of the incentive and on group cohesiveness. At low cohesiveness or low rivalness, participation in collective action will increase with the value of the incentive. When rivalness and cohesiveness are both high, participation will fall as the incentive becomes more valuable, because of peer enforcement of antisocial norms.

These hypotheses represent global patterns of model behavior that are extremely robust to variations of the simulation protocol and peripheral assumptions. Among the subtle features of the response surface in Figure 3, one of the most intriguing is the nonmonotonic relationship between collective action and the value of the incentive observed in the middle range of cohesiveness: participation crashes when the incentive exceeds a critical value, then slowly rises again as the incentive grows stronger. A replication of Experiment 1 over the full range of rivalness shows that in this middle range of cohesiveness, the shape of the relationship between the incentive and participation in collective action depends decisively on the rivalness parameter.[21] A striking bifurcation appears, in which the concavity of the relationship between participation and incentive value reverses at a key value of rivalness. However, a second computational experiment here provides an accessible view, investigating the two dimensions of incentive value and rivalness, within this middle range of cohesiveness.

Experiment 2

This experiment maps participation in collective action across the entire range of rivalness (λ) and the same broad range of incentive value (μ). Instead of manipulating the strength of peer influence as a system-level parameter, I allow heterogeneity, implementing α as a uniform random variable in the range (0,1). It plays the same role as in Equation 9 and I use the same symbol. However, I now interpret α_i as actor *i*'s *responsiveness to influence.* Thus the actors in Experiment 2 are heterogeneous in two aspects:

[20] This paradoxical drop in antisocial norms for high incentive value would not obtain for the extreme case in which all actors had unlimited scope of influence, but is quite robust if all actors have a limited scope.

[21] To see this surface animated over the range of rivalness, from $\lambda = 0$ (as in Figure 2) to $\lambda = 1$ (as in Figure 3), see the author's supplementary materials (http://faculty.washington.edu/kitts/).

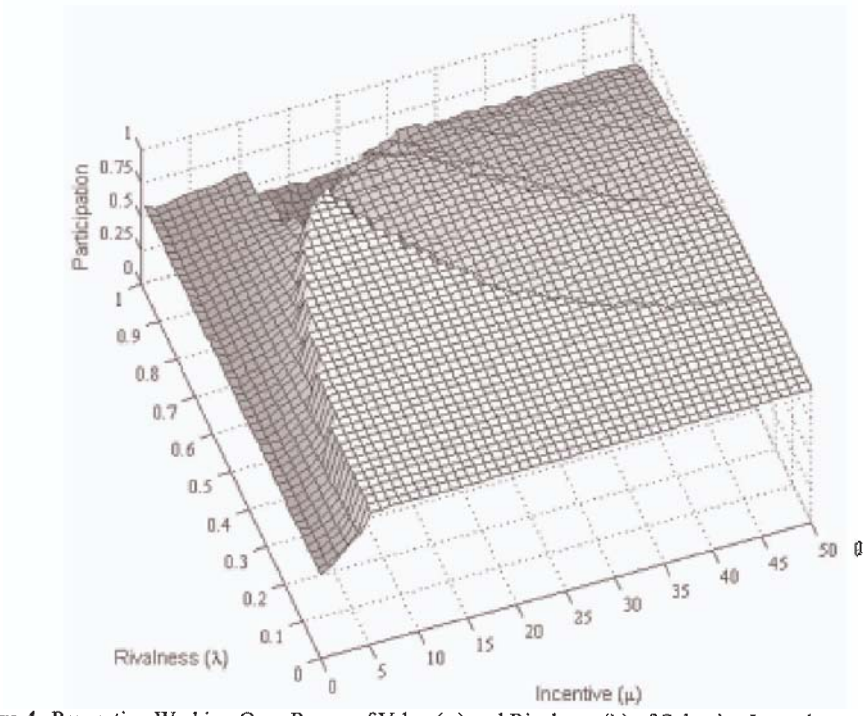

Figure 4. Proportion Working Over Range of Value (μ) and Rivalness (λ) of Selective Incentives

Note: $g = 1$, $c = 5$, $e = 2$, $N = 10$.

their expected scope of influence (θ_i) and their responsiveness to peer pressure (α_i).[22]

I map the level of collective action as I manipulate the value (μ) and rivalness (λ) of the selective incentive in a 51 × 51 factorial design, with the expected scope of influence (θ_i) and susceptibility to peer influence (α_i) heterogeneous across actors and all other parameters uniform ($g = 1$, $c = 5$, $e = 2$, $N = 10$).[23] The simulation begins with universal free riding [$w_i = 0$ and $v_i = 0$ for all actors $i \in \{1, 2, \ldots N\}$], and uses the same decision protocol as in Experiment 1. Figure 4 shows the proportion working for the collective good over the space of incentive value (μ) and rivalness (λ).

First, notice a fin-shaped region in which all group members choose to work. On the left side, there is a drop in participation when the individual share of the incentive falls below the amount needed to compensate all group members for working. Cooperation in this left region depends on prosocial norms, and promotion of peers' work jumps from zero to a high level below this ridge. The collapse in participation in the high-value, high-rivalness corner reflects the operation of antisocial norms. The author's supplementary materials include the corresponding response surfaces for prosocial and antisocial norms (http://faculty.washington.edu/kitts/).

Exploring the full continuum of rivalness in Experiment 2 allows a more nuanced view of the relationships between the selective incentive, informal norms, and collective action, now with the additional scope condition that peer influ-

[22] I do not assume a correlation between these two variables. Making α variable does not imply that actors have differing powers to influence the group norm, only that they care more or less about the norm. Furthermore, there is no reason to assume that actors know each other's α levels.

[23] That is, I manipulate both parameters in fine increments, where $\lambda \in \{0, 0.02, 0.04, \ldots 1.0\}$ and $\mu \in \{0, 1, 2, \ldots 50\}$, performing 250 replications as in Experiment 1. In this experiment, approximately 69.3% converged to a stable fixed point within five rounds; approximately 9.8% of simulations converged at a later point; and 20.9% never converged to actor-level stability (i.e., followed a stochastic closed orbit around the given mean productivity) in 10,000 rounds.

ence is moderately strong. At low rivalness, collective action increases monotonically with the value of the selective incentive. At moderately high rivalness, collective action initially increases with the value of incentives, then crashes as the incentive grows very valuable, because of antisocial norms. At very high rivalness, collective action initially crashes at a low incentive value because of antisocial norms, then increases with value. The existence of a bifurcation point on the dimension of rivalness where the shape of the relationship of the incentive to collective action qualitatively shifts is not at all obvious, but represents a robust observation of the model's qualitative behavior. In fact, this nonmonotonic relationship of incentive value to collective action (and its interaction with rivalness) reflects very basic and generic properties of the assumed choice processes.

Recall that Experiment 2 implements susceptibility to influence (α_i) as uniformly distributed in (0,1) over actors for each simulation. Group-level cohesiveness (i.e., average strength of in-group influence) is then normally distributed over simulations, with an observed mean of .500 and a standard deviation of .091. Thus, we can think of Experiment 2 as a slice of the middle range of cohesiveness exploring the dimensions of rivalness and value of incentives. Comparing the middle slice of Figure 3 (where α is in the neighborhood of .5 and $\lambda = 1$) to the far slice of Figure 4 (where the group mean α is in the neighborhood of .5 and $\lambda = 1$) suggests that heterogeneity of susceptibility to influence does not alter the shape of the response surface, although the level of productivity is marginally lower when susceptibility is heterogeneous over actors.

Sensitivity Analysis

Mathematical analysis has explicated the simple models of work inclinations and regulatory interests. Computational experiments combined these models and explored their joint behavior across the space of key parameters. These simulations used a sequential decision algorithm, a simulation tool widely used in the relevant sociological literature. Given the remarkable robustness of the simulation results, we can be confident that the reported conclusions from the computational experiment do not reflect "flukes" attributable to the stochastic component of the model. However, numeric investigation does not itself allow inference beyond the particular parameter values and initial conditions implemented by the researcher. To map the space defined by two key properties of selective incentives (value and rivalness) and also cohesiveness, the numeric investigation of the model has held constant the size of the group (N), the cost of working (c) and enforcing (e), and the individuals' productive capacity (g). In the Discussion and in the author's supplementary materials (http://faculty.washington.edu/kitts/), I explore the scope of my conclusions as these assumptions are relaxed.

Given the difficulty of estimating the efficacy of one's own influence, I assign scope values randomly in the computational experiment. Although I leave a detailed study investigating the subjective scope of influence for future research, sensitivity analyses have demonstrated that hypotheses are also robust to specification of scope, in which all actors share the same scope from $\theta = 3$ to $\theta = 9$, and also allow θ to be heterogeneous. Of course, if all actors have $\theta_i \leq 2$ (more generally, $\theta_i \leq e/g$), this implies that promoting work will never yield an expected profit for any actor.[24] Thus, when the entire group has a negligible scope, no prosocial norms will emerge, and we thus will no longer see a positive relationship between cohesiveness and productivity for a weak or nonrival incentive. This result is obvious, so I implement a cost of enforcement and distribution of scope to allow that prosocial influence is possible, but still problematic.

To investigate the robustness of the main findings, I implemented a variety of alternative simulation protocols. For example, I replicated the sequential decision model shown here (using random sequencing of actors' updates) with a repeating sequence over actors, then replicated it again with a synchronous update protocol (updating all actors' choices at once). These alternative simulation protocols did not change the qualitative conclusions.

[24] Given the numeric values used in this study ($N = 10$, $e = 2$, $g = 1$), uniform random draws will yield such an impoverished distribution of scope with a miniscule probability (2.28×10^{-6}).

DISCUSSION

The propositions in this article are statements about actors' regulatory interests and inclinations to participate in collective action that can be proven true under highly general conditions. The hypotheses are robust observations of the computational model's qualitative behavior that serve as novel predictions for empirical research on collective action. This thorough exploration of the model allows us to draw scope conditions on these hypotheses to guide empirical research. For example, the hypotheses require that pressuring peers be not so costly (relative to group members' expectation of their efficacy in influencing peers) that members would never bother to pressure peers. Equivalently, the hypotheses require that at least some group members should expect influence efforts to have an effect on peer behavior that may at least in some cases exceed the cost of sanctioning. Animations of Figures 2, 3, and 4, with scope varying from the minimum to the maximum of its range, are provided in the author's supplementary materials (http://faculty.washington.edu/kitts/).

Reports in the author's supplementary materials also investigate a broad range of assumptions about costs of enforcement, animating Figures 2, 3, and 4 as enforcement costs increase from 0 to 5. Results shown here are representative and conclusions are robust. Notably, the hypotheses also follow if peer sanctioning is perfectly costless, as some scholars assume for gossip or "esteem sanctions" (McAdams 1997:365). However, subtle patterns in the response surface may depend on the presence of a nontrivial sanctioning cost. For example, the paradoxical decrease in antisocial norms when rivalness and incentive value are both very high and cohesiveness is low depends explicitly on sanctioning cost, because it occurs due to second-order free riding. If sanctioning is free, this finding disappears. This "second-order free-riding *benefit*" is one of many intriguing and novel observations offered by the model. Because it depends on auxiliary assumptions (specification of sanctioning cost and subjective scope of influence), however, I suggest a closer formal investigation of this phenomenon before a hypothesis is derived for empirical investigation.

The propositions are proved under the model for any finite group size. However, the simulations used to derive the hypotheses have used auxiliary assumptions that are most plausible for small groups. The effectiveness of peer enforcement may decrease and its cost may increase with group size (Heckathorn 1988:552). Similarly, the marginal impact of peer competition shrinks as the number of workers grows large, making the prediction of antisocial norms sensitive to the initial conditions for large groups. Finally, the assumptions of undifferentiated influence structure and uniform distribution of influence scope grow less plausible with increasing group size. For these reasons, small group size is an important scope condition for my hypotheses. Group size is held constant at 10 to facilitate comparison with related models (Heckathorn 1990; Whitmeyer 2002).

This article focuses on the standard case of positive incentives (rewards for working toward the collective good), but may be elaborated to include punishments for shirking. The derived propositions could easily be restated for the case of negative incentives, in which rivalness implies a diminishing potency or probability of punishment as more peers shirk and thus share the punishment. However, the simulations are not general proofs and must implement an explicit incentive regime. Thus, hypotheses derived from the simulations apply to positive incentive systems.

The simplifying assumption that work choices are binary and all workers receive an identical selective incentive focuses on an important and conventional case, but future work may consider other cases such as differential rewards based on effort, skill, or favoritism. I expect that this will not alter my main conclusion, that incentive value and rivalness interact to create perverse regulatory interests and antisocial norms, but that such elaborations may affect which actors choose to participate or pressure peers.

Qualitatively different collective action problems are described by linear, accelerating, decelerating, or sigmoid production functions (Heckathorn 1996; Marwell and Oliver 1993). This article uses a simple linear function that represents a classic N-person prisoner's dilemma. Investigating the interplay of nonlinear production functions with alternative reward functions is beyond the scope of this article, but offers a promising direction for further work. Such further work also can allow other sources of heterogeneity among actors such as produc-

tive capacity, costs of influence, and interpersonal power (Heckathorn 1993; Oliver et al. 1985).

A more realistic dynamic model would allow actors to incorporate some feedback about the effectiveness of their influence efforts and thus adjust their subjective scope over time. This more sophisticated treatment of influence scope will become more relevant when we allow for important heterogeneity in actors' power to influence peers, such that very powerful actors may correctly perceive their own power.

Although experimental control prevents "spurious" effects in this model,[25] empirical research may be confounded by processes not considered in this discussion. For example, increasing the rivalness of an incentive system may diminish cohesiveness (Blau 1963; Bothner, Stuart, and White 2004; Crombag 1966; Raven and Eachus 1963) or increase the subjective value of an incentive. Increasing cohesiveness may inflate the value of the collective good (as individuals value each other's welfare) or selective incentives (as prestige becomes more salient before a valued audience). The model can be elaborated to investigate the implications of such feedback systematically.

CONCLUSION

This study demonstrates some counterintuitive implications of hybrid systems of control, in which centralized selective incentives to work for the collective good are combined with opportunities for decentralized peer influence. A core innovation is the point that many selective incentives to engage in collective action (especially rewards of esteem, status, or prestige) are *rival.* That is, the value of incentives to ego may decrease as a function of the number of ego's peers who share the incentives. This innovation relaxes the ubiquitous assumption in collective action theory that the value of the selective incentive to ego is independent of peers' choices.

I have proved that this negative interdependence implies perverse regulatory interests that may lead recipients of rival incentives to oppose peers' work for the collective good. Where incentives are too weak to justify compliance, of course, all actors will have a regulatory interest in forcing peers to work. In this standard scenario, collective action depends on effective peer influence and will be undermined by second-order free riding. When presented with potent rival incentives, however, members who receive the incentives will have a perverse interest in opposing work among peers. In this novel scenario, effective social influence can undermine collective action, and second-order free riding can save it. The computational experiments have mapped out the conditions under which these two opposite scenarios should obtain.

The first experiment mapped collective action as the values of selective incentives and group cohesiveness (the strength of peer influence) vary, comparing this surface for minimal rivalness with the same surface given maximal rivalness. This analysis showed that either strong centralized rewards or strong peer influence may lead to high participation in collective action, but that their combination is volatile when incentives are rival, due to the emergence of antisocial norms. Equilibrium in this region entails a mixture of strategies at both the first and second order, wherein a surprising normative dissensus obtains even in a homogeneous population that uniformly values a collective good.[26]

The second experiment allowed for heterogeneity in peer influence and mapped the model's behavior over both parameters of the selective incentive, showing an intriguing interaction of the value and rivalness of the reward. While affirming the results of the first experiment, it investigated the model over the range of rivalness and also showed that conclusions are robust to heterogeneity in susceptibility to influence.

[25] Note that μ represents the subjective value of the incentive to the recipient, and α represents susceptibility to influence after all causes are taken into account. Both are exogenous independent variables, manipulated directly as part of the experimental design, and do not vary during a simulation run.

[26] A replication of Experiment 1 with all actors having homogeneous scope also found this diversity in both work and enforcement choices. Those results are available in the author's supplementary materials (http://faculty.washington.edu/kitts/).

I have proved the tendency for rivalness and value of selective incentives to pull regulatory interests away from the collective good under the model for any finite group size. Whether these regulatory interests translate into observable social pressure may depend on finer details of the model such as enforcement costs, subjective scope of influence, and group size, as I have shown. Although extensive sensitivity analyses have demonstrated robustness to auxiliary assumptions, I have held group size constant and refrained from deriving hypotheses for the effect of group size, while acknowledging that the various simplifying assumptions will be most plausible for small groups.

This project has offered a distinct contribution to contemporary models of formal and informal control. In his seminal paper, Heckathorn (1990) noted that "virtually all sanctions generate externalities" (p. 367). However, he meant that sanctions have corresponding "spillover" externalities (such as pride enjoyed by the family of a laureate or shame suffered by the family of a convict). Heckathorn did not consider the possibility that compliance (and receipt of incentives) by any actor creates negative externalities for other compliant actors, as I have shown to be generally true where selective incentives are rival. Heckathorn did consider "oppositional" norms, but his model did not allow actors to negatively influence peers' work, only to neutralize others' control efforts. Finally, oppositional norms in Heckathorn's model did not turn against the collective good, but solved the problem of "overcontrol," in which peer influence compels compliance even where marginal cost exceeds marginal productivity.[27] Heckathorn's oppositional norms thus promoted Pareto-optimal outcomes for the group. Of course, my results neither subsume nor contradict Heckathorn's findings. My findings depend on the assumption of rivalness, which was not part of his model. I have argued that many of the paradigmatic examples of selective incentives in collective action literatures are significantly rival, but this is not true for all incentives.

Note that I have not *argued* that members are homogeneous in their beliefs, values, and power, that individuals are unaffected by history or by special relations to others, or that monitoring and assignment of selective incentives are free of error in the empirical world. There are many well-known obstacles to collective action, and I have disregarded these to investigate a fundamental problem that has not been recognized in the general theories of norms and collective action. My analysis has demonstrated that even in a system without friction (biases, errors, or transaction costs), norms deriving from egoists' regulatory interests may diverge from the collective good in the presence of a valuable and rival incentive. My simulations have explored implications for collective action while allowing for enforcement costs and actor-level heterogeneity.

Formalizing the theory has allowed rigorous derivation of hypotheses within well-defined scope conditions. After extensive exploration, I have emphasized the most robust, meaningful, and nonobvious predictions that follow from the basic model. The appropriate next step in direct empirical validation is to design critical experiments that fit the scope conditions specified in this discussion, and to investigate my hypotheses in the laboratory. Whereas the formalization allows us to ask the right questions and thus design decisive empirical tests, it also sensitizes us to the simplifying assumptions required to derive those hypotheses. I expect that elaboration of this general model (in ways suggested throughout the article) should inform development of richer "middle-range" theories for more specific empirical contexts. The simulation has strengthened the link between theory and empirical research by making the derivation of hypotheses both rigorous and transparent.

The presented model does not give formal predictions for empirical problems that violate the basic assumptions of the model.[28] It may be

[27] Such a scenario—termed the "altruists dilemma" (Heckathorn 1991), in which prosocial behavior actually diminishes the group's welfare and selfish opportunism benefits the group—represents a different theoretical problem. This article focuses on the core class of prisoner's dilemma models, in which cooperation is collectively optimal and yet individuals may free ride on others' contributions.

[28] By contrast, informal discursive theory may naively offer hypotheses for a broad range of empirical contexts that, unbeknown to the authors, violate crucial assumptions required to derive their hypotheses. Without rigorous derivation of hypotheses, empir-

elaborated incrementally to derive hypotheses for cases in which decisions are continuous rather than binary, social relations vary among group members, and multiple dimensions of status differentiation exist, building on the simple foundation offered in this article.

The computational experiments also have generated a wealth of intriguing and counterintuitive conjectures, which deserve much more formal elaboration than I have been able to perform in this study. A few examples include the surprising dip in antisocial norms as the rival incentive grows very valuable (at low cohesiveness) and the contingent effect of second-order free riding, which may enhance or diminish collective action depending on the value and rivalness of the centralized selective incentive. Such elaborations should show the distinct scope conditions for these ancillary results and investigate their interplay with the very general processes described in this article.

James Kitts *is an Assistant Professor in the Department of Sociology at the University of Washington. He investigates the emergence and stability of norms using formal models and empirical studies of voluntary associations. He is broadly interested in the relations between individuals through their participation in groups and the relations between groups through members' participation choices. His recent projects have focused on social networks, the dynamics of affiliation, and participation in social movement organizations.*

MATHEMATICAL APPENDIX

It is simple to prove the presented propositions by partial differentiation of the key functions.

Proposition 1: Actors' inclination to work increases with the value of the incentive.

To see how inclinations to work depend on the value of the selective incentive, we can examine the partial derivative of *IW* with respect to μ:

$$\frac{\partial IW}{\partial \mu} = 1 - \left(\frac{n}{n+1}\right)\lambda \qquad (10)$$

See in Equation 10 that *IW* will always be a strictly increasing function of μ, for any $\lambda \in [0,1]$ and positive finite number of peers working (n). Thus, Proposition 1 follows.

Proposition 2: If, and only if, the incentive is valuable ($\mu > 0$) and peers are working ($n > 0$), actors' inclination to work decreases as the rivalness of the incentive increases.

Similarly, $\partial IW/\partial\lambda$ shows how actors' inclination to work depends on rivalness.

$$\frac{\partial IW}{\partial \lambda} = -\left(\frac{n}{n+1}\right)\mu \qquad (11)$$

See in Equation 11 that *IW* will always be a strictly decreasing function of λ when $\mu > 0$ and $n > 0$. If $\mu = 0$ or $n = 0$, then $\partial IW/\partial\lambda = 0$. Thus, Proposition 2 follows.

Proposition 3: If, and only if, the selective incentive is valuable ($\mu > 0$) and rival ($\lambda > 0$), actors' inclination to work decreases as a greater number of peers work.

To investigate the dependence of inclinations to work on the number of peers currently working, we can differentiate *IW* with respect to n:

$$\frac{\partial IW}{\partial n} = -\frac{\lambda\mu}{(n+1)^2} \qquad (12)$$

Given that the denominator must be positive for any level of work among peers (n), it can be seen that this derivative will be strictly negative, as long as $\mu > 0$ and $\lambda > 0$. If $\mu = 0$ or $\lambda = 0$, then $\partial IW/\partial n = 0$. Thus, Proposition 3 follows.

Proposition 4: If, and only if, the selective incentive is rival ($\lambda > 0$), workers' regulatory interests in production decrease with the value of the incentive.

To investigate the dependence of regulatory interests on the value of the selective incentive, we can differentiate *RI* with respect to μ:

$$\frac{\partial RI}{\partial \mu} = -w_i \frac{\lambda}{(n+1)^2} \qquad (13)$$

Because the denominator must be positive, this derivative must be negative when $\lambda > 0$ and $w_i = 1$. The relationship of workers' regulatory interests to the value of the incentive is then strictly negative. Also, because the derivative is negative for all positive finite values of n, regulatory interests must decrease with the value of the incentive for workers regardless of

ical research may give little decisive leverage on theory, offering us little guidance for interpreting contradictory empirical findings.

peers' participation (n) or i's expected scope of influence (θ). Of course, $\partial RI/\partial\mu = 0$ for nonrival incentives ($\lambda = 0$) and members who do not earn the incentive ($w_i = 0$). Thus, the remainder of Proposition 4 follows.

Proposition 5: If, and only if, the selective incentive is valuable ($\mu > 0$), workers' regulatory interests in production decrease with the rivalness of the incentive.

Similarly, $\partial RI/\partial\lambda$ shows how actors' regulatory interests depend on the rivalness of the selective incentive:

$$\frac{\partial RI}{\partial \lambda} = -w_i \frac{\mu}{(n+1)^2} \qquad (14)$$

Given that the denominator must be positive, $\partial RI/\partial\lambda$ must be strictly negative when $\mu > 0$ and $w_i = 1$. The relationship of workers' regulatory interests to the rivalness of the incentive is then strictly negative. As for incentive value, this does not depend on n or θ. Of course, $\partial RI/\partial\lambda = 0$ for valueless incentives ($\mu = 0$) or for shirkers ($w_i = 0$). Thus, the rest of Proposition 5 follows.

REFERENCES

Allison, Paul D. 1992. "The Cultural Evolution of Beneficent Norms." *Social Forces* 71:279–301.

Axelrod, Robert. 1984. *The Evolution of Cooperation.* New York: Basic Books.

———. 1986. "An Evolutionary Approach to Norms." *American Political Science Review* 80:1095–111.

Back, Kurt W. 1951. "Influence through Social Communication." *Journal of Abnormal and Social Psychology* 46:9–23.

Bendor, Jonathan and Dilip Mookherjee. 1987. "Institutional Structure and the Logic of Collective Action." *American Political Science Review* 81:290–307.

Bendor, Jonathan and Piotr Swistak. 1997. "The Evolutionary Stability of Cooperation." *American Political Science Review* 91:290–307.

———. 2001. "The Evolution of Norms." *American Journal of Sociology* 106:1493–545.

Berkowitz, Leonard. 1955. "Group Standards, Cohesiveness, and Productivity." *Human Relations* 7:509–19.

Bicchieri, Cristina and Yoshitaka Fukui. 1999. "The Great Illusion: Ignorance, Informational Cascades, and the Persistence of Unpopular Norms." *Business Ethics Quarterly* 9:127–55.

Blau, Peter M. 1963. *The Dynamics of Bureaucracy: A Study of Interpersonal Relations in Two Government Agencies.* Chicago, IL: University of Chicago Press.

Bothner, Matthew S., Toby E. Stuart, and Harrison C. White. 2004. "Status Differentiation and the Cohesion of Social Networks." *Journal of Mathematical Sociology* 28:261–95.

Boyd, Robert, Herbert Gintis, Samuel Bowles, and Peter J. Richerson. 2003. "The Evolution of Altruistic Punishment." *Proceedings of the National Academy of Sciences* 100:3531–35.

Boyd, Robert and Peter J. Richerson. 1992. "Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups." *Ethology and Sociobiology* 13:171–95.

———. 2001. "Norms and Bounded Rationality." Pp. 281–96 in *Bounded Rationality: The Adaptive Toolbox,* edited by G. Gigerenzer and R. Selten. Cambridge, MA: MIT Press.

Cartwright, Dorwin. 1968. "The Nature of Group Cohesiveness." Pp. 91–109 in *Group Dynamics: Research and Theory,* edited by Dorwin Cartwright and Alvin Zander. New York: Harper & Row.

Chong, Dennis. 1991. *Collective Action and the Civil Rights Movement.* Chicago, IL: University of Chicago Press.

Coleman, James S. 1990. *Foundations of Social Theory.* Cambridge, MA: Harvard University Press.

Crombag, Hans F. 1966. "Cooperation and Competition in Means-Interdependent Triads: A Replication." *Journal of Personality and Social Psychology* 4:692–95.

Deutsch, Morton. 1949. "An Experimental Study of the Effects of Cooperation and Competition upon Group Process." *Human Relations* 2:199–231.

Eggertsson, Thrainn. 2001. "Norms in Economics, with Special Reference to Economic Development." Pp. 76–104 in *Social Norms,* edited by Michael Hechter and Karl-Dieter Opp. Berkeley, CA: University of California Press.

Ellickson, Robert C. 1991. *Order Without Law: How Neighbors Settle Disputes.* Cambridge, MA: Harvard University Press.

———. 2001. "The Evolution of Social Norms: A Perspective from the Legal Academy." Pp. 35–75 in *Social Norms,* edited by Michael Hechter and Karl-Dieter Opp. Berkeley, CA: University of California Press.

Festinger, Leon. 1957. *A Theory of Cognitive Dissonance.* Stanford, CA: Stanford University Press.

Festinger, Leon, Stanley Schachter, and Kurt Back. 1950. *Social Pressures in Informal Groups.* New York: Harper.

Hanneman, Robert A., Randall Collins, and Gabriele Mordt. 1995. "Discovering Theory Dynamics by Computer Simulation: Experiments on State Legitimacy and Imperialist Capitalism." *Sociological Methodology* 25: 1–46.

Harary, Frank. 1959. "Status and Contrastatus." *Sociometry* 22:23–43.

Healy, Patrick. 2001. "Harvard's Quiet Secret: Rampant Grade Inflation." *Boston Globe,* October 7, pp A1.

Hechter, Michael. 1987. *Principles of Group Solidarity.* Berkeley, CA: University of California Press.

Hechter, Michael and Elizabeth Borland. 2001. "National Self-Determination: The Emergence of an International Norm." Pp. 186–233 in *Social Norms,* edited by Michael Hechter and Karl-Dieter Opp. Berkeley, CA: University of California Press.

Hechter, Michael and Karl-Dieter Opp. 2001. "What We Have Learned about the Emergence of Social Norms?" Pp. 394–415 in *Social Norms,* edited by Michael Hechter and Karl-Dieter Opp. Berkeley, CA: University of California Press.

Heckathorn, Douglas D. 1988. "Collective Sanctions and the Creation of Prisoner's Dilemma Norms." *American Journal of Sociology* 94:535–62.

———. 1990. "Collective Sanctions and Compliance Norms: A Formal Theory of Group-Mediated Social Control." *American Sociological Review* 55:366–84.

———. 1991. "Extensions of the Prisoner's Dilemma Paradigm: The Altruist's Dilemma and Group Solidarity." *Sociological Theory* 9:34–52.

———. 1993. "Collective Action and Group Heterogeneity: Voluntary Provision versus Selective Incentives." *American Sociological Review* 58:329–50.

———. 1996. "The Dynamics and Dilemmas of Collective Action." *American Sociological Review* 61:250–77.

Hirshleifer, David and Eric Rasmusen. 1989. "Cooperation in a Repeated Prisoners' Dilemma with Ostracism." *Journal of Economic Behavior and Organization* 12:87–106.

Homans, George C. 1950. *The Human Group.* New York: Harcourt, Brace, & World.

———. 1961. *Social Behavior: Its Elementary Forms.* New York: Harcourt, Brace, & World.

Horne, Christine. 2001a. "The Enforcement of Norms: Group Cohesion and Meta-Norms." *Social Psychology Quarterly* 64:253–66.

———. 2001b. "Sociological Perspectives on the Emergence of Norms." Pp. 3–34 in *Social Norms,* edited by Michael Hechter and Karl-Dieter Opp. Berkeley, CA: University of California Press.

Ingram, Paul and Karen Clay. 2000. "The Choice-Within-Constraints New Institutionalism and Implications for Sociology." *Annual Review of Sociology* 26:525–46.

Kanazawa, Satoshi. 1997. "A Solidaristic Theory of Social Order." *Advances in Group Processes* 14:81–111.

Kim, Hyojoung and Peter S. Bearman. 1997. "The Structure and Dynamics of Movement Participation." *American Sociological Review* 62:70–93.

Kitts, James A. 2003. "Egocentric Bias or Information Management? Selective Disclosure and the Social Roots of Norm Misperception." *Social Psychology Quarterly* 66:222–37.

Kuran, Timur. 1995. *Private Truths, Public Lies: The Social Consequences of Preference Falsification.* Cambridge, MA: Harvard University Press.

Loch, Christoph H., Bernardo A. Huberman, and Suzanne Stout. 2000. "Status Competition and Performance in Work Groups." *Journal of Economic Behavior and Organization* 43:35–55.

Lott, Albert J. and Bernice E. Lott. 1965. "Group Cohesiveness as Interpersonal Attraction: A Review of Relationships with Antecedent and Consequent Variables." *Psychological Bulletin* 64:259–309.

Macy, Michael W. 1993. "Backward-Looking Social Control." *American Sociological Review* 58:819–36.

Macy, Michael W. and Robb Willer. 2002. "From Factors to Actors: Computational Sociology and Agent-Based Modeling." *Annual Review of Sociology* 28: 143–166.

March, James G., Martin Schulz, and Xueguang Zhou. 2000. *The Dynamics of Rules: Change in Written Organizational Codes.* Stanford, CA, Stanford University Press.

Marwell, Gerald and Pamela E. Oliver. 1993. *The Critical Mass in Collective Action: A Micro-Social Theory.* New York: Cambridge University Press.

McAdams, Richard H. 1997. "The Origin, Development, and Regulation of Norms." *Michigan Law Review* 96:338–433.

Nee, Victor. 1998. "Norms and Networks in Economic and Organizational Performance." *American Economic Review* 88:85–89.

Oliver, Pamela E. 1980. "Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Investigations." *American Journal of Sociology* 85:1356–75.

Oliver, Pamela E., Gerald Marwell, and Ruy Teixeira. 1985. "A Theory of Critical Mass: I. Interdependence, Group Heterogeneity, and the Production of Collective Action." *American Journal of Sociology* 91:522–56.

Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups.* Cambridge, MA: Harvard University Press.

Opp, Karl-Dieter. 2001. "Social Networks and the Emergence of Protest Norms." Pp. 234–73 in *Social Norms,* edited by Michael Hechter and Karl-Dieter Opp. Berkeley, CA: University of California Press.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* New York: Cambridge University Press.

Raven, Bertram H. and H. Todd Eachus. 1963.

"Cooperation and Competition in Means-Interdependent Triads." *Journal of Abnormal and Social Psychology* 67:307–16.

Schachter, Stanley. 1951. "Deviation, Rejection, and Communication." *Journal of Abnormal and Social Psychology* 46:190–207.

Schachter, Stanley, Norris Ellertson, Dorothy McBride, and Doris Gregory. 1951. "An Experimental Study of Cohesiveness and Productivity." *Human Relations* 4:229–38.

Sherif, Muzafer. 1966. *The Psychology of Social Norms.* New York: Harper & Row.

Shibutani, Tomatsu. 1978. *Derelicts of Company K: A Sociological Study of Demoralization.* Berkeley, CA: University of California Press.

Taylor, Michael. 1982. *Community, Anarchy, and Liberty.* New York: Cambridge University Press.

———. 1987. *The Possibility of Cooperation.* New York: Cambridge University Press.

Ullmann-Margalit, Edna. 1977. *The Emergence of Norms.* Oxford, UK: Oxford University Press.

Voss, Thomas. 2001. "Game-Theoretic Perspectives on the Emergence of Social Norms." Pp. 105–36 in *Social Norms,* edited by Michael Hechter and Karl-Dieter Opp. Berkeley, CA: University of California Press.

Whitmeyer, Joseph M. 2002. "The Compliance You Need for a Cost You Can Afford: How to Use Individual and Collective Sanctions?" *Social Science Research* 31:630–52.

Yamagishi, Toshio. 1986. "The Provision of a Sanctioning System as a Public Good." *Journal of Personality and Social Psychology* 51:110–16.