Punish Liars, Not Free-Riders

Brenton Kenkel

November 30, 2017

University of Rochester/Vanderbilt University

Motivation

Uncertainty about willingness to contribute to collective action or public goods.

- Climate efforts
- Refugee crisis
- Military coalitions

Private information makes collective action even harder.

- Is the project feasible?
- How to divide the labor?

Central Question

When and how can *communication* promote cooperation in collective action when actors have private information?

- Won't solve all problems of collective action
- Can it help with those that stem from private information?

Main Findings

In a simple model of repeated collective action, communication can bring us up to complete-information second-best if:

1. Participants care enough about the future
2. Set of potential projects is risky enough (sufficiently high chance of failure)
3. We punish giving *more* than you claim to be willing

Stage Game: Primitives

- Players $i \in \{1, 2\}$
- Types $\omega_i \in \Omega_i = \{0, 1, 2\}$
    - i.i.d., nonzero prior probabilities $(p_0, p_1, p_2) \in \Delta^2$
    - determine marginal cost of contribution
- Messages $m_i \in M_i = \{0, 1, 2\}$
    - cheap talk
- Contributions $x_i \in X_i = \{0, 1, 2\}$
    - voluntary, non-refundable

Stage Game: Timing

1. Each player learns her own ω_i
2. *Messaging stage:* simultaneously send m_i
3. Observe messages, update beliefs about t_j to $\lambda_i(m_j) \in \Delta^2$
4. *Contribution stage:* simultaneously choose x_i

Stage Game: Payoffs

Public good worth 1, costs 2 units effort to produce:

$$u_i(x_i, x_j, \omega_i) = \mathbf{1}\{x_i + x_j \geq 2\} - c(\omega_i)x_i$$

Player is willing to contribute up to ω_i if pivotal:

- $c(0) = c_0 > 1$
- $c(1) = c_1 \in (1/2, 1)$
- $c(2) = c_2 \in (0, 1/2)$

Efficient Provision

An assessment satisfies *efficient provision* if:[1]

- Supply whenever feasible: If $\omega_1 + \omega_2 \geq 2$, then $x_1 + x_2 \geq 2$
- No wasted contributions: $x_1 + x_2 \in \{0, 2\}$
- Efficient distribution of costs: if $\omega_i < \omega_j$, then $x_i = 0$

[1] See also Palfrey, Rosenthal, and Roy (2017).

Efficient Provision: Necessary Condition

Lemma

In any PBE of the stage game that satisfies efficient provision, at least one player's messaging strategy is fully separating.

Proof. Let m_i, m_j be messages sent on the equilibrium path.

$\{0,1\} \subseteq \operatorname{supp} \lambda_j(m_i) \Rightarrow$ when $t_j = 1$, must violate supply whenever feasible or no wasted contributions

$\{1,2\} \subseteq \operatorname{supp} \lambda_j(m_i) \Rightarrow$ when $t_j = 1$, must violate supply whenever feasible or efficient distribution of costs

$\{0,2\} \subseteq \operatorname{supp} \lambda_j(m_i) \cap \operatorname{supp} \lambda_i(m_j) \Rightarrow$ when $t_i = t_j = 2$, must violate supply whenever feasible or no wasted contributions

Inefficiency in the Stage Game

Proposition

No PBE of the stage game satisfies efficient provision.

Proof. Suppose the contrary. WLOG let 1's message separate.

2's equilibrium response to 1's message:

	$t_2 = 0$	$t_2 = 1$	$t_2 = 2$
$m_1 = 0$	$x_2 = 0$	$x_2 = 0$	$x_2 = 2$
$m_1 = 1$	$x_2 = 0$	$x_2 = 1$	$x_2 = 2$
$m_1 = 2$	$x_2 = 0$	$x_2 = 0$	$x_2 \leq 2$

Strictly profitable to deviate from $m_1 = 2$ to $m_1 = 1$.

Repeated Extension

- Discrete periods $t = 0, 1, \ldots, \infty$
- Common discount factor $\delta \in [0, 1)$
- Period-specific types ω_i^t
  - i.i.d. across players *and* periods
  - interpret as new problems/projects arising over time
- Messages m_i^t
- Contributions x_i^t
- Total discounted payoffs $\sum_{t=0}^{\infty} \delta^t u_i(x_i^t, x_j^t, \omega_i^t)$

Symmetric Truth-Telling

Let *symmetric truth-telling* be defined by the finite automaton:

- State space $\{h, b\}$ (*h*onest, *b*abbling)
- Initial state $s^0 = h$
- In state h:
  - In messaging stage, announce $m_i^t = \omega_i^t$
  - In contribution stage, play corresponding symmetric efficient provision equilibrium
- In state b:
  - In messaging stage, randomize uniformly
  - In contribution stage, play a fixed BNE of the game without communication
  - Can be *any* such BNE, no need for minmaxing

Symmetric Truth-Telling, cont.

- Transition function
  - If $s^t = h$, go to $s^{t+1} = h$ if contributions consistent with announcements, $s^{t+1} = b$ otherwise
  - If $s^t = b$, go to $s^{t+1} = b$

Critical feature — move to babbling if a player contributes *more* than claimed (i.e., $x_i^t > m_i^t$).

Without this, types $t_i = 2$ would always prefer to screen by sending $m_i = 1$.

Ex Ante Superiority

Lemma

If $p_0 \geq c_2/(1 - c_2)$, the *ex ante* expected utility of symmetric truth-telling in the stage game is strictly greater than that of any babbling equilibrium.

Proof sketch. Best babbling outcome for i is $x_j = \omega_j$.

$t_i = 0$ is indifferent.

$t_i = 1$ gets same contributions, but can condition on t_j — strict improvement.

$t_i = 2$ is better off in case $t_j = 0$ (reduce risk of failure), worse off in case $t_j > 0$ (less burden sharing). So a sufficient condition is high enough chance $t_j = 0$.

What We Have

If $p_0 \geq c_2/(1-c_2)$, the continuation value of the honest stage in symmetric truth-telling strictly exceeds that of the babbling stage.

As $\delta \to 1$, a one-shot deviation that pushes us into babbling can never be profitable.

So for great enough p_0 and δ, never an incentive to deviate from the contributions *given* the messages.

What We Don’t Have (Yet)

Still need to worry about incentive compatibility in the messaging stage.

Player can avoid detection by misrepresenting type, then adhering to that type’s prescribed contributions.

Need to worry about $t_i = 2$ mimicking $t_i = 1$.

Incentive Compatibility

Lemma

There is an unobservable profitable deviation from symmetric truth-telling if and only if $p_0 < c_2/(1-c_2)$.

Proof. Consider $t_i = 2$ who mimics $t_i = 1$ and follows the prescribed contributions.

If $t_j = 0$, she saves $2c_2$ but the project fails, for a net difference of $2c_2 - 1$.

If $t_j \in \{1, 2\}$, she saves c_2 and the project still succeeds.

Therefore, the deviation is profitable if and only if

$$p_0(2c_2 - 1) + (1 - p_0)c_2 > 0 \quad \Leftrightarrow \quad p_0 < \frac{c_2}{1 - c_2}.$$

Main Finding

Proposition

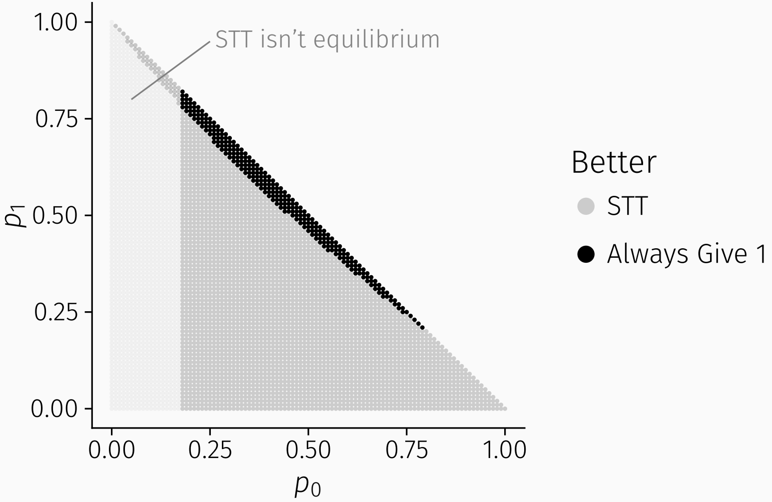
If $p_0 \geq c_2/(1 - c_2)$ and δ is sufficiently large, symmetric truth-telling is an equilibrium.

Observations

- In a repeated setting, cheap talk can substantially reduce inefficiencies due to private information
- Equilibrium sustained solely by threat of communication breakdown
- Requires punishing “unanticipated generosity”
- Critical condition is positive risk of project failure

Welfare Comparison

Difference between symmetric truth-telling and no-communication strategy where all $x_1^t = x_2^t = 1$.



Can We Do Better?

For cases where always-give-1 beats symmetric truth-telling, imagine a strategy where:

- Players honestly announce types
- If $\omega_1^t + \omega_2^t < 2$, contribute $x_1^t = x_2^t = 1$
- If $\omega_1^t + \omega_2^t \geq 2$, contribute according to symmetric efficient provision

Call this *efficient assured completion*.

We Can't Do Better

Proposition

There is no equilibrium with efficient assured completion.

Proof: 2's equilibrium response to 1's message:

	$t_2 = 0$	$t_2 = 1$	$t_2 = 2$
$m_1 = 0$	$x_2 = 1$	$x_2 = 1$	$x_2 = 2$
$m_1 = 1$	$x_2 = 1$	$x_2 = 1$	$x_2 = 2$
$m_1 = 2$	$x_2 = 0$	$x_2 = 0$	$x_2 = 1$

Because of assured completion, now no downside to mimicking message *and* contributions of $t_1 = 1$.

Closing Thoughts

Conclusions:

- Under uncertainty, cannot simply punish free-riders
- Honest communication is sustainable if:
  - Interaction is repeated
  - "Too high" contributions are punished
  - Real risk of failure if dishonest

Future directions:

- Historical application to alliances?
- Endogenize project selection?
- Lab experiment?