DISCUSSION PAPER SERIES

IZA DP No. 2106

Mutual Monitoring in Teams: Theory and Experimental Evidence on the Importance of Reciprocity

Jeffrey Carpenter
Samuel Bowles
Herbert Gintis

April 2006

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Mutual Monitoring in Teams: Theory and Experimental Evidence on the Importance of Reciprocity

Jeffrey Carpenter
Middlebury College
and IZA Bonn

Samuel Bowles
Santa Fe Institute
and University of Siena

Herbert Gintis
Central European University
and Santa Fe Institute

Discussion Paper No. 2106
April 2006

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 2106
April 2006

ABSTRACT

Mutual Monitoring in Teams: Theory and Experimental Evidence on the Importance of Reciprocity

Monitoring by peers is often an effective means of attenuating incentive problems. Most explanations of the efficacy of mutual monitoring rely either on small group size or on a version of the Folk theorem with repeated interactions which requires reasonably accurate public information concerning the behavior of each player. We provide a model of team production in which the effectiveness of mutual monitoring depends not on these factors, but rather on strong reciprocity: the willingness of some team members to engage in the costly punishment of shirkers. This alternative does not require small group size or public signals. An experimental public goods game provides evidence for the behavioral relevance of strong reciprocity in teams.

JEL Classification: C92, H41, J41, J54, Z13

Keywords: team production, public good, monitoring, punishment, experiment

Corresponding author:

Jeffrey P. Carpenter
Department of Economics
Middlebury College
Middlebury, VT 05753
USA
Email: jpc@middlebury.edu

1 Introduction[1]

Monitoring by peers in work teams, credit associations, partnerships, local commons situations, and residential neighborhoods is often an effective means of attenuating incentive problems that arise where individual actions affecting the well being of others are not subject to enforceable contracts. Most explanations of the incentives to engage in mutual monitoring (Varian 1990, Stiglitz 1993) rely either on the small size of the interacting group, or on repeated interactions and low discount rates, allowing the Folk theorem to be invoked (Fudenberg and Maskin 1986, Fudenberg et al. 1994). Neither of these is completely satisfactory, since work teams are not always small and the Folk theorem has little explanatory power.[2] Other treatments leave the incentive to engage in mutual monitoring unexplained (Arnott 1991, Weissing and Ostrom 1991).[3]

We provide a model of team production in which the effectiveness of mutual monitoring depends not on these factors, but rather on 'strong reciprocity': the willingness of some team members to engage in the costly punishment of shirkers. The key conditions supporting mutual monitoring are (a) contributing to production becomes a team norm, and (b) a fraction of team members are 'reciprocators' who punish violators of team norms even when this is costly to do. We call this altruistic punishment and those practicing it strong reciprocators.

An experimental public goods game provides evidence for the behavioral relevance of strong reciprocity in teams. In a treatment approximating a one-shot interaction, and on the terminal round of the game, shirkers are punished. We also find that shirkers respond to punishment by increasing their contribution to the public good. Altruistic motivations cannot account for our results. We conclude with some results specifying conditions under which mutual monitoring motivated by strong reciprocity provides an effective solution to incentive problems arising from incomplete contracting, as well as conditions under which mutual monitoring is likely to fail.

The problem of free riding in teams has been addressed by two standard models. The first, due to Alchian and Demsetz (1972), holds that residual claimancy should be assigned to an individual designated to monitor team members' inputs, thus ensuring the incentive compatibility for the (non-contractible) activity of monitoring itself, while addressing the members' incentive to free ride by the threat of dismissal by the monitor. They contrast this view of the 'classical firm,' as they call it, with an alternative in which team members are residual claimants and monitoring is performed, if at all, by salaried personnel. Alchian and Demsetz correctly observe that group residual claimancy would dilute incentives, but simply

[1] We would like to thank Mark Howard for assistance with the experiment and the John D. and Catherine T. MacArthur and the National Science Foundations for financial support.

[2] The repeated game solution to the problem of sustaining cooperative behavior in teams has several weaknesses, including: (a) there are a multiplicity of equilibria, most of which do not exhibit high levels of cooperation; (b) subgame perfection (i.e., the credibility of threats to punish non-cooperators) requires an implausible degree of coordination among team members.

[3] Dong and Dow (1993b) and Legros and Matthews (1993) assume the team can impose collective sanctions on shirkers. This assumption is reasonable if shirking is easily detected and team members have more effective or lower cost forms of punishment than are available to a traditional firm. We do not make this assumption. Dow and Dong (1993a) assume shirking can be controlled by the threat of non-shirkers to exit the team. However the threat of exiting is credible only if team members have very high fallback positions—in Dong and Dow's model, this takes the form of independent production—which generally is not the case.

posit the allocational superiority of the classical firm: "we assume that if profit sharing had to be relied on for all team members, losses from the resulting increase in central monitor shirking would exceed the output gains from the increased incentives of other team members not to shirk." (1972:786) As we will see, their invocation of the so-called "$1/n$ problem" to justify this assumption is not entirely adequate.

The second approach, pioneered by Holmström (1982), demonstrates that in principal multi-agent models one can achieve efficiency or near-efficiency through contracts that make individual team members residual claimants on the effects of their actions without conferring ownership rights on them. Contracts of this type typically impose large penalties for shirking and require large lump-sum up-front payments on the part of agents, or they pay each team member the entire team output minus a large constant and thus, in the presence of stochastic influences on output, entail negative payments in some periods, or at best a substantial variance of income to team members. These arrangments are infeasible if team members have insufficient wealth. Moreover, where contributions (e.g., work effort) are continuously variable these incentive mechanisms support large numbers of Nash equilibria, thus rendering breakdown of cooperation likely.

These approaches do not explain how mutual monitoring works, but rather why it may be unnecessary. The limited applicability of the owner-monitor and optimal contracting approaches provides one motivation for exploring the relationship between residual claimancy and mutual monitoring in teams. Another motivation is empirical. There is some evidence that group residual claimancy is effective, by comparison with payments unrelated to group output, even in quite large teams (Ghemawat 1995, Hansen 1997, Knez and Simester 2001). Mutual monitoring based on residual claimancy appears to be effective in the regulation of common pool resources such as fisheries, irrigation, and grazing lands (Ostrom 1990), in the regulation of work effort in producer cooperatives (Greenberg 1986, Craig and Pencavel 1995) and in the enforcement of non-collateralized credit contracts (Banerjee, Besley and Guinnane 1994). Experimental studies (Frohlich et al., 1998) provide additional support for the effects of residual claimancy in inducing lower supervision costs and higher productivity in (small) work teams. Further, the fact that residual claimancy may provide incentives for monitoring even in quite complex settings and large groups is suggested by evidence that in the United States home ownership is a significant predictor of participation in community organizations (Glaeser and DiPasquale 1999) and local politics but, significantly, not national politics (Verba et al., 1995), as well as willingness to monitor and sanction coresidents who transgress social norms (Sampson et al., 1997).

Making team members residual claimants can have positive incentive effects, since team members may have privileged access to information concerning the activities of other team members, and may have means of disciplining shirkers and rewarding hard work that are not available to third parties. As residual claimants, moreover, team members may have the incentive to use this information and exercise their sanctioning power, even if the team is large. Thus while Alchian and Demsetz are surely correct in saying that residual claimancy in large teams does not substantially reduce the direct incentive to free ride, it may support superior means of sanctioning and hence discouraging free riding through mutual monitoring. Our experiments suggest the motive to monitor team members includes a positive utility attached to punishing wrong-doers, independently of any expectation of material gain which might accrue to the punisher as a result of modification of the subsequent behavior of the punished shirkers. Monitoring is costly, however, and if the desire to monitor is not sufficiently widespread, we shall see, mutual monitoring will fail.

2 Strong Reciprocity

We will show that under certain conditions, residual claimancy by team members can provide sufficient incentives for mutual monitoring, and thus support high levels of team performance. A key element in our approach, one shared by recent contributions of Kandel and Lazear (1992), Rotemberg (1994), Banerjee et al. (1994), and Besley and Coate (1995) is that our model is based on 'social preferences' which, while unconventional, are well supported by recent experimental and other research.

We assume that though team members observe one another in their productive activity, they cannot design enforceable contracts on actions because this information is not verifiable (cannot be used in courts). In this situation we show that under appropriate conditions the assignment of residual claimancy to team members will attenuate incentive problems even when teams are large.

Two common characteristics of successful mutual monitoring are uncontroversial: the superior information concerning non-verifiable actions of team members available to other team members and the role of residual claimancy in motivating members to acquire and use this information in ways that enhance productivity. Less clear is whether residual claimancy motivates costly monitoring in large groups.[4]

A parsimonious explanation of mutual monitoring is provided, however, by the notion of *strong reciprocity*: the well-documented human propensity to cooperate with those who obey, and to punish those who violate social norms, even when this behavior cannot be justified in terms of self-regarding, outcome-oriented preferences (Campbell, 1983).[5] We distinguish this from *weak reciprocity*, namely reciprocal altruism, tit-for-tat, exchange under complete contracting, and other forms of mutually beneficial cooperation that can be accounted for in terms of self-regarding outcome-oriented preferences. The commonly observed rejection of substantial positive offers in experimental ultimatum games is consistent with this interpretation.[6] Moreover the fact that offers generated by a computer rather than another person are significantly less likely to be rejected suggests that those rejected offers at to cost to themselves are reacting to violations of norms rather than simply rejecting disadvantageous offers (Blount, 1995). More directly analogous to the team production case, however, are findings in n-player public goods experiments. These provide a motivational foundation for mutual monitoring in teams whose members are residual claimants, since these experiments show that agents are willing to incur a cost to punish those whom they perceive to have

[4] The problem of motivating the peer-monitors would not arise, of course, if team members were sufficiently altruistic towards teammates. In this case members would simply internalize the benefits conferred on others by their monitoring. Rotemberg (1994) develops a model of this type. However were team members sufficiently altruistic in this sense to motivate mutual monitoring, there would be no initial free rider problem either.

[5] Kandel and Lazear (1992), which is otherwise closest to our approach to modelling mutual monitoring, do not admit a reciprocity motive, but rather assume that members monitor and punish to increase their individual material payoffs.

[6] See Roth (1995) for a survey.

treated them or a group to which they belong badly.[7] In these experiments, which allow subjects to punish non-cooperators at a cost to themselves, the moderate levels of contribution typically observed in early play often rise in subsequent rounds to near the maximal level, rather than declining to insubstantial levels as in the case where no punishment is permitted. It is also significant that in the experiments of Fehr and Gächter, punishment levels are undiminished in the final rounds, suggesting that disciplining norm violators is an end in itself (de Quervain et al. 2004) and hence will be exhibited even when there is no prospect of modifying the subsequent behavior of the shirker or potential future shirkers (Walker and Halloran 2004, Carpenter and Matthews 2005).

Reciprocal preferences depend on one's belief about the behavior of the individuals with whom one deals. To model behaviors in a large $n-$player public goods setting, we say that individual i's utility depends on his own material payoff and the payoff of other individuals $j = 1 \ldots n$ according to:

$$u_i = \pi_i + \sum_j \frac{a_i + \lambda_i a_j}{1 + \lambda_i} \pi_j, \tag{1}$$

where $a_i, a_j \in [-1, 1]$ and $\lambda_i \geq 0$ (Rabin 1993, Levine 1998). The parameter a_i is i's level of unconditional good will or ill will (altruism or spite) towards others, and a_j is i's judgement of j's good will, while λ_i indicates the balance of unconditional vs. conditional elements in i's return to other members' material payoffs. If $a_i = 0$, then i is a non-altruistic reciprocator (exhibits neither good will nor spite towards others, but conditions behavior on the goodness or spitefulness of others). If $\lambda_i = 0$, then i exhibits unconditional altruism or spite, depending on the sign of a_i.

In the model below we assume that individuals behave as if they were maximizing a function such as (1) above and form their judgments of others on the basis of the others' contributions to the public good. Thus, if one is altruistic and believes that the others are also altruistic, one may engage in conditional generosity, by contributing to the public good. But if the generosity is not reciprocated by one or more of the relevant population, one alters one's judgement of others' generosity. In this case utility maximization may lead the individual to reduce his own contribution, or to punish low contributors if this is feasible, and not too costly. Note that this motivation for punishing a shirker values the punishment *per se* rather than the benefits likely to accrue to the punisher if the shirker responds positively to the punishment. For this reason the motivation to punish is independent of the size of the team.

The willingness to engage in costly punishment provides a basis for linking residual claimancy with mutual monitoring, even in large teams. An individual who shirks inflicts harm on the other members of the team if (and only if) they are residual claimants. Members may then see this violation of reciprocity as reason to punish the shirker. We should note that our model requires only that a certain fraction of team members be reciprocators. This is in line with the evidence from experimental economics, which indicates that in virtually every experimental setting a certain fraction of the subjects do not act reciprocally, either

[7] See Ostrom et al. (1992) on common pool resources, Fehr et al. (1997) on efficiency wages, and Fehr and Gächter (2000) on public goods. Coleman (1988) develops the parallel point that free riding in social networks can be avoided if network members provide positive rewards for cooperating.

because they are self-interested, or they are purely altruistic.[8]

In support of the analytical model developed in Section 3 we report in Section 4 an experiment carried out by the authors involving a public goods game with costly punishment. This experiment replicates Fehr and Gächter (2000) in that there is a positive level of punishment in all periods, and the level of cooperation does not decay when costly punishment is repeated. In addition, we show that the level of punishment directed towards a team member increases with the social cost that member imposes on the group due to shirking. In particular, the level of punishment and cooperation increase with the degree of team residual claimancy, and do not decrease when team size increases. Moreover, players appear to understand quite clearly the incentives they create and to which they respond, since after the first few rounds of play, punishment is directed virtually exclusively towards shirkers, and punishment in one round leads shirkers to increase their contribution in the next round. Finally, we show that these results are not due to altruism on the part of punishers, since players punish shirkers even when the costs of punishing shirkers exceed the increase in group earnings afforded by punishing.

3 Mutual Monitoring in Teams

Consider a team with n members ($n > 3$), each of whom can supply an amount of effort $1 - \sigma \in [0, 1]$. We call σ_i the *level of shirking* of member i. We assume the members of the team share their output equally. For convenience, if we refer to i or j, we assume they are team members in $\{1, \dots, n\}$ unless otherwise stated, and if we refer to both i and j, we assume $i \neq j$. Also we write $n_{-i} = \{k = 1, \dots, n | k \neq i\}$.

Let σ_j be j's level of shirking, so $\bar{\sigma} = \sum_{j=1}^{n} \sigma_j / n$ is the average level of shirking. We assume the cost of working (not shirking) is one dollar, and working adds q dollars to team output. We call q the *social productivity of cooperation.* Each member's payoff is then given by $q(1 - \bar{\sigma})$. The payoff loss to the team from one member shirking completely is q, of which the shirker's share is q/n, so the shirker's net gain from shirking is

$$g = 1 - \frac{q}{n}, \tag{2}$$

which we assume is strictly positive. We also assume $q > 1$, otherwise universal shirking would be optimal.[9]

Consider a single team member j. Another member $i \in n_{-j}$ can punish j at cost $c_i(s_{ij}(\sigma_j)) > 0$, where s_{ij} is the punishment i imposes on j if j shirks at level σ_j. The cost $c_i(s_{ij})$ may involve public criticism, shunning, threats of physical harm and the like. We assume that acts of punishment, like work effort, are non-verifiable and hence not subject to contract. We also assume $c_i(s_{ij})$ is increasing and strictly convex, and $c_i'(0) = 0$. Alternatively, i can trust and never punish j, which costs zero.

Member i can judge member j only on the basis of j's the level of shirking. Therefore, we make $a_j = 1 - 2\sigma_j$ in (1), so $a_j = -1$ if j completely shirks, and $a_j = 1$ if j does not

[8] For an especially clear example, see Blount (1995). Fehr and Schmidt (1999) provides a survey.

[9] Most of the homogeneity assumptions we make can be dropped, at the expense of complicating the notation and the descriptions of the model.

shirk at all. It is easy to check that the gain to i from punishing j is given by

$$\frac{\lambda_i(2\sigma_i - 1) - a_i}{1+\lambda_i} s_{ij} - c_i(s_{ij}). \tag{3}$$

Member i will then choose s_{ij} to maximize utility in (1), giving rise to the first order condition (assuming an interior solution)

$$c_i'(s_{ij}) = \frac{\lambda_i(2\sigma_j - 1) - a_i}{1+\lambda_i} \equiv \alpha_i(\sigma_j). \tag{4}$$

Note that when $\alpha_i(\sigma_j) < 0$, or equivalently, when

$$\sigma_j \leq \sigma_j^0 = \frac{1}{2}\left[\frac{a_i}{\lambda_i} + 1\right],$$

the maximization problem has a corner solution in which i does not punish. For $\sigma_j \geq \sigma_j^0$, the level of punishment s_{ij} is increasing in σ_j. Let $s_{ij}^*(\sigma_j)$ be the level of punishment of j that maximizes i's utility when j shirks at level σ_j. Then clearly

$$\rho_i(\sigma_j) \equiv u_i(\sigma_j) = \frac{\lambda_i(2\sigma_i - 1) - a_i}{1+\lambda_i} s_{ij}^*(\sigma_j) - c_i(s_{ij}^*(\sigma_j)) \tag{5}$$

is i's *subjective gain* from punishing j when j shirks at level j and i chooses a utility-maximizing level of punishment.[10]

If μ_{ij}, $i \in n_{-j}$, the probability that i punishes j, is chosen to be a best response, we have

$$\mu_{ij} \begin{cases} = 0, & \rho_i(\sigma_j) < 0 \\ \in (0,1), & \rho_i(\sigma_j) = 0 \\ = 1, & \rho_i(\sigma_j) > 0 \end{cases} \tag{6}$$

Let s_j be the expected punishment inflicted by all $i \in n_{-j}$ on j if j shirks. We have

$$s_j(\sigma_j) = \sum_{i \in n_{-j}} \mu_{ij} s_{ij}^*(\sigma_j). \tag{7}$$

From (2), we see that the expected gain to j from shirking, including the expected cost of punishment, is $g - s_j(\sigma_j)$. Therefore if σ_j is chosen as a best response, we have

$$\sigma_j \begin{cases} = 0, & g < s_j(\sigma_j) \\ \in (0,1), & g = s_j(\sigma_j) \\ = 1, & g > s_j(\sigma_j) \end{cases} \tag{8}$$

[10] Note that, unlike a member's share of the firm's net revenue, the subjective gain from punishing does not decline with the size of the team. We motivate this assumption and discuss the effects of team size below. For simplicity we have assumed that i's propensity to punish, $\rho_i(\sigma_j)$, is not affected by the propensities to punish or the observed rates of punishing of other members of the team. Replacing this with the assumption that punishing propensities are positively related opens the possibility of multiple equilibria, some involving high levels of punishing and some low. We explore this alternative below.

If $g < 0$ there is a unique, Pareto efficient, Nash equilibrium in which no members shirk and no member punishes. In this case residual claimancy alone is sufficient to ensure efficiency. The more interesting case, however, is where group size is sufficiently large that $q/n < 1$, so shirking is an individual best response in the absence of punishing. In this case any Nash equilibrium involves positive shirking, since if $\sigma_j = 0$ for some j then by (6) and $\rho_i(0) = 0$, we see that $\mu_{ij} = 0$ for $i \in n_{-j}$. But then by (8), $\sigma_j = 1$, a contradiction. Thus we must investigate conditions under which $0 < \sigma_j < 1$ for some j in equilibrium, requiring

$$g = s_j(\sigma_j). \tag{9}$$

We call such a situation a *working equilibrium*. Note that, since $s_j(\sigma_j)$ is an increasing function, the solution to (9) is unique.

We say agent i is *self-interested* if $\rho_i(1) = 0$. A sufficient condition for being self-interested is, of course, that $a_i = \lambda_i = 0$, so i's utility is a function of his material payoff alone. We say agent i is a *reciprocator at shirking level* σ if $\rho_i(\sigma) > 0$, and we say i is a *reciprocator* if i is a reciprocator for some level of shirking. We say reciprocators are *homogeneous* if all share the same parameters $a_i = a$, $\lambda_i = \lambda$, and the cost of punishing schedule $c(s) = c_i(s)$. If reciprocators are homogeneous, it is clear that they also have the same subjective gain from punishing schedule $\rho_i(\sigma)$, and the same punishment schedule $s(\sigma)$.

Consider a group in which a fraction $f(\sigma)$ of members are reciprocators at level σ. Notice that if j is not a reciprocator, j has $f(\sigma)n$ potential punishers, whereas if j is a reciprocator, j has $f(\sigma)n-1$ potential punishers. In the interest of simplicity of exposition, we will ignore this difference, assuming all agents face $f(\sigma)n$ potential punishers.[11] We have

Theorem 1 *Suppose reciprocators are homogeneous, and let $f(\sigma)$ be the fraction of reciprocators at shirking level σ.*

1. *If there are no reciprocators (i.e., $f(1) = 0$), there is a unique Nash equilibrium, in which $\sigma_j = \sigma^* = 1$ and $\mu_{ij} = \mu^* = 0$ for all $i, j = 1, \ldots, n$; i.e., all members shirk and no member punishes.*

2. *If there is a strictly positive fraction of reciprocators (i.e., $f(1) > 0$), then there is a unique $\sigma = \hat{\sigma} \in (0, 1)$ such that $\rho(\hat{\sigma}) = 0$ for all reciprocators.*

3. *If there is a strictly positive fraction of reciprocators and if $g > f(\hat{\sigma})ns(\hat{\sigma})$, where $\hat{\sigma}$ is as defined in (2), there is a Nash equilibrium in which $\sigma_j = \sigma^* = 1$ and $\mu_{ij} \in [0, 1]$ is arbitrary, for all $i, j = 1, \ldots, n$; i.e., all workers shirk and all reciprocators punish.*

4. *If there is a strictly positive fraction of reciprocators and if $g < f(\hat{\sigma})ns(\hat{\sigma})$, where $\hat{\sigma}$ is as defined in (2), then*

   (a) *there is a mixed strategy Nash equilibrium in which all members shirk at level $\hat{\sigma}$, and reciprocators punish each agent with probability*

$$\mu^* = \frac{g}{f(\hat{\sigma})ns(\hat{\sigma})}; \tag{10}$$

[11] The effect of dropping this assumption is in all cases quite transparent, since in effect, the model is the union of n independent games, in each of which one agent is the worker and the other $n-1$ agents are the punishers.

(b) an interior equilibrium remains interior when team size increases;

(c) the social welfare difference per team member between a first best world with no shirking and the equilibrium of this game is $\hat{\sigma}(q-1)$. This does not decline when team size increases.

The proof is straightforward.

The intuition behind Theorem 1 is as follows. The equilibrium level of shirking, $\hat{\sigma}$, equates the net benefits of punishing and trusting, while the equilibrium probability of punishing equates the net benefits of working and shirking. Thus when $\sigma > \hat{\sigma}$, the expected benefits of punishing, $\rho(\sigma)$ is positive. Members who punish with high probability will then receive higher payoffs than members who punish with low probability, inducing some to increase their punishing probability. As the punishing probability increases, the gains to shirking decline, leading members to reduce σ. This dynamic continues until $\sigma = \hat{\sigma}$. A similar dynamic occurs when σ is less than its equilibrium value. When μ is greater than its equilibrium value (10), the expected costs of shirking $f(\sigma)\mu ns(\sigma)$ exceed the benefits g. Members who do not shirk much will then be receiving higher payoffs than members who shirk a lot inducing some to decrease their level of shirking. As the shirking rate declines, the gains to punishing decline, leading to a reduction in μ. This dynamic will continue until (10) is satisfied. A similar dynamic occurs when μ is less than the equilibrium value given by (10).

Behavioral traits such as a work ethic or a willingness to punish co-members for inflicting harm on the team are, of course, strongly norm-governed and as such need not be proximately determined by the explicit optimization of any agent but rather may be the expression of behavioral rules. Thus the model underlying Theorem 1 may be interpreted as the basis of a dynamic treatment of work and punishment norms, with the updating of norms responding to the observed payoffs of others. For example, as our description of the intuition behind part (4a) of Theorem 1 suggests, the determination of σ and μ may be represented as dynamic processes based on the differential replication of norms governing the working, shirking, and punishing behaviors we have modeled, the equilibrium values σ^* and μ^* simply representing outcomes that are stationary in the underlying dynamic. We do not develop this extension here.

Now suppose there is no homogeneity. Then for each σ, a certain fraction of agents don't punish, a certain fraction punish with certainty, and for the isolated values of σ such that $\rho_i(\sigma) = 0$ for some i, $k > 0$ members are indifferent between punishing and not punishing. Let $\hat{\sigma}(\sigma)$ be the total punishment per shirker by agents who punish with probability one. Then total punishment is given by $s(\sigma) = \hat{\sigma}(\sigma) + \mu k(\sigma)s^{**}(\sigma)$, where $s^{**}(\sigma)$ is the amount of punishment inflicted by an agent i for which $\rho_i(\sigma) = 0$, or is zero if no such agent exists, and $k(\sigma)$ is the number of agents i for which $\rho_i(\sigma) = 0$.

It is easy to see that the function $s(\sigma)$ defined above, while not continuous, is increasing and has a continuous inverse. Therefore given g, as long as $f(1)ns(1) > g$, there is a unique $\hat{\sigma}$, a unique $\hat{\mu}$, and a well-defined set of agents C, $k \geq 0$ in number, such that (a) $g = s(\hat{\sigma}) + \hat{\mu}k(\hat{\sigma})s^{**}(\hat{\sigma})$; (b) $\rho_i(\hat{\sigma}) = 0$ for $i \in C$; (c) agents in C punish with probability $\hat{\mu}$, while all other agents j punish with probability zero or one, according as $\rho_j(\hat{\sigma})$ is negative or positive. This is the unique Nash equilibrium.

4 Experiments in Mutual Monitoring

Our model of mutual monitoring in teams embodies an essential behavioral assumption, namely that under some conditions strong reciprocity motives will induce sufficient punishment of free riding to sustain high levels of team output. To test the plausibility of this assumption, we conducted an experimental public goods game, extending the standard protocol by making each player's contribution to the public good known to all team members at the end of each round, and allowing players to punish other players based on this information, at a cost to themselves. Fehr and Gächter (2000) used a similar experimental setting to show that there is indeed a propensity to punish, and that allowing costly punishment in a multiperiod setting prevented the decay of cooperation usually found in public goods experiments (see Ledyard 1995).[12] In addition to replicating Fehr and Gächter, we investigate the motives for punishment and the subjects' response to punishment, as well as the effect of group size on the propensity to punish.[13]

We deliberately created an experimental environment in which contributions would be difficult to sustain by implementing the so-called *strangers treatment*, in which subjects are randomly reassigned to a new group at the beginning of each round of play.[14] We also make punishing shirkers costly; the cost of inflicting a penalty of two experimental monetary units, EMUs, is one EMU for the punisher.

Suppose there are n players, each player receives w EMUs at the beginning of each round, and player i contributes $w(1-\sigma_i)$ to the public good. These contributions are revealed to the players, who then can punish others at a cost of 1 EMU per sanction.[15] Let s_{ij} be the expenditure on sanctions assigned by player i to player j (we assume $s_{ii}=0$). Then the payoff to player i is

$$\pi_i = w\left[\sigma_i + nm(1-\bar{\sigma})\right] - \sum_{j=1}^{n} s_{ij} - 2\sum_{j=1}^{n} s_{ji}, \tag{11}$$

where $\bar{\sigma}=\sum_{j=1}^{n}\sigma_j/n$ is the average shirking rate, $\sum_j s_{ij}$ is player i's expenditure on sanctions and $2\sum_j s_{ji}$ is the reduction in i's payoffs due to the total sanctions received from the rest of the team. Note that we have set $\alpha=1$ and defined a new variable $m=q/n$, since we do not use the concept of residual claimancy in the experiment, but rather vary m, which is the per-member payoff to a player contribution of one EMU, and n, the team size. Also, the unit endowment in the model developed in the previous two sections is $w=25$ EMUs in our experiment.

To examine how subjects' contributions and punishment allocations are affected by team size and the degree of harm caused by shirking we used two group sizes (four and eight) and two values of m (0.3 and 0.7) allowing us to compare across treatments for similarities

[12] Other contributions to the literature on the ability of sanctions to control free riding in social dilemma settings include Barr (2001), Cinyabuguma et al. (2005), Masclet et al. (2003) and Walker and Halloran (2004).

[13] The instructions for a typical session appear in the Appendix.

[14] The more common *partners treatment*, in which groups remain together throughout the experiment tends to foster more cooperation than the stranger treatment Croson 1996, Fehr and Gächter 2000, Keser and van Winden 2000).

[15] The instructions to participants refer neutrally to "reductions," with no interpretation supplied.

in behavior based on the cost that shirkers inflict on their teammates. Our underlying behavioral assumptions concerning reciprocity imply that a team member's punishment of a teammate will vary both with the teammate's amount of shirking and the harm caused by a unit of shirking, the latter depending on the size of the group and the marginal per person return from contributions to the public account. There are two ways to measure the harm done by a shirking team member. The first, which we term the *private cost of shirking*, is the reduction in a given teammate's payoffs associated with an act of shirking by individual i or $mw\sigma_i$. By contrast, z_i the *social cost of shirking* by member i, takes account of the costs borne by every team member other than the shirker and is thus $(n-1)mw\sigma_i$.

A total of twelve sessions were conducted (three per treatment) with 172 participants. Figure 1 summarizes our experimental design.[16]

	Four-person teams	Eight-person teams
m=0.30	10 teams 40 subjects	6 teams 48 subjects
m=0.70	9 teams 36 subjects	6 teams 48 subjects

Figure 1: Experimental Design

Each session lasted ten periods. In each period (a) subjects were randomly reassigned to a group, given an endowment of $w = 25$ EMUs, and allowed to contribute, anonymously, any fraction of the endowment to a public account, the remainder going to the subject's private account; (b) the total group contribution, the subject's gross earnings, and the contributions of other team members (presented in random order) were revealed to each subject, who was then permitted to assign sanctions to others; (c) payoffs were calculated according to (11), and subjects were informed of their net payoffs for the period.

If we assume standard preferences for participants (i.e. each player cares only about his or her personal payoff) then punishing is not a best response because it is costly and can have only limited effect on future payoffs to the punisher, given the stranger treatment. Hence the unique subgame perfect Nash equilibrium of the game is that no one punishes and therefore no one contributes to the public good. To assess the validity of the behavioral assumptions on which our model founded we want to test the following hypotheses.

Hypothesis 1: Punishing Occurs in Response to Shirking. Punishment occurs in all periods and under all treatment conditions when $\sigma_i > 0$ for some i. Further, those who shirk more receive more punishment, $(\partial \sum s_{ji} / \partial \sigma_i > 0)$.

Hypothesis 2: The Level of Punishment Increases with the Level of Shirking and the Harm that Shirking Inflicts on Other Team Members. The level of punishment

[16] The number of participants, and therefore teams, per treatment vary due to no-shows. All subjects were recruited by email from the general student population and none had ever participated in a public goods experiment before. Each subject was given a five dollar show-up fee upon arrival and then was seated at a partially isolated computer terminal so that decisions were made in privacy. Each session took approximately 45 minutes from sign-in to payments and subjects earned \$20.58 on average, including the show-up fee.

directed toward player i increases with (a) the cost imposed on individual punishers, $wm\sigma_i$, and (b) the cost imposed on all other team members, $wm(n-1)\sigma_i$.

Hypothesis 3: Shirkers Respond to Punishment. Punishment in one round leads shirkers to increase their contributions in subsequent rounds.

Hypothesis 4: Punishment Fosters Cooperation. The overall level of cooperation does not decay when costly punishment is permitted.

Hypothesis 5: Altruism Does Not Provide and Adequate Explanation of Punishment.

Hypotheses 1 and 4 replicate Fehr and Gächter (2000). Hypothesis 2 tests the assertions of the previous section, since it implies that the tendency to punish increases with the team's residual claim, and does not decline with increasing group size. Hypotheses 3 and 4 are important for the welfare implications of mutual monitoring, and Hypothesis 5 tests an alternative explanation for punishment.

The results of our experiments confirm most but not all of the above hypotheses. Average contributions are higher in our mutual monitoring treatments than in the standard voluntary contribution game, and most importantly, they do not decline over time.[17] The reason behind sustained levels of cooperation in our experiment is punishment. Overall, 89% of the participants punished another participant at least once.[18]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All periods				First 5 periods	Last 5 periods	Last period
Constant	-3.68***	-5.19***	-2.77***	-2.81***	-1.50***	-3.64***	-1.99**
	(0.39)	(0.73)	(0.36)	(0.47)	(0.56)	(0.61)	(1.02)
Shirking level ($w\sigma_i$)	0.38***	0.39***					
	(0.02)	(0.02)					
Residual claim (m)		2.85***					
		(1.16)					
Private cost imposed			0.66***	0.66***	0.56***	0.67***	0.32***
by shirker ($wm\sigma_i$)			(0.04)	(0.05)	(0.06)	(0.06)	(0.11)
Team size				0.08	-0.30	0.42	1.14
				(0.49)	(0.56)	(0.61)	(1.03)
Log likelihood	-3671	-3668	-3684	-3684	-1954	-1731	-355

Table 1: Why are Shirkers Punished? All regressions are Tobits and (except (7)) include random effects. The dependent variable is the total punishment received (standard errors of the estimates are in parentheses). *indicates significant at 0.10, ** 0.05, ***0.01 level.

[17] This assertion is based on comparing average contributions in the current experiment to the contributions in other strangers experiments including Croson (1996) and Keser and van Winden (2000).

[18] The frequency of punishers is high in all treatments. 98% punished in the Low m, Small treatment; 88% punished in the Low m, Large treatment; 81% punished in the High m, Small treatment; and 90% punished in the High m, Large treatment.

Because our hypotheses concern how behavior changes over time as players learn more about the consequences of their actions, we used the panel nature of our data to estimate a number of the implied learning models. We used random effects methods to control for cross-sectional differences and any effects of experience. A summary of our analysis is presented in Tables 1 and 2.

Concerning Hypotheses 1, equation (1) of Table 1 shows the results of a regression analysis of punishment decisions, where the dependent variable is the total expenditure on punishment by the other team members and the independent variable is individual i's level of shirking, $w\sigma_i$. We see that shirking triggers punishment and therefore we can not reject Hypothesis 1, team members react to shirking by paying to punish free riders.

Hypothesis 2 states that the propensity to punish is increasing in the cost shirkers impose on individual punishers, $wm\sigma_i$, and in the social cost of shirking, $wm(n-1)\sigma_i$. This means that the propensity to punish should be greater for the High m treatment than for the Low (part a), greater for the Large than the Small (part b), and should be directed toward subjects who shirk more rather than less (both parts). We use equations (2) - (4) in Table 1 to assess these hypotheses. From equation (1) we know that shirking incites retribution, but we are uncertain whether team members react to the costs of shirking imposed on them and others. Equation (2) adds m to the analysis and demonstrates that members of groups in which a unit of shirking imposes high costs tend to punish more. Equation (2) establishes that punishers are responding to the degree of harm done by the shirker, not simply responding to the act of shirking *per se*. Equation (3) confirms that the private cost imposed on team members is a motive for punishment. Finally, (4) adds a dummy variable for Large teams. The effect of team size is positive, small, and insignificant, implying that the harm done to others is not what is motivating the punishment of shirkers. In columns (5) through (7) we test to see if our results depend on the fact that we are pooling our observations. In (5) we limit attention to the first five rounds and in (6) we look only at the last five rounds. As one can see, the coefficient on the $wm\sigma_i$ is similar for the two time periods (the difference is not significant, $p = 0.65$); however, while free riders are still punished in the last round, the motivation is lower (see (7)). Thus part (a) of Hypothesis 2 is not rejected, while part (b) is rejected.

Does shirking pay once the costs of punishment are considered? To answer we ran equation (1) separately for each treatment.[19] Note that the act of shirking deprives the shirker of the returns from the group project, so the net benefit of shirking in the absence of punishment is just $1-m$. In the Low m, Small treatment the estimated punishment induced by a unit of private allocation is 0.53 which means shirking pays quite well $(0.7-0.53)$. In the Low m, Large treatment, punishment induced by shirking declines to 0.14 implying shirking results in an even higher net return $(0.7-0.14)$. In the other two treatments shirking actually does not pay. For High m, Small treatment the net payoff is negative but close to zero (-$0.11 = 0.3 - 0.41$) and in High m, Large shirkers do considerably worse $(0.3 - 0.62)$. These differences in the net return to shirking help explain differences in contributions between our four treatments.

Hypothesis 3 asserts that being punished leads shirkers to increase their next-round

[19] In each case the coefficient on the punished player's private allocation was significant at the $p < 0.01$ level. Further, a Chow test suggests there are structural differences in our treatments ($\chi^2 = 73.54$, $p < 0.01$).

contribution. We test this by regressing subjects' public contributions in period t on the punishment points received by the shirker in period $t-1$ and other variables. The results are shown in Table 2. In equation (1) we see that the coefficient on the punishment received last period term is positive, and significant and that there is a lot of inertia in contributions demonstrated by the fact that a player's contribution in period $t-1$ is a strong, robust predictor of what he or she will do in period t. We also see that how much one deviates from the average contribution matters significantly: subjects move toward the mean.

	(1)	(2)	(3)
Constant	2.90***	3.48***	6.24***
	(0.88)	(0.93)	(1.65)
Lag Contribution	0.80***	0.81***	0.76***
	(0.06)	(0.06)	(0.06)
Lag Punishment	0.14***	0.03	0.03
	(0.04)	(0.08)	(0.08)
Deviation from average contribution	0.54***		
	(0.06)		
Shirker deviation from average		0.55***	0.51***
		(0.09)	(0.09)
Contributor deviation from average		-0.27***	-0.22**
		(0.09)	(0.10)
Shirker dev × Lag Pun		0.05***	0.05***
		(0.02)	(0.02)
Contributor dev × Lag Pun		-0.03***	-0.03***
		(0.01)	(0.01)
Cost of contributing (1-m)			-4.92***
			(1.65)
Number of other teammates (n-1)			0.07
			(0.16)
Log likelihood	-4561	-4545	-4540

Table 2: How do Participants Respond to Punishment? All regressions are Tobits and include random effects. The dependent variable is one's contribution in round t(standard errors of the estimates are in parentheses). *indicates significant at 0.10, ** 0.05, ***0.01 level.

To explore whether mean seeking behavior is symmetrical, we define a free-rider in a given round to be a player who contributed less than the team average in that round. We call "contributors" those players who contribute more than the average (and therefore have negative deviations). We hypothesize that contributors and free-riders respond differently to punishment. Because the experimental instructions refer to "contributions to a group project," it is soon clear to participants why shirkers are punished, but when it occurs punishment is probably confusing for contributors. To study possible differences in responses to punishment, equation (2) separates shirkers from contributors and interacts lagged punishment with players' deviations from the average to test whether punishment, *per se*, matters or if punishment only matters when one deviates from the norm. Equation (2) confirms our suspicions. We see that for both free riders and contributors movement towards the mean is increased by punishment and the effect of punishment is greater the farther one is from the mean. The behavior of free riders and contributors is different. In the case of contributors there is little regression to the mean unless one is punished, while free riders move towards the mean even if not punished, but more strongly so when they are. Consistent with the experimental results of Hopfensitz and Reuben (2006), a reasonable interpretation is that contributors respond spitefully to being punished.[20]

Equation (3) examines the effect on contributions of m, the per person benefit of contributing, and the size of teams. If each team member benefits by m from a contribution, then $1-m$ is the net cost of contributing to the group project. We see that our participants respond significantly to the cost of contributing, reducing contributions when the cost rises. This result could arise because team members are motivated to help others in the group and the larger is m the greater good will their contribution do. But this is not the case. Controlling for inertia, punishment, position in the distribution of contributions, and cost of contributing, we find that subjects neither contribute more nor less in larger groups.

Does punishment of free riders work? We know from Table 2 that free riders respond to punishment, but do they respond enough to offset the costs of punishing? The thought experiment we will consider is to compare a situation in which one EMU has been devoted to punishment of a representative free-rider and another in which that one EMU was instead contributed to the public account. Suppose the representative free-rider deviates from the mean by the average of free-riders, 5.08 EMUs. Using equation (3) from Table 2, a unit of punishment would raise his contributions next period by $0.05(5.08) = 0.254$. This is just more than half of the cost to the punisher to deploy the sanction, since inflicting a punishment of one unit costs 0.5 units. But this short term effect would be carried over into subsequent periods by the inertia in contributions, giving a long term effect of 1.06 which more than twice the cost of inflicting the punishment.[21] We conclude that punishment effects subsequent contributions, and when directed towards egregious free-riders it is cost effective in the social sense (In the private sense it is not cost effective because in the strangers treatment the likelihood of being paired with the person one punishes is low). Punishing those close to the mean or above the mean is not socially cost effective, however. The estimated response of a free rider who was 2.40 below the mean would just offset the cost of punishment. Punishment of less serious free riders does not pay.

Hypothesis 4 asserts that the level of contribution does not decay when punishment is permitted. We would like to address the following questions concerning the temporal pattern of behavior: (i) is there a decline over time in contributions; (ii) is there a decline in contributions conditional on the individual's experience in the most recent period of the game; (iii) does punishment decline conditioned on the amount of shirking occurring; (iv) do subjects exhibit an end game effect with respect to either punishing or giving?

[20] Although contributors give more than the average, they do not always contribute fully. Therefore, some contributors may have interpreted punishment as being sanctioned for not fully contributing. However, the negative coefficient on the contributor interaction term is consistent with our resentment explanation.

[21] Because the relationship between current contributions and future contributions is strongly positive, punishing to increasing a free-rider's contributions by 0.254 next period will also increase his contributions by $0.76(0.254) = 0.19$ in the succeeding period. The limit of this punishment multiplier is $0.254/(1-0.76) = 1.06$.

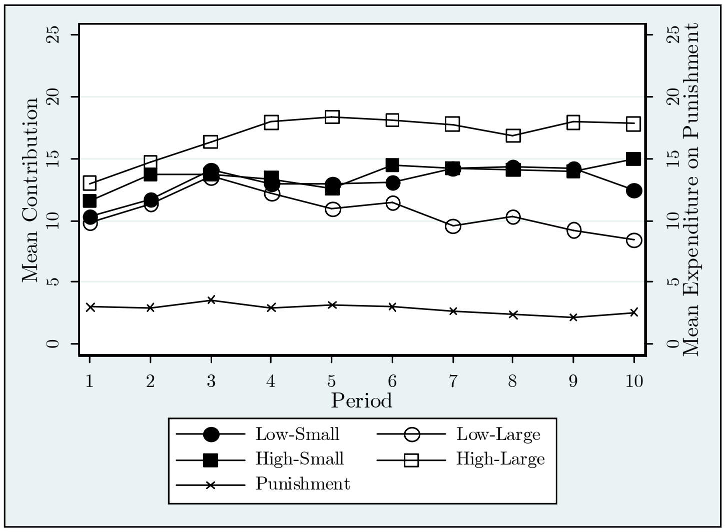

Figure 2: Mean Behavior by Treatment and Round (Note: Low or High refers to the team's residual claim, Small or Large refers to the team size and Punishment refers to the mean expenditure on puinshment pooled across treatments).

Figure 2 addresses the first question about the pattern of contributions. Clearly the aggregated contributions in the current experiment do not exhibit the standard decline (Ledyard 1995). Average contributions pooled across treatments start at 45% of the endowment in period one, rise slightly, level off, and end at 54% in period ten. In response to the second question, we ran equation (3) in Table 2 with period dummies and found, controlling for punishment, contributions are significantly higher only in periods two and three, but periods four through nine are no different from period ten. The third question asks whether punishment declines with experience. Figure 2 suggests that punishment does not decline (even in period 10) but to be sure we re-ran equation (3) from Table 1 with period dummies. Here we find that punishment is significantly higher only in period three. Punishment does not decline between periods four and ten. These additional regressions also demonstrate that there is no end-game effect in either punishment or contributions (i.e., comparing periods nine and ten). In sum, we conclude that contributions increase initially and are maintained at significant levels. Further, this is largely due to the punishment imposed on free riders by reciprocating contributors.

Finally, Hypothesis 5 asserts that players punish shirkers whether or not punishment leads to higher group earnings, and cannot be accounted for as an instrumental strategy used by altruists who want to increase the payoffs to others in the team. Were punishment such an altruistic act, punishers would both contribute more in larger groups (because for a given m more benefits to others are distributed in large groups) and punish more in large groups (because if successful in inducing the free-rider to contribute more punishment would generate more aggregate benefits.) The fact that group size has no effect on either punishment (Table 1) or contributions (Table 2) suggests that while altruism towards other team members may be involved it is not sufficient to explain the high levels of punishment of free-riders. Instead we conclude that punishment *per se* is a motive.[22]

[22] There is also no point in punishing at the end of the last period if one is an altruist.

5 Impediments to Mutual Monitoring: Team Size and 'Shirking Cliques'

Would it be plausible to treat ρ as a decreasing function of team size, on the grounds that strong reciprocity may weaken when the team becomes larger, and thus the propensity to punish any given act of shirking would fall? Though this is possible, there is to our knowledge no clear evidence in support of this notion, and there are many 'stylized facts' contradicting it. For instance, citizens of large nations appear no less willing to sacrifice for their compatriots than those of small nations. Similarly, people are often observed to support their local sports team, their regional sports team, and their national sports team with equal commitment. Examples of this type abound. The experimental evidence described above, moreover, suggests that ρ does not depend on team size.

There are of course additional paths through which increasing team size might weaken the mutual monitoring mechanism. Increased n might lower the average cost s a monitor can impose on a shirker, since the 'average social distance' between a pair of workers can be expected to increase as the team becomes more numerous. The ability to detect shirking among randomly selected pairs of team members may also decline. Notice, however, that adding team members also increases the number of potential punishers so it might well increase the likelihood that any given act of shirking would be detected, and hence might increase the amount of punishment which a shirker would expect (Carpenter 2004).

The intuition behind Theorem 1 is that as long as increasing team size does not reduce the number of team members that one may see the effectiveness of mutual monitoring does not decline as team size increases. The proof is as in Theorem 1, with minor changes. Notice, however, that the effectiveness of mutual monitoring depends on the extent to which members may be informed about one another's activities: if team members are able to see relatively few co-members, then the condition $fs > g$ is not likely to obtain, and universal shirking will be the equilibrium.

However large teams often do not have the informational homogeneity assumed in Theorem 1, since with increased team size often comes a more refined division of labor in which there are specialized 'work groups' whose members all see one another, and who are not seen by other team members. If members of such a group have an incentive to make credible commitments involving the reciprocators in the group not punishing, so all members of the group can shirk without penalty, we call the group a *shirking clique.* We have

Corollary 2 *Suppose there is a subgroup C of members that is isolated—members of C see one another but are not seen by non-C members. Then if the conditions of Theorem 1 hold for C—in particular, if the frequency of reciprocators f in C is the same as in the team—C is not a shirking clique.*

This result follows from the observation that Theorem 1 holds, and this theorem assumes nothing concerning the pattern of seeing. The point is that a group that sees only its own members cannot agree that the reciprocators will not punish, since there is no way to enforce such an agreement within the model and reciprocators would benefit from violating the agreement.

Yet shirking cliques may exist under less restrictive conditions. We will mention, but not develop formally, two possibilities. First, if group composition is not random with respect to the frequency of reciprocators, it is clear from Theorem 1 that an isolated group may have a sufficiently low frequency of reciprocators that it will form a shirking clique. The critical frequency of reciprocators f_c for an isolated group of size k is

$$f_c = \frac{g}{ks},$$

indicating that shirking cliques may be more difficult to sustain in larger groups.

A second type of shirking clique emerges if we admit asymmetric information and collusion. Suppose a work group is isolated and colludes to impose costs on reciprocators, making it unattractive to punish. Then the group can form a shirking clique. Even a non-isolated group whose members collude to punish reciprocators may constitute a shirking clique as long as there are not too many reciprocators among non-clique members who can see the members of the group. Indeed, our experimental data indicate that shirkers do occasionally punish reciprocators - the average contribution of a person punished by a shirker (i.e., someone who contributed nothing) is 12.56 EMUs.

6 Conclusion

Our model and experimental results suggest that under appropriate conditions, strong reciprocity can support mutual monitoring even in large teams, unless the frequency of reciprocators is too low or the technology and organization of the production process favors the formation of shirking cliques. Furthermore, mutual monitoring allows levels of member effort that are closer to first best, thus enhancing team welfare.

The case we have modeled appears to describe a production team in which the noncontractible action is work effort. But the model may equally depict a range of analogous problems. The 'team' might be composed of family, neighbors and friends engaged in informal insurance to supplement market-supplied insurance as in Stiglitz (1993) and Arnott(1991), or members of an informal borrowing group where team members borrow from a financial institution with a renewal of credit being contingent on all members repaying at the end of the first round. Both cases conform to the assumptions of the above model, namely, superior information held by team members, combined with interdependence of members welfare on other members' actions, and opportunities to punish members who impose costs on others in the team.

Another application is to residential home owners in which the team consists of neighbors whose residential amenities, and hence the value of their housing assets, are affected by the noncontractual actions of others in the neighborhood. Sampson et al. (1997) provide empirical evidence of such mutual monitoring in neighborhoods. In this case, monitoring and punishment may consist of admonitions favoring anything from maintaining the appearance of one's property to joining in collective actions to gain safer streets or better schools for the neighborhood. Finally, we think the model may illuminate a characteristic of the foraging bands which constituted human society during most of its history, namely, widespread hunting, foraging, and food sharing, and punishment of those who violated the underlying reciprocity norms (Woodburn 1982, Knauft 1991, Boehm 1993, Bowles and Gintis 2003).

Given the apparently widespread nature of the problems of non-contractibility which it addresses and the welfare benefits it may make possible, it may be wondered why mutual monitoring is not ubiquitous in modern economies. A reason suggested by this model is that residual claimancy by team members is essential to the underlying monitoring motivations, and for many of the relevant production teams the fact that members are asset poor effectively precludes assignment of any but trivial levels of residual claimancy to team members. Transferring residual claimancy over the income streams of an asset but not ownership itself to team members creates incentives for the team to depreciate the assets, the costs of which may more than offset any gains from mutual monitoring. Thus prohibitive costs may arise if residual claimancy is separated from ownership, and outright ownership may be precluded by borrowing limitations and possibly high levels of risk aversion characteristic of low wealth team members.

The role of residual claimancy in motivating mutual monitoring thus provides another case in which differing distributions of wealth may support differing equilibrium distributions of contracts and systems of governance. Other cases include forms of agricultural and residential tenancy (Laffont and Matoussi 1995) access to self employment and human investment (Galor and Zeira 1993, Loury 1981, Blanchflower and Oswald 1998, Black et al. 1996), and the extent of cooperative forms of ownership of team assets (Legros and Newman 1996).[23] Because a given distribution of contracts and incentives may be constrained Pareto-optimal under some distributions of wealth but not under others, a particular distribution of wealth may preclude the evolution of allocationally superior systems of contract and incentives. Thus the assertion that the assignment of residual claimancy and control rights in market economies may be deduced from considerations of allocative efficiency is not generally valid.[24]

7 References

Alchian, A. and H. Demsetz (1972) Production, Information Costs, and Economic Organization. American Economic Review 62(December), pp. 777-95.

Arnott, R. (1991) Moral Hazard and Nonmarket Institutions. American Economic Review 81(1), pp. 180-190.

Banarjee, A., T. Besley, et al. (1994) Thy Neighbor's Keeper: The Design of a Credit Cooperative with Theory and a Test. Quarterly Journal of Economics 109(May), pp. 491-515.

Bardhan, P., S. Bowles, et al. (2000) Wealth Inequality, Wealth Constraints and Economic Performance. in: A. Atkinson and F. Bourguignon Handbook on Income Distribution, (Dortrecht: North Holland).

Barr, A. (2001) Social Dilemmas and Shame-Based Sanctions: Experimental Results from Rural Zimbabwe. Center for the Study of African Economies Working Paper. WPS2001.11.

Besley, T. and S. Coate (1995) Group Lending, Repayment Incentives and Social Collateral. Journal of Development Economics 46, pp. 1-18.

[23] We survey these cases in Bardhan et al. (2000).

[24] This presumption is expressed in Alchian and Demsetz (1972), Kihlstrom and Laffont (1979), Grossman and Hart (1986), Holmström and Tirole (1988) and Hart and Moore (1990).

Black, J., D. de Meza, et al. (1996) House Prices, the Supply of Collateral and the Enterprise Economy. The Economic Journal 106, pp. 60-75.

Blanchflower, D. and A. Oswald (1998) What Makes a Young Entrepreneur? Journal of Labor Economics 16(1), pp. 26-60.

Blount, S. (1995) When Social Outcomes Aren't Fair: The Effect of Causal Attribution on Preferences. Organizational Behavior & Human Decision Processes 62(2), pp. 131-44.

Boehm, C. (1993) Egalitarian Behavior and Reverse Dominance Hierarchy. Current Anthropology 34(3), pp. 227-40.

Bowles, S. and H. Gintis (2004) The Evolution of Strong Reciprocity. Theoretical Population Biology 65, pp. 17-28.

Campbell, D. T. (1983) Two Distinct Routes beyond Kin Selection to Ultra-Society: Implications for the Humanities and Social Sciences. in: D. Bridgeman The Nature of Prosocial Development, (New York: Academic Press), pp. 11-41.

Carpenter, J. (2004) Punishing Free-Riders: how group size affects mutual monitoring and the provision of public goods. Games and Economic Behavior, forthcoming.

Carpenter, J. and P. Matthews (2005) Norm Enforcement: Anger, Indignation, or Reciprocity. Department of Economics, Middlebury College, Working Paper 0503.

Cinyabuguma, M., T. Page, et al. (2005) Cooperation under the Threat of Expulsion in a Public Goods Experiment. Journal of Public Economics 89(8), pp. 1421-35.

Craig, B. and J. Pencavel (1995) Participation and Productivity: A Comparison of Worker Cooperatives and Conventional Firms in the Plywood Industry. Brookings Papers: Microeconomics, pp. 121-160.

Croson, R. (1996) Partners and Strangers Revisited. Economics Letters 53, pp. 25-32.

de Quervain, D., U. Fischbacher, et al. (2004) The Neural Basis for Altruistic Punishment. Science 305(27 August), pp. 1254-1258.

Dong, X.-y. and G. Dow (1993a) Does Free Exit Reduce Shirking in Production Teams? Journal of Comparative Economics 17, pp. 472-484.

Dong, X.-y. and G. Dow (1993b) Monitoring Costs in Chinese Agricultural Teams. Journal of Political Economy 101(3), pp. 539-553.

Fehr, E. and S. Gaechter (2000) Cooperation and Punishment in Public Goods Experiments. American Economic Review 90(4), pp. 980-994.

Fehr, E., S. Gaetcher, et al. (1997) Reciprocity as a Contract Enforcement Device. Econometrica 65(4), pp. 833-60.

Fehr, E. and K. Schmidt (1999) A Theory of Fairness, Competition, and Cooperation. Quarterly Journal of Economics 114(3), pp. 769-816.

Frohlich, N., J. Godard, et al. (1998) Employee vs. Conventionally Owned and Controlled Firms: an experimental analysis. Managerial and Decision Economics 19, pp. 311-326.

Fudenberg, D., D. Levine, et al. (1994) The Folk Theorem with Imperfect Public Information. Econometrica 62, pp. 997-1039.

Fudenberg, D. and E. Maskin (1986) Games with Discounting or with Incomplete Information. Econometrica 54(3), pp. 533-554.

Ghemawat, P. (1995) Competitive Advantage and Internal Organization: Nucor Revisited. Journal of Economics and Management Strategy 3(4), pp. 685-717.

Glaeser, E. and D. DiPasquale (1999) Incentives and Social Capital: Are Homeowners Better Citizens? Journal of Urban Economics 45(2), pp. 354-384.

Greenberg, E. (1986) Workplace Democracy: The Political Effects of Participation. (Ithaca: Cornell University Press).

Grossman, S. and O. Hart (1986) The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. Journal of Political Economy 4(691-719).

Hansen, D. (1997) Worker Performance and Group Incentives: a case study. Industrial and Labour Relations Review 51(1), pp. 37-49.

Hart, O. and J. Moore (1990) Property Rights and the Nature of the Firm. Journal of Political Economy 98(6), pp. 1119-1158.

Holmstrom, B. (1982) Moral Hazard in Teams. Bell Journal of Economics 13, pp. 324-40.

Holmstrom, B. and J. Tirole (1989) The Theory of the Firm. in: R. Willig Handbook of Industrial Organization, (Amsterdam: North-Holland), pp. 61-133.

Hopfensitz, A. and E. Reuben (2006) The importance of emotions for the effectiveness of social punishment. University of Amsterdam, CREED Working Paper.

Kandel, E. and E. Lazear (1992) Peer Pressure and Partnerships. Journal of Political Economy 100(4), pp. 801-17.

Keser, C. and F. van Winden (2000) Conditional Cooperation and Voluntary Contributions to Public Goods. Scandinavian Journal of Economics 102(1), pp. 23-29.

Kihlstrom, r. and J.-J. Laffont (1979) A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion. Journal of Political Economy 87(4), pp. 719-748.

Knauft, B. (1991) Violence and Sociality in Human Evolution. Current Anthropology 32(4), pp. 391-409.

Knez, M. and D. Simester (2001) Firm-wide Incentives and Mutual Monitoring as Continental Airlines. Journal of Labor Economics 19(4), pp. 743-772.

Laffont, J.-J. and M. S. Matoussi (1995) Moral Hazard, Financial Constraints and Share Cropping in El Oulja. Review of Economic Studies 62, pp. 381-399.

Ledyard, J. (1995) Public Goods: a survey of experimental research. in: J. Kagel and A. Roth The Handbook of Experimental Economics, (Princeton: Princeton University Press), pp. 111-94.

Legros, P. and A. Newman (1996) Wealth Effects, Distribution, and the Theory of Organization. Journal of Economic Theory August.

Levine, D. (1997) Modeling Altruism and Spitefulness in Experiments. Review of Economic Dynamics 1(3), pp. 593-622.

Loury, G. (1981) Intergenerational Transfers and the Distribution of Earnings. Econometrica 49, pp. 843-867.

Masclet, D., C. Noussair, et al. (2003) Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism. American Economic Review 93(1), pp. 366-380.

Ostrom, E. (1990) Governing the Commons: the evolution of institutions for collective action. (Cambridge: Cambridge University Press).

Ostrom, E., J. Walker, et al. (1992) Covenants With and Without a Sword: self-governance is possible. American Political Science Review 86, pp. 404-17.

Rabin, M. (1993) Incorporating Fairness into Game Theory and Economics. American Economic Review 83(5 December), pp. 1281-1302.

Rotemberg, J. J. (1994) Human Relations in the Workplace. Journal of Political Economy 102(4), pp. 684-717.

Roth, A. (1995) Bargaining Experiments. in: J. Kagel and A. Roth The Handbook of Experimental Economics, (Princeton: Princeton University Press), pp. 253-348.

Sampson, R., S. Raudenbush, et al. (1997) Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy. Science 277(August 15), pp. 918-924.

Stiglitz, J. (1993) Peer Monitoring and Credit Markets. in: K. Hoff, A. Braverman and J. Stiglitz The Economics of Rural Organization: Theory, Practice, and Policy, (New York: Oxford University Press), pp. 70-85.

Varian, H. (1990) Monitoring Agents with Other Agents. Journal of Institutional and Theoretical Economics 46(1), pp. 153-174.

Verba, S., K. Schlozman, et al. (1995) Voice and Equality: Civic Voluntarism in American Politics. (Cambridge: Harvard University Press).

Walker, J. and M. Halloran (2004) Rewards and Sanctions and the Provision of Public Goods in One-Shot Settings. Experimental Economics 7(3), pp. 235-247.

Weissing, F. and E. Ostrom (1991) Irrigation Institutions and The Games Irrigators Play: Rule Enforcement Without Guards. in: R. Selten Game Equilibrium Models II: Methods, Morals, and Markets, (Berlin: Springer), pp. 188-262.

Woodburn, J. (1982) Egalitarian Societies. Man 17(3), pp. 431-451.

8 Appendix - Participant instructions

You have been asked to participate in an economics experiment. For participating today and being on time you have been paid $5. You may earn an additional amount of money depending on your decisions in the experiment. This money will be paid to you, in cash, at the end of the experiment. By clicking the BEGIN button you will be asked for some personal information. After everyone enters this information we will start the instructions for the experiment.

Please be patient while others finish entering their personal information. The instructions will begin shortly.

During the experiment we will speak in terms of Experimental Francs instead of Dollars. Your payoffs will be calculated in terms of Francs and then translated at the end of the experiment into dollars at the following rate: 30 Francs = 1 Dollar.

Each participant receives a lump sum payment of 15 Francs at the beginning of the experiment (on top of the $5.00 show-up payment). This one-time payment may be used to offset any losses that are incurred during the experiment. However, it should be noted that you can ALWAYS avoid losses through your own decisions.

The experiment is divided into 10 different periods. In each period participants are divided into groups of 5. You will therefore be in a group with 4 other participants. The composition of the groups will change randomly at the beginning of each period. Therefore, in each period your group will consist of different participants.

Each period of the experiment consists of two stages. In the first stage you will decide how many francs you want to invest in each of two investment accounts. One account is a Private Account, which only you benefit from. The second account is a Public Account, the benefits of which are shared equally by all members of your group. In the second stage of the period you will be shown the investment behavior of the other members of your group.

You can then decide whether you want to distribute points to members of your group. If you distribute points to other members of your group, their earnings will be reduced.

Now we will explain the two stages in more depth.

Stage One

At the beginning of every period each participant receives and endowment of 20 francs. You have to decide how much of this endowment you want to invest in each of the two accounts mentioned above. You are asked to invest in whole franc amounts (i.e. an investment of 5 francs is alright, but 3.75 should be rounded up to 4).

To record your investment decision, you will type the amount of francs you want to invest in the Public and/or the Private account by typing in the appropriate text-input box which will be yellow. Once you have made your decision, there will be a green Submit button that will record your investment decision.

After all the members of your group have made their decisions, each of you will be informed of your Gross Earnings for the period.

Your Gross Earnings will consist of two parts:

1. Your return on your Private Account. Your Private Account returns 1 franc for each franc invested. That is, for each franc invested in the Private Account you get 1 franc back.

2. Your return from the Public Account. Your earnings (and everyone else's in your group) is equal to 0.3 times the total investment by all members of the group to the Public Account.

Your Earnings can be summarized as follows:

$1\ \times$(Investment in Private Account) + $.3\times$(Group Total Investment in Public Account)

The income of each group member from the Public Account is calculated the same way. This means that each group member receives the same amount from the total investment in the Public Account. For example, consider the case of groups with 5 members, if the total investment in the Public Account is 75 francs (e.g. first group member invests 15 francs, the second 20, the third 10 and the fourth and fifth 15 each) then each group member will receive $.3 \times 75 = 22.5$ francs. If the total investment was 30 francs then each group member would receive $.3 \times 30 = 9$ francs.

For each franc you invest in the Private Account you get 1 franc back. Suppose however you invested this franc in the Public Account instead. Your income from the Public Account would increase by $.3 \times 1 = .3$ francs. At the same time the earnings of the other members of your group would also increase by .3 francs, so the total increase in the group's earnings would be 1.5 francs. Your investment in the Public Account therefore increases the earnings of the other group members. On the other hand your earnings increase for every franc that the other members of your group invest in the Public Account. For each franc invested by another group member you earn $.3 \times 1 = .3$ francs.

Stage Two

In stage two you will be shown the investment decisions made by other members of your group and they will see your decision. Also at this stage you can reduce the earnings made by other member of your group, if you want to. You will be shown how much EACH member of your group invested in both the Public and Private Accounts. Your investment decision will also appear on the screen and will be labeled as 'YOU.' Please remember that the composition of your group will change at the beginning of each period and therefore you will not be looking at the same people all the time.

You must now decide how many points (if any) you wish to give to each of the other member of your group. You distribute points by typing them into the input-text box that appears below the investment decision of each of the other group members.

You will have to pay a cost to distribute points to other group members. This cost increases as you distribute more points to another participant. You can distribute between 0 and 10 points to each other member of your group. Your total cost of distributing points is the sum of all the costs you incur for distributing points to each of the other group members. The following table illustrates the relationship between the points distributed to each group member and the costs of doing so in francs.

Points:	0	1	2	3	4	5	6	7	8	9	10
Cost of Points (in francs):	0	1	2	4	6	9	12	16	20	25	30

Consider the case where there are 5 people per group. Suppose you assign 2 points to a group member. This costs you 2 francs. If you assign 9 points to another group member, it will cost you 25 francs and if you assign 0 points to the rest of the members of your group, you do not incur any cost. In this case your Total Cost of assigning points is $(2+25+0+0)$ or 27 francs. At any time you will be able to calculate your total cost of distributing points by clicking the orange Calculate Cost button that will appear on the screen. When you have finished distributing points you will click the blue Done button.

If you assign 0 points to a particular group member you do not change his or her earnings. However, for each point you assign to a group member, you reduce his or her Gross Earnings in the current period by 10 percent. Hence, if you assign one group member 2 points, his or her Gross Earnings for the period will be reduced by 20%. Assigning 4 points reduces Gross Earnings by 40% etc.

How much a participant's earnings from the first stage are reduced is determined by the Total amount of points he or she receives from all the other group members. If a participant receives a total of 3 points (from all the other group members in the current period) then his or her Gross Earnings would be reduced by 30 percent. If someone is assigned 4 points in total his or her Gross Earnings would be reduced by 40 percent. If anybody is assigned 10 or more points their Gross Earnings will be reduced by 100 percent. In this case the Gross Earnings of this person would be 0 francs for the current period.

For example, if a participant had Gross Earnings of 30 francs from the first stage and was assigned 3 points in the second stage, then his or her earnings would be reduced to $30 - (.3 \times 30) = 30 - 9 = 21$ francs.

In general, your earnings after the second stage will be calculated as follows:

Total Earnings at the end of the Second Stage:

1. If you received fewer than 10 points then Total Earnings equal

(Gross Earnings from Stage One)-[Gross Earnings?(.1×received points)]−(the cost of points you distributed)

2. If you receive 10 or more points then Total Earnings equal −(the cost of the points you distributed)

Please note that your earnings at the end of the second stage can be negative, if the cost of the points you distribute exceeds your (possibly reduced) earnings from stage one. However, you can avoid such losses by the decisions you make.

After all participants have made their decisions in the second stage, your final earnings for the period will be displayed in a manner similar to what follows:

Earnings Screen at the end of the Period

Your Gross Profits in the Current Period: The Total Cost of the Points You Assigned to Others: Number of Points Assigned to You by Others: Reduction of Gross Profit due to Points Assigned to You: % Current Period Payoff after Subtractions: Your Accumulated Earnings Including this Period:

When you have finished reviewing your earnings for the current period you will click the orange Proceed to Next Period button and wait for others to finish. When everyone is done, the experiment will proceed to the next period starting with stage one.

If you have any questions please raise your hand. Otherwise, click the red Finished button when you are done reading.

This is the end of the instructions. Be patient while everyone finishes reading.