Peer Monitoring, Ostracism and the Internalization of Social Norms

with Rohan Dutta and Salvatore Modica

Introduction

- build on work showing the importance of self-enforcing social norms in enabling groups to overcome public goods problems (Olson, Ostrom)
- social norms are endogenous: Boyd-et-al cross-cultural experiments

Our Model

elaborate on the model of peer incentives from Kandori, Levine/Modica and Levine/Mattozzil

an environment where monitoring is difficult (few monitors)

- individual behavior: Nash equilibrium with respect to selfish preferences
- collective decisions: groups can coordinate on a mutually advantageous equilibrium
- monitoring and penalties for anti-social behavioral
- internalization of social norms
- stickiness of social norms

Issues

- cost of punishing the monitor depends social closeness of monitor and producer: trade-off between information and incentives; rotation, supervisor versus peer review, police versus doctors
- optimality of social norms outside the laboratory may lead to the failure of procedures such as double-blind designed to reduce or eliminate possibility of outside influence
- tradeoff between social benefit and the social cost of monitoring: external incentives for public good contribution – substitute or complement? Perverse effects with fixed costs
- more general Lucas critique of experiments (lab, field, natural) – interventions may (or may not) change social norms depending upon circumstances
- does internalization complement or substitute incentives?
- cultural norms and strategic subsidies of internalization

The Base Model

- large group where monitoring is difficult in the sense that each production decision is observed by at most one other person.
- continuum of pairs with a unit mass
- pair consists of a producer and monitor

Technology

producer effort $e \in \{0, 1\}$ with cost ec where $c > 0$

value of public good: fraction of pairs producing ϕ per capita benefit ϕV

monitor costlessly observes noisy signal $z \in \{0, 1\}$: with probability π the signal is wrong; makes report $x \in \{0, 1\}$

social interaction: population is rematched into social subgroups of size $N \geq 4$; producer and monitor in same subgroup h

exactly one of the N members of each subgroup randomly chosen to be presenter and may volunteer to share an interesting story

$N - 1$ members of anonymous audience observe the report by or about the presenter and vote whether to ostracize; $1 < K < N - 1$ votes in favor lead to ostracism

presentation has value of N to the presenter and βN to each audience member

Truthful Strategies

truthful strategy:

- choice of whether or not to produce as a producer
- whether to send the message equal to the signal if a monitor
- always volunteer a story conditional on having one
- rule for ostracizing the presenter

social norm: a truthful strategy that if followed by everyone is a Nash equilibrium

collective decision: group chooses *optimal social norm* that maximizes the ex ante per capita utility of the identical group members (*social utility*)

Two Types of Social Norms

default norm

no effort

all stories to be volunteered

nobody ostracized

utility from only the social interaction $U = 1 + \beta(N-1)$.

implementation of production

$e = 1$

monitor tells the truth

all stories are volunteered

incentive compatible ostracism rule

note that all ostracism rules are incentive compatible for the audience because nobody is decisive

Implementing Production

potential social norms denoted by s correspond to ostracism probabilities $p(x), q(x)$ as function of the report $x \in \{0, 1\}$.

ostracizing one member of a pair imposes in expectation a cost of 1 on that person and a cost of $h\beta$ on the partner.

per capita probability of ostracism [on the equilibrium path]

$$\Pi(s) = (1/2)[\pi p(0) + (1-\pi)p(1)] + (1/2)[\pi q(0) + (1-\pi)q(1)].$$

social utility $W(s)$ is per capita payoff from production V minus the per capita cost of production (half the producer cost) plus utility from the social interaction minus the expected cost of ostracism:
$W(s) = V - c/2 + (1 - \Pi(s))U$.

Cost of Implementing Production

$\mathrm{W(s)=U+V-[\Pi}(s)U + c/2]$

cost of implementation $C(s) = \Pi(s)U + c/2$

monitoring cost plus production cost

optimal social norm must minimize implementation

implementation will be optimal if and only if $V \geq \min_s C(s)$.

Cost Minimizing Social Norms

Theorem: *If and only if the* implementation condition

$$\frac{c}{(1-2\pi)(1-h^2\beta^2)} \leq 1$$

is satisfied can production be implemented. In the cost minimizing social norm producers who are reported to have taken the bad action ($x = 0$*) are ostracized with probability* $p(0) = P$ *and monitors who report the good action (*$x = 1$*) are ostracized with probability* $q(1) = Q$ *and there is no other ostracism. The ostracism probabilities are*

$$P = \frac{c}{(1-2\pi)(1-h^2\beta^2)}, \quad Q = h\beta P,$$

and the cost of implementation is

$$C = \left[\frac{U}{2}\frac{\pi + (1-\pi)h\beta}{(1-2\pi)(1-h^2\beta^2)} + \frac{1}{2}\right] c.$$

Further Discussion

- note the discontinuity: implementation fails abruptly
- feedback effect: a bigger punishment for the producer implies a bigger punishment for the monitor. The feedback effect is that the latter reduces the incentive for the producer to produce: by not producing she can reduce the probability the monitor is punished for sending a good report.
- must punish the monitor for good reports even though that is the only kind submitted and they are known to be true
- only way to get the monitor to tell the truth is to make her indifferent between the two reports. There is no mechanism or social norm in which the monitor strictly prefers to tell the truth
- malicious gossip is valued in the sense that a monitor is less likely to be ostracized for filing a bad report.
- cost of implementation is proportional to c the incentive to cheat on the social norm; standard result in peer monitoring

Alternative Monitoring Technologies

a fraction of monitors randomly assigned to a fraction of producers

producer may have no monitors, one monitor, or many monitors, randomly determined

who knows what about whom?

two extremes:

1. very few monitors so that the number of monitors per producer can as a good approximation be taken to be either zero or one, with the producer unaware of whether a monitor is present,

2. very many monitors all of whom observe exactly the same signal

our benchmark case lies between these two extremes

Few Monitors

η probability monitor is present to witness a production decision

only effect is to change the incentive constraint for the producer

$$\eta P = \frac{1}{1 - h^2\beta^2} \frac{c}{1 - 2\pi},$$

implementability accordingly harder to satisfy, but implementation cost does not change since larger punishments are used with smaller frequency

Many Monitors

many monitors who observe exactly the same signal

ostracize all monitors with probability one for disagreement

if all tell the truth all strictly prefer to tell the truth

in equilibrium no punishment of monitors

same as $h = 0$.

Applications

- rotation and expertise: trade-off with decreasing $\pi = f(h)$; police external monitors, surgeons internal monitors
- urban slum versus poor rural village – shops versus restaurants
- double-blind/dictator in the laboratory
- fixed cost/stickiness, external incentives and discontinuous response

Generalized Lucas Critique

small interventions are unlikely to change social norms hence conclusions drawn from small interventions may mislead as the effect of large interventions

for example: subsidizing mosquito netting in a few villages is unlikely to change religion practices, but doing over an entire region may

the point is: in doing interventions it is generally assumed social norms are fixed and have no particular reason for being what they are

in fact: religious practices may be a well-chosen social norm to respond to circumstances

Investment in Social Norms

as before the group/principal announces a pure strategy σ called the social norm.

after this announcement and before matching, production and monitoring individuals may choose to invest (or specialize) in a pure strategy s of their choice

cost investment:

$a \geq 0$ if the strategy chosen $s = \sigma$ is the social norm,

$a + \Gamma$ if the strategy chosen $s \neq \sigma$ is not the social norm, where $\Gamma \geq 0$ is the benefit of conformity

it is less costly to learn the language used by everyone else than to invent your own language

choice of investment is known only to the investor: no punishment is possible based on the investment decision

Consequences of Investment

an investor gets utility from the strategy invested in

if s is chosen and the terminal node is consistent with s the investor receives a bonus of $B \geq 0$ the *value of commitment*

we assume $a + \Gamma \leq B$ so that investing in a strategy and following it is profitable

internalization means that individuals choose to invest in the social norm

observe that the group/principal should never choose a social norm that will not be internalized: it would always be better to announce as the social norm the equilibrium strategy chosen by members

Essential versus Inessential Indifference

solution of the basic model involved several forms of indifference

the producer is indifferent between producing and not producing

- inessential: can be made strict by punishing a little more for a bad signal

the monitor is indifferent between reporting 0 and 1

- essential: cannot be made strict; model not robust to introducing a small cost of observing the signal

the audience members are indifferent to ostracizing or not ostracizing

- essential: cannot be made strict; weakly dominant not to ostracize; model not robust to small probability unanimity is required for ostracism.

$B, \Gamma > 0$ makes all indifference inessential and the model robust

Conformity and Ostracism: Complements or Substitutes?

- need both B, Γ; both large enough get complete internalization
- may need both internalization and incentives to implement production (complements)
- once production is implemented bigger B, Γ substitutes for incentives
- can say which constraint (monitor, producer) you should "spend" you Γ on
- for $\pi \leq .333$ spend it on the monitor (producer strictly prefers to produce)

Social versus Cultural Norms

- individuals choose social norms
- cultural norms are generally derived at a young age from others, especially parents and peers
- cultural norms require a much larger investment and have a much greater value of commitment
- should be part of the same theory as that of social norms
- investment in strategies can be subsidized by interested parties
- public schools teach national myths; fight over curriculum is over history, language, religion – not arithmetic or reading
- combine with Bisin-Verdier horizontal/vertical models?