Peer Discipline and the Strength of Organizations

David K. Levine and Salvatore Modica

The Issue

- groups do not act as individuals
- Olson and other have emphasized: incentives within groups matter
- how does internal group discipline work and what are the consequences?
- introduce a model of costly peer punishment
- homogeneous group and abstract from the issue of coordination failure
- focus on minimizing the cost to the group of enforcing particular actions
- measure the strength of the group as ability raise funds to provide a public good
- dependence on size of the group and size of the prize

Conclusions

- public goods problem are not important
- fixed costs per member due to peer punishment are
- when the overall stakes are small, small groups are more effective than large groups
- if small groups are too greedy in their demands, they will lose in competition with larger groups

The Model: Initial Round

- $N > 2$ identical players $i = 1, \ldots, N$ in group
- initial round 0: group members choose primitive actions $a^i \in A$ representing production decisions and the like
- $a^R \in A$ action of a representative member of the group
- consequence of the primitive actions of group members are binary signal of individual behavior $z_0^i \in \{G, B\}$
- probability of a ``bad'' signal B is $\pi_0(a^i, a^R)$ (non-binary signals later)
- plus utility consequence of primitive actions $u(a^i, a^R)$
- (do not specify what happens if more than one player deviates from a common action chosen by group members as it doesn't matter)

The Model: Peer Punishment Rounds

sequence of audit rounds $t = 1, 2, \ldots$

players may be matched in pairs as *auditor* and *auditee*

in round $t \geq 1$ an auditor i assigned to audit member j observes a signal $z_{t-1}^{j} \in \{G, B\}$ of the behavior of the auditee

two choices b^i: to recommend punishment (p) or not to recommend punishment (n)

auditee j does not get a move.

member i's behavior as an auditor, another signal $z_t^i \in \{G, B\}$ is generated

bad signal is recommended for punishment or a good signal is not recommended for punishment, then the bad signal is generated with probability $\pi(a^R)$, otherwise with probability $\pi^p(a^R) \geq \pi(a^R)$

Error Symmetry

$$\Pr(z_t^i = B \mid z_{t-1}^j = G, b^i = n) = \Pr(z_t^i = B \mid z_{t-1}^j = B, b^i = p) = \pi(a^R)$$

similarly with $\pi^p(a^R)$

- distribution of z_t^i depends only on whether the player “follows the social norm” (punish on bad signal or not punish on good)
- does not depend on which right thing she does
- symmetry of errors simplifies the analysis considerably and makes exact computations possible.
- general results hold also in the asymmetric case

Costs and Punishments

- payoffs additively separable between the initial primitive utilities and costs incurred or imposed during auditing: quasi-linearity
- no discounting: rounds take place relatively quickly
- following a recommendation of punishment a punishment may (or may not) be imposed.
- If imposed both auditor and auditee suffer a cost, plus an additional social cost to players who do not participate in that particular match
- auditor i suffers a utility loss of C_t^i
- auditee suffers a utility loss of $P_t^i \geq 0$
- rest of group suffers a utility loss of C_t^{iG} evenly divided among the $N-2$ players who do not participate in the match.

Nature of Punishment

punishment may have many possible forms

- if auditee is fired from his job, removed from the organization or demoted can have an adverse effect on the organization and lower utility of those group members who are not directly involved
- punishments may involve the collaboration of the entire group - for example shunning or refusing to speak to a group member
- “avoidable” by an individual group who may refuse to go along with the “social norm” of carrying out the punishment
- rather than giving each player several decisions: whether to punish in a particular audit and also whether to carrying out their own “share” of a punishment, we compress the decision into a single decision “whether to follow the social norm”
- kind of a Jehiel “analogy-based” equilibrium

Repeated Punishments

- individuals potentially punished and pay cost of punishment more than once
- with indivisible punishments such as being fired from a job viewed as "demerits" or probabilities that are cumulated to the end of the game at which point they determine the chance the player is fired
- it makes sense for probabilities of indivisible punishments that utility is additively separable.
- auditee must have non-negative cost
- other costs may be either positive or negative
- allows possibility that particular individuals may benefit from the punishment : if auditee is demoted, some other group member may be promoted

Enforceability

initial primitive round probability of “bad”signal B is $\pi_0(a^i, a^R)$ and utility is $u(a^i, a^R)$

as in repeated game literature: does a punishment scheme based on the signal exist such that a^R is incentive compatible?

enforceability if for some punishment P_1 and for all a^i

$$u(a^R, a^R) - \pi_0(a^R, a^R)P_1 \geq u(a^i, a^R) - \pi_0(a^i, a^R)P_1$$

if for all a^i we have $u(a^i, a^R) - u(a^R, a^R) \leq 0$ we say that a^R is *static Nash* (no peer discipline needed) define $\sigma(a^i, a^R) \equiv \mathrm{sgn}(\pi_0(a^i, a^R) - \pi_0(a^R, a^R))$

for $\sigma(a^i, a^R) = 0$ (actions indistinguishable from a^R)

if $u(a^i, a^R) = u(a^R, a^R)$ define *gain function*

$$\tilde{G}(a^i, a^R) = 0$$

otherwise $\tilde{G}(a^i, a^R) = \mathrm{sgn}(u(a^i, a^R) - u(a^R, a^R)) \cdot \infty$

(actions that are distinguishable from a^R) define gain function

$$\tilde{G}(a^i, a^R) = \frac{u(a^i, a^R) - u(a^R, a^R)}{\pi_0(a^i, a^R) - \pi_0(a^R, a^R)}.$$

also define

$$G(a^R) \equiv \max_{\sigma(a^i, a^R) \geq 0} \tilde{G}(a^i, a^R)$$

$$G^-(a^R) \equiv \min_{\sigma(a^i, a^R) < 0} \tilde{G}(a^i, a^R)$$

Note: $G^{-}(a^R) \geq 0$ if and only if $\sigma(a^i, a^R) \geq 0$ for all $u(a^i, a^R) - u(a^R, a^R) > 0$.

Lemma [Enforceability]: *The group action a^R is enforceable with the punishment P_1 if and only if $\max\{0, G(a^R)\} \leq P_1 \leq G^{-}(a^R)$ hence it is enforceable if and only if $\max\{0, G(a^R)\} \leq G^{-}(a^R)$*

enforceability only concerns first audit round

Implementations

- peer punishment environment taken as an exogenous economic fundamental
- does not completely specify a game
- also must specify the matches take place and how the punishments and costs are determined
- a complete specification called an *implementation*
- start with simple example: *two-stage implementation*

The Two-Stage Implementation

- game rounds continue until a randomization device brings end
- beginning of the first audit round (equivalently: end of the initial primitive round)
- fixed probability $0 < \delta_0 < 1$ that game will continue with first audit round
- beginning of the second audit round and all subsequent rounds continuation probability $0 < \delta < 1$
- during each audit round each player audits exactly one other and is audited by one other
- matching takes place by randomly placing players on a circle and having each player audit the adjacent opponent in the clockwise direction

Exogeneity

key property of matching procedure:

- chances of future matches independent of the actions taken by players
- together with symmetric errors implies that in deciding what action to take a player need only consider the chances of being punished in the immediately following audit round

Completion of Specification of Implementation

punishments take place whenever they are recommended

punishments are fixed constants P_1, C_1, C_1^G same for all players

assume $C_1 + C_1^G + P_1 \geq 0$: no net benefit to the group from carrying out a punishment

Histories and Equilibrium

public histories: previous realizations of the matchings

private information: initial primitive action, signals received, audit actions taken

the signal about the player herself, may or may not be part of the private history of that player.

pure strategy a map from histories and opponent signals to punishment recommendations

a profile of strategies are *Nash Equilibrium* if given the strategies of the others no player can improve his payoff

equilibrium is a *peer discipline equilibrium* if all players follow the strategy of punishing on the bad signal and not punishing on the good signal

a^R is *incentive compatible in the implementation* if a peer discipline equilibrium with a^R as common initial action

Implementations with Social Consensus

two-stage punishment game a special case

in audit rounds both the matching and the punishments are determined endogenously through *social consensus*

matching determined at the beginning of the round, punishments after recommendations

in the general case no assumption of anonymity and players may be treated differently based on their name

- may be that some people are audited less frequently than others, so must be punished more when “caught”
- or only a subset of the population carry out audits
- or a hierarchy: only “managers” conducting audits

Social consensus

a simultaneous move subgame

starts with operation of a public randomization device

a set of alternatives $\mathfrak{B}$

a default alternative $\mathfrak{b}_0 \in \mathfrak{B}$,

each player simultaneously chooses a particular alternative depending on the realization of the device and possibly on previous social consensus

a given number $N/2 < K < N$

outcome of the game is the unique alternative that is the consensus of K or more players, or the default alternative $\mathfrak{b}_{\mathfrak{o}}$ if there is no consensus

if all players agree on the same rule because $K < N$ no player is decisive and so no player can change the consensual decision. So: every alternative in $\mathfrak{B}$ is part of an equilibrium regardless of payoffs

The General Linear Case

recall: C_t^{iG} is the total cost incurred by the $N-2$ players not in the match due to punishment in the match in which player i is auditor at time t and that C_t^i, P_t^i are the cost to i and punishment to the auditee when i is the auditor at time t

feasible set of punishments costs is

$$P^i, C^i = \theta P^i, C^{iG} = \psi P^i | P^i \geq 0$$

where $\theta + \psi + 1 \geq 0$ (no net benefit to the group from carrying out a punishment)

Characterizaton

Theorem: *The non static-Nash enforceable initial action a^R is incentive compatible for some implementation if and only if $|\theta| < \Delta\pi$. In this case to maximize the average expected utility of the group it is necessary and sufficient that the incentive constraints hold with equality for each positive probability public history. The average expected equilibrium utility level per person is*

$$u(a^R) - \left[\pi_0 + \frac{|\theta|}{\Delta\pi - |\theta|}\pi\right](\theta + \psi + 1)G(a^R)$$

this is achievable with the two-stage implementation

Group Size and the Strength of Groups

- measure of group effectiveness by its ability to mobilize resources
- here willingness to pay
- in thinking about the problem of group strength bear in mind that the size of effective groups is often quite large - for example in the U.S. there are about 3 million farmers and 2 million farms
- if there is really a public goods problem that must be overcome by the group, the problem for the farm lobby should be nearly insurmountable

Willingness to Pay

prize worth S

divided equally among the group, each group member getting a benefit of S/N

how much effort is the group willing to provide to get the prize?

Considerations

- in the one-person model of willingness to pay assumed that commitment to pay - for example by bidding in an auction - will in fact be honored
- less evident with a group without peer enforcement: group will be happy to submit a high bid - but when the time comes for group members to provide the promised effort each will wish to shirk, and there is no effective mechanism for forcing them not to
- with peer enforcement the group can credibly commit - for example by social consensus - to providing and enforcing the promised effort provision. So we will focus on a group which has available a linear peer punishment technology.

Divisibility

if the bid is successful an individual who provides effort a^i receives the net benefit $u(a^i, a^R) = (S/N) - a^i$ where we normalize the unit cost of effort to one

with perfectly divisible effort:

to submit a bid of b the group can have each individual member pay b/N

if the prize is contingent on the group fulfilling the promise of effort, each individual is decisive - if any member fails to pay their share the bid falls short and the prize is lost

hence the group is willing to bid any amount up to S and there is no public good, peer enforcement or other problem

Indivisibility

as a practical matter effort is not indivisible

- lobbying, protesting, bribing and so forth require an overhead cost of thinking about and organizing oneself to participate in the activities
- not feasible to spend two minutes a year contributing to a group effort in an effective way
- we focus on the case where each group member can provide either 0 or 1 unit of effort
- to bid b the group should appoint a subset of b members each to provide an effort level of 1.
- with perfect observability again each appointed member is decisive - however those chosen to contribute will do so only if $S/N \geq 1$ (this seems to be the case Olson has in mind)
- in particular if $N > S$ peer discipline is needed if the group is to make a non-trivial bid

Imperfect Monitoring

- idea that individuals are decisive in a large group is neither very interesting or relevant
- monitoring is not so perfect
- basis of the anti-folk theorem is the idea that with even a small amount of noise in observing individual behavior the use of decisiveness breaks down completely in a large group

The Model

- decentralized process with imperfect monitoring for determining who contributes
- each individual receives a random signal $\{0, 1\}$ of whether or not to provide effort
- social consensus determines the probability μ of the “effort” signal of 1
- signal observed by the auditor with fixed garbling:
- signal observed by the auditor is the same as that of the initial round auditee with probability $1 - \epsilon$ and is the opposite with probability $\epsilon \leq 1/2$
- auditor also observes whether or not the auditee has contributed

Enforceability and Signal Compression

four possible strategies:

A: contribute on 1 do not contribute on 0

B: never contribute

C: always contribute

D: contribute on 0 do not contribute on 1

interested in the enforceability of A

Linearity and Randomization

(action,signal)

four possible punishments $P_{11}, P_{01}, P_{10}, P_{00}$

$P_1 = \max\{P_{00}, P_{01}, P_{10}, P_{11}\}$

$\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11}$ are probabilities the one corresponding to the $\operatorname{argmax}\{P_{00}, P_{01}, P_{10}, P_{11}\}$ being equal to 1

set $P_{00} = \beta_{00} P_1, P_{01} = \beta_{01} P_1, P_{10} = \beta_{10} P_1, P_{11} = \beta_{11} P_1$

compress the underlying signal with four outcomes to a single binary signal: if the underlying signal has the value (e^i, s^i) we can assign B with probability β_{e^i, s^i} and neither incentive or costs are changed.

Compression Through Social Consensus

β's are chosen by social consensus to maximize welfare

$$S/N - \mu - \left[\pi_0 + \tfrac{|\theta|}{\Delta\pi - |\theta|}\pi\right](\theta + \psi + 1)G(a^R)$$

Theorem: *Welfare maximization implies* $\beta_{00} = 0, \beta_{01} = 1, \beta_{10} = 0, \beta_{11} = 0$*. Correspondingly* $\pi_0 = (1-\mu)\epsilon$ *and* $G(a^R) = 1/(1-\epsilon)$*with welfare being equal to*

$$\frac{S}{N} - \mu - \left[(1-\mu)\epsilon + \frac{|\theta|}{\Delta\pi - |\theta|}\pi\right](\theta + \psi + 1)\frac{1}{1-\epsilon}$$

Willingness to Pay

Maximum willingness to pay with outside option of 0 is max of 0, min of S and:

$$N\mu = \frac{S - N\left[\epsilon + \frac{|\theta|}{\Delta\pi - |\theta|}\pi\right](\theta + \psi + 1)\frac{1}{1-\epsilon}}{1 - (\theta + \psi + 1)\frac{\epsilon}{1-\epsilon}}$$

assume ϵ small enough that $\mu < 1$

$$\frac{\epsilon}{1-\epsilon} < \frac{1}{\theta + \psi + 1}$$

Result

Corollary: *The maximum μ the group is willing to enforce is positive iff*

$$s > (\theta + \psi + 1)\,\frac{\frac{|\theta|}{\Delta\pi - |\theta|}\pi + \epsilon}{1 - \epsilon}$$

in which case the ratio $N\mu/S = \mu/s$ is given by

$$\frac{1 - \frac{1}{s}\frac{\frac{|\theta|}{\Delta\pi - |\theta|}\pi + \epsilon}{1 - \epsilon}(\theta + \psi + 1)}{1 - (\theta + \psi + 1)\frac{\epsilon}{1 - \epsilon}}$$

which is increasing in per-capita value s

Competition Between Groups

consider a “small” group lobbying against a larger interest

the “small” group chooses the size of the prize S (how big farm subsidies should be (since s is larger it is willing to bid more)

it pays the bid of the larger group (second price auction) so gets

$$S - \frac{S - N_L \frac{\frac{|\theta|}{\Delta\pi - |\theta|}\pi + \epsilon}{1-\epsilon}(\theta + \psi + 1)}{1 - (\theta + \psi + 1)\frac{\epsilon}{1-\epsilon}}$$

this is decreasing in S if the larger group pays a positive amount, so it should choose

$$S = N_L \frac{\frac{|\theta|}{\Delta\pi - |\theta|}\pi + \epsilon}{1 - \epsilon}(\theta + \psi + 1)$$