Social norms and incentives in repeated games

Mihaela van der Schaar
Electrical Engineering Department, UCLA
Applications

Providing services/information
Providing/sharing resources

Social networks, societal networks, crowdsourcing platforms, social cloud computing, social networks, expert networks, peer review systems, online labor markets, P2P networks, multi-user mobile communication, cognitive radio networks, smart grids, etc.

Files  Knowledge  Opinions  Labor
Canonical (Gift-Giving) Game

- **Actions:**
  - Requester: no action to choose
  - Worker: \( a \in \mathcal{A} = \{S, NS\} \)
    - **S:** High level of effort/resources
    - **NS:** Low level of effort/resources

- **Game:**

<table>
<thead>
<tr>
<th>Requester ( b )</th>
<th>( S )</th>
<th>( NS )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(-c)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

  \( b - c > 0 \)

  Individual vs. society goals!

  Incentives needed!
## Canonical (Gift-Giving) Game

- Requester always pays the same amount $q$ (flat-rate pricing)
  - The worker receives $\mu q$.
  - The website charges $(1 - \mu)q$ as the transaction fee.
- Actions:
  - Requester: no action to choose
  - Worker: $a \in \mathcal{A} = \{S, NS\}$
    - S: High level of effort
    - NS: Low level of effort

### Game

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requester</td>
<td>$b - q$, $\mu q - c$</td>
<td>$-q$, $\mu q$</td>
</tr>
</tbody>
</table>

Individual vs. society goals! Incentives needed!
How to provide incentives?

Incentive provision through rewards and punishments

By what?
- Payment (Pricing / Credit)
- Differential service (by history)

By whom?
- Users (Repeated interaction)
- System/Policer (Intervention)

Direct/Personal reciprocity (Identity)
Indirect/Social reciprocity (Anonymous)
Social norms

- A social norm is a rule that defines “appropriate” and “inappropriate” behaviors
  - Compliance
  - Rewards (present and future)
  - Punishments (present and future)
- A social norm defines a strategy for the repeated game
- History in anonymous settings?
  - Ratings
- **Social norms: 3 parts**
  - Rating levels
  - Plan based on current ratings
  - Update rule based on current ratings and behavior
Moral hazard and adverse selection

• “Free-riding” is a moral hazard problem (what is good for the individual is not good for society)
  – Reciprocity partly solves moral hazard problems by providing additional incentives to work

• Another issue: some workers are better than others (adverse selection)
  – Reciprocity partly solves adverse selection problems as well by differentiating good workers from bad workers

• This talk: focuses on moral hazard
Focus

- Seminal work: Kandori
  - Focus on folk theorems
  - Not constructive
  - No reporting errors
  - Patient players etc.

- Agenda here:
  - Optimal social/system performance
  - Constructive – how do norms and ratings look like?
  - How information about others shapes design?
  - Reporting errors
  - Not patient players
## Folk Theorems: Good and Bad ...

<table>
<thead>
<tr>
<th>Good</th>
<th>Bad</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can get full cooperation</td>
<td>Can get “anything”</td>
</tr>
<tr>
<td>Can get “anything”</td>
<td>Players must be patient</td>
</tr>
<tr>
<td></td>
<td>Full information</td>
</tr>
</tbody>
</table>
Folk Theorem: Bad ...

In most real situations

• players not “infinitely patient”

  • designer’s task harder

  • players do not / cannot know full history

  • designer may control information – designer has more tools
Selfish agents want to follow the protocol = incentive compatibility = sustainability
Anonymity

• players do not know who they played in the past
• players know only summary of past history (rating)

information controlled by designer
Learning the information

How would players learn information?

- manager monitors play
- manager makes announcements

public announcements

private announcements
errors in reports
- to monitor
- from monitor

imperfect monitoring
Imperfect Monitoring

Players do not see full history

Each period
- action profile $a \in A$
- signal $s \in S$
  stochastic $\pi(s|a)$

Two possibilities
- all players observe same $s$
- different players observe different $s$

We focus on imperfect public monitoring
Using the information

Example

Each player knows

- own rating
- rating of opponent
- distribution of ratings (announced by designer)

Why does this matter?

Because current rating distribution affects

- future distribution
- who player will meet in future / future opponents
- what future opponents will do
Y. Xiao and M. van der Schaar,
"Socially-Optimal Design of Service Exchange Platforms with Imperfect Monitoring,"
A general resource exchange system:
- A set of users \( \mathcal{N} \triangleq \{1, \ldots, N\} \)
- Users have resources valuable to the others
- Users are long-lived in \( t = 0, 1, 2 \ldots \)
- At each period \( t \):
  - Each user (as a client) requests resources
  - Each user (as a server) is matched to a client
  - Each server chooses the effort level in providing resources

Assumptions:
- Two effort levels: “low” and “high” \( \{0, 1\} \)
- Clients have no cost in requesting
- Homogeneous users:
  - Servers’ costs to exert high (or low) effort same across users
  - Clients’ benefits from high (or low) effort same across users
- No monetary exchange
(Stage-game) Model:
- A gift-giving game between a client and a server
  high effort low effort
  request \[ (b, -c) \] \[ (0, 0) \]
- A matching: \( m : \mathcal{N} \rightarrow \mathcal{N}, i \) (server) \( \leftrightarrow m(i) \) (client)
- Set of proper matchings:
  \[ M = \{ m : \text{m bijective and } m(i) \neq i, \forall i \in \mathcal{N} \} \]
- A matching rule: \( \mu : M \rightarrow \Delta(M) \)
- Focus on uniform random matching

Rating mechanisms:
- Assign each user \( i \) with a rating \( \theta_i \in \Theta \)
- Rating profile (Unknown to users): \( \theta \in \Theta^N \)
- Rating distribution (Known to users): \( s(\theta) \)
In each period $t$:

- The platform displays $s(\theta)$, and announces the recommended plan $\alpha_0 : \Theta \times \Theta \rightarrow \{0, 1\}$
- Each user $i$ requests resources
- Each user $i$ is matched as a server to user $m(i)$ with probability $\mu(m)$
- Each user $i$ is informed by the platform of the client’s rating $\theta_{m(i)}$
- Based on its plan $\alpha_i : \Theta \times \Theta \rightarrow \{0, 1\}$, each user $i$ chooses its effort level $\alpha_i(\theta_{m(i)}, \theta_i)$
- Each client reports its (erroneous) assessment of the effort level to the platform
- The platform updates the rating profile based on the rating update rule $\tau : \Theta \times \Theta \times \{0, 1\} \rightarrow \Theta$. 
Design Objectives

- “Simple” rating mechanisms
  - *Binary* rating
  - A *small* set of *(three)* plans

- What does the designer tell the users?
  - Only the rating distribution -> Leads to constraints on strategies adopted by users

- *Construct* the equilibrium strategy to achieve the desired outcome
Design Challenges

- Folk theorems not useful
  - Users not infinitely patient
  - Not constructive

Design Surprise

- Nonstationary equilibrium strategies are essential
“Simple” rating mechanisms

- Altruistic plan $\alpha^a(\theta_c, \theta_s) = 1, \forall \theta_c, \theta_s \in \{0, 1\}$
- Selfish plan $\alpha^s(\theta_c, \theta_s) = 0, \forall \theta_c, \theta_s \in \{0, 1\}$
- Fair plan $\alpha^f(\theta_c, \theta_s) = \begin{cases} 
0 & \theta_s > \theta_c \\
1 & \theta_s \leq \theta_c 
\end{cases}$

- Erroneous report (with error probability $\varepsilon$)

$$R(z'|z) = \begin{cases} 
1 - \varepsilon, & z' = z \\
\varepsilon, & z' \neq z
\end{cases}$$

- Rating update rule

$$\tau(\theta'_s|\theta_c, \theta_s, z) = \begin{cases} 
\beta^+_\theta_s, & \theta'_s = 1, z \geq \alpha_0(\theta_c, \theta_s) \\
1 - \beta^+_\theta_s, & \theta'_s = 0, z \geq \alpha_0(\theta_c, \theta_s) \\
1 - \beta^-_{\theta_s}, & \theta'_s = 1, z < \alpha_0(\theta_c, \theta_s) \\
\beta^-_{\theta_s}, & \theta'_s = 0, z < \alpha_0(\theta_c, \theta_s)
\end{cases}, \text{ for } \theta_s = 0, 1.$$
“Simple” rating mechanisms

- Altruistic plan $\alpha^a(\theta_c, \theta_s) = 1, \forall \theta_c, \theta_s \in \{0, 1\}
- Selfish plan $\alpha^s(\theta_c, \theta_s) = 0, \forall \theta_c, \theta_s \in \{0, 1\}
- Fair plan $\alpha^f(\theta_c, \theta_s) = \begin{cases} 0 & \theta_s > \theta_c \\ 1 & \theta_s \leq \theta_c \end{cases}
- Erroneous report (with error probability $\varepsilon$)

$$R(z'|z) = \begin{cases} 1 - \varepsilon, & z' = z \\ \varepsilon, & z' \neq z \end{cases}$$

- Rating update rule: Rating goes up when the reported effort level exceeds the recommended effort level

$$\tau(\theta'_s|\theta_c, \theta_s, z) = \begin{cases} \beta^+_{\theta_s}, & \theta'_s = 1, z \geq \alpha_0(\theta_c, \theta_s) \\ 1 - \beta^+_{\theta_s}, & \theta'_s = 0, z \geq \alpha_0(\theta_c, \theta_s) \\ 1 - \beta^-_{\theta_s}, & \theta'_s = 1, z < \alpha_0(\theta_c, \theta_s) \\ \beta^-_{\theta_s}, & \theta'_s = 0, z < \alpha_0(\theta_c, \theta_s) \end{cases}$$, for $\theta_s = 0, 1$. Can achieve the social optimum
“Simple” rating mechanisms

- **Altruistic plan** $\alpha^a(\theta_c, \theta_s) = 1, \forall \theta_c, \theta_s \in \{0, 1\}$
- **Selfish plan** $\alpha^s(\theta_c, \theta_s) = 0, \forall \theta_c, \theta_s \in \{0, 1\}$
- **Fair plan** $\alpha^f(\theta_c, \theta_s) = \begin{cases} 0 & \theta_s > \theta_c \\ 1 & \theta_s \leq \theta_c \end{cases}$
- **Erroneous report (with error probability $\varepsilon$)**
  
  \[ R(z'|z) = \begin{cases} 1 - \varepsilon, & z' = z \\ \varepsilon, & z' \neq z \end{cases} \]

- **Rating update rule**

  \[ \tau(\theta'_s|\theta_c, \theta_s, z) = \begin{cases} \beta^+_s, & \theta'_s = 1, z \geq \alpha_0(\theta_c, \theta_s) \\ 1 - \beta^+_s, & \theta'_s = 0, z \geq \alpha_0(\theta_c, \theta_s) \\ 1 - \beta^-_s, & \theta'_s = 1, z < \alpha_0(\theta_c, \theta_s) \\ \beta^-_s, & \theta'_s = 0, z < \alpha_0(\theta_c, \theta_s) \end{cases}, \text{ for } \theta_s = 0, 1. \]

- **Serve everybody**
- **Serve nobody**
- **Only serve users with higher or equal ratings**

Rating goes up when the reported effort level exceeds the recommended effort level.
Under “good” behavior: (the reported effort level is higher or equal to recommended effort level)

Under “bad” behavior: (the reported effort level is lower than recommended effort level)

Different rewards under different behaviors

Design parameters: Punishments for bad behavior
Illustration of rating update rules

Under “good” behavior:
(the reported effort level is higher or equal to recommended effort level)

Under “bad” behavior:
(the reported effort level is lower than recommended effort level)

Design parameters:
- Rewards for good behavior
- Punishments for bad behavior
Stochastic game formulation:

- **Players**: the users and the platform $\mathcal{N} \cup \{0\}$
- **State**: rating profile $\theta \in \Theta^N$
- **Action set (set of plans)**: $A \triangleq \{\alpha | \alpha : \Theta \times \Theta \rightarrow \{0, 1\}\}$
- **Stage-game payoff**: $u_i(\theta, \alpha_0, \alpha)$
- **History at period $t$**: $h^t = (\theta^0, \ldots, \theta^t) \in \mathcal{H}^t$
- **Strategy**: $\pi_i \in \Pi : \sum_{t=0}^{\infty} \mathcal{H}^t \rightarrow A, i = 0, 1, \ldots, N$
- **Strategy profile $\pi = (\pi_1, \ldots, \pi_N)$**
- **Recommended plan $\alpha_0$ and Recommended strategy $\pi_0$**
- **Overall payoff**

$$U_i(\theta^0, \pi_0, \pi) = \mathbb{E}_{h^\infty} \left\{ (1 - \delta) \sum_{t=0}^{\infty} \delta^t u_i(\theta^t, \pi_0(h^t), \pi(h^t)) \right\}.$$
Restrictions on the strategies

• The users know the rating distribution only
• The strategies should depend on rating distributions only
• Rating distributions are announced at each period → Public Announcement (PA) strategy
• Different from public strategies
  – Public strategies: depends on the history of states
  – PA strategies: depends on the history of rating distributions
• Users condition on own rating as well
Public announcement strategies

- Symmetric strategy profile: \( \pi \cdot \mathbf{1}_N \)
- Public announcement (PA) strategy

**Definition (Public announcement Strategy)**

A strategy \( \pi \) is a public announcement strategy, if for all \( t \geq 0 \) and for all \( h^t, \tilde{h}^t \in \mathcal{H}^t \), we have

\[
\pi(h^t) = \pi(\tilde{h}^t), \text{ if } s(\theta^k) = s(\tilde{\theta}^k), \quad k = 0, 1, \ldots, t.
\]

We write the set of all PA strategies as \( \Pi_{PA} \).

- Set of symmetric PA strategies restricted on the subset of plans \( \bar{B} \subset A \): \( \Pi_{PA}(B) \)

Examples of subset B:
B={altruistic, fair, selfish},
or B={altruistic, selfish}
Which equilibrium to use?

Public Equilibrium $\sigma$
- each $\sigma_i$ depends only on history of public signals
- each $\sigma_i$ optimal given $\sigma_{-i}$ and information

Public Perfect Equilibrium (PPE)
- each $\sigma_i$ depends only on history of public signals
- each $\sigma_i$ optimal given $\sigma_{-i}$ and information after every public history
PPE vs (Stationary) Markov Equilibrium

Markov?
  • Might think of last public signal as state
  • **Markov strategy**
    • condition only on current state
  • Public strategy
    • condition on history of states

**Markov Equilibrium**
  • players condition only on current state
  • stationary

**Public Equilibrium**
  • players condition on history of states
  • might not be stationary

We will gain a lot by using *non-stationary strategies*
Public Announcement Equilibrium (PAE)

- Every PAE is a PPE (Public Perfect Equilibrium)
- **More stringent** requirement than PPE
- Allow users to consider deviating to ANY strategy!
  - deviation-proof against the users with the knowledge of rating profiles

---

Continuation strategy: \( \pi_i|_{h^k}(h^t) = \pi_i(h^k h^t) \)

**Definition (Equilibrium Definition)**

A pair of an PA recommended strategy and a symmetric PA strategy profile \((\pi_0, \pi \cdot 1_N) \in \prod_{PA}(B) \times \prod_{PA}^N(B)\) is a PAE restricted on subset \(B\), if for all \(t \geq 0\), for all \(h^t \in \mathcal{H}^t\), and for all \(i \in \mathcal{N}\), we have \(\forall \pi_i|_{h^t} \in \Pi\)

\[
U_i(\tilde{\theta}^t, \pi_0|_{h^t}, \pi|_{h^t} \cdot 1_N) \geq U_i(\tilde{\theta}^t, \pi_0|_{h^t}, (\pi_i|_{h^t}, \pi|_{h^t} \cdot 1_{N-1})).
\]
Maximize the social welfare at the equilibrium in the worst case (with respect to different initial rating profiles)

\[
\max_{\tau,(\pi_0,\pi \cdot 1_N) \in \prod_{PA} \times \prod_{PA}^N} \min_{\theta^0 \in \Theta^N} \frac{1}{N} \sum_{i \in \mathcal{N}} U_i(\theta^0, \pi_0, \pi \cdot 1_N)
\]

s.t.

\[
(\pi_0, \pi \cdot 1_N) \text{ is a PAE.}
\]
Stationary mechanisms

Rating mechanisms with stationary recommended strategies:

- A fixed recommended plan after each rating distribution

Some simple and intuitive stationary recommended strategies:

- Use the fair plan (F) all the time
- Threshold-based stationary recommended strategies

An example threshold-based stationary strategy for a system with 5 users:

<table>
<thead>
<tr>
<th>State (rating distribution)</th>
<th>Recommend the altruistic plan (A) when there are at least 3 good users</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,5)</td>
<td></td>
</tr>
<tr>
<td>(1,4)</td>
<td></td>
</tr>
<tr>
<td>(2,3)</td>
<td></td>
</tr>
<tr>
<td>(3,2)</td>
<td>Recommend the selfish plan (S) when there are at least 3 bad users</td>
</tr>
<tr>
<td>(4,1)</td>
<td></td>
</tr>
<tr>
<td>(5,0)</td>
<td></td>
</tr>
</tbody>
</table>
Why do errors matter?

- Punishment is applied when it should not be.
- Incentives to follow are weakened (because users may get punished anyway, so why should users bother to follow?)
Strategy 1: always F

Rating mechanism 1: Recommend the fair plan all the time

Normalized social welfare goes to 0 when the error increases
Strategy 2: threshold-based (A+S)

Rating mechanism 2: Recommend altruistic and selfish plans (A+S)

Recommend = \begin{cases} 
\text{Altruistic, when } \# \text{ of good users no smaller than a threshold} \\
\text{Selfish, otherwise} 
\end{cases}

Given a discount factor, a rating update error, and the rating update rule, choose the optimal threshold:

\[
\max_\kappa \text{ Social welfare } \\
\text{subject to } \text{Sustainability}
\]
Strategy 2: threshold-based (A+S)

Rating mechanism 2: Recommend altruistic and selfish plans (A+S)

Recommend = \[
\begin{cases}
\text{Altruistic, when # of good users no smaller than a threshold} \\
\text{Selfish, otherwise}
\end{cases}
\]

Normalized social welfare goes to 0 when the error increases

Optimal threshold for each error probability

Performance loss due to wrongly-triggered punishments

Normalized social welfare goes to 0 when the error increases
Inefficiency is inevitable

Given the rating update error, can we achieve social optimum by optimizing the rating update rule when the discount factor goes to 1?

Intuitively:

Answer: No!

However:

Cannot use arbitrarily mild punishments, otherwise the users may want to deviate!
Strategies 3-4: A+F and F+S

Other threshold based recommended strategies:

A+F:

\[
\text{Recommend} = \begin{cases} 
\text{Altruistic, when # of good users no smaller than a threshold} \\
\text{Fair, otherwise}
\end{cases}
\]

F+S:

\[
\text{Recommend} = \begin{cases} 
\text{Fair, when # of good users no smaller than a threshold} \\
\text{Selfish, otherwise}
\end{cases}
\]
Inefficiency of threshold-based strategies

Performance of threshold-based recommended strategies:

Summary so far: stationary strategies are inefficient

Incentive provision requires severe enough punishments, which cause inefficiency
Inefficiency of threshold-based strategies

Summary so far: stationary strategies are inefficient

Reason: punishments are not refined enough
  • punishments depend only on the current rating distribution
  • to provide enough incentives, punishments are too often wrongly-triggered!

Solution: Nonstationary policies
  • punishments depend on the history of rating distributions → more refined punishments
  • adaptively adjust the frequency of punishments
A simple nonstationary strategy

Nonstationary policies using the altruistic and selfish actions only:

**Proposition**

Starting from any initial rating profile \( \theta \), the maximum social welfare achievable by \((\pi_0, \pi \cdot 1_N) \in \prod_{PA}(A^{as}) \times \prod_{PA}^N(A^{as})\) is at most

\[
b - c - c \cdot \rho(\theta, \alpha^{0*}_0, S^*_B) \sum_{s' \in S^*_B} q(s' | \theta, \alpha^{0*}_0, \alpha^a \cdot 1_N),
\]

where \( \alpha^{0*}_0 \), the optimal recommended action, and \( S^*_B \), the optimal subset of rating distributions, are the solutions to an optimization problem.

\( S^*_B \): the set of “bad” rating distributions in which the selfish plan is recommended as punishments

\( S^*_B \) empty \( \rightarrow \) always use the altruistic plan \( \rightarrow \) not an equilibrium

In an equilibrium strategy, \( S^*_B \) nonempty \( \rightarrow \) performance loss
A simple nonstationary strategy

Nonstationary policies using the altruistic and selfish actions only:

Proposition

Starting from any initial rating profile $\theta$, the maximum social welfare achievable by $(\pi_0, \pi \cdot 1_N) \in \prod_{PA}(A^{as}) \times \prod_{PA}^N(A^{as})$ is at most

$$b - c - c \sum_{s' \in S_B^*} \rho(\theta, \alpha_0^*, S_B^*) q(s' | \theta, \alpha_0^*, \alpha^a \cdot 1_N),$$

Always $> 0$

$= 0$ if and only if $S_B^*$ is empty

where $\alpha_0^*$, the optimal recommended action, and $S_B^*$, the optimal subset of rating distributions, are the solutions to an optimization problem.

$S_B^*$: the set of “bad” rating distributions in which the selfish plan is recommended as punishments

$S_B^*$ empty $\implies$ always use the altruistic plan $\implies$ not an equilibrium

In an equilibrium strategy, $S_B^*$ nonempty $\implies$ performance loss
Performance loss compared to the optimal rating mechanism:

Conclusion: we need nonstationary strategies with differential punishments! (Fair plan is needed!)
Socially-Optimal Strategic Design

**Theorem**

Given any rating update error $\varepsilon \in [0, 0.5)$,

- **(Design rating update rules):** A rating update rule $\tau(\varepsilon)$ that satisfies
  - **Condition 1:** $\beta_1^+ > 1 - \beta_1^-$ and $\beta_0^+ > 1 - \beta_0^-$,
  - **Condition 2:** $x_1^+ \triangleq (1 - \varepsilon)\beta_1^+ + \varepsilon(1 - \beta_1^-) > \frac{1}{1 + (N-1)\beta}$,
  - **Condition 3:** $x_0^+ \triangleq (1 - \varepsilon)\beta_0^+ + \varepsilon(1 - \beta_0^-) < \frac{1 - \beta_1^+}{(N-1)\beta}$.

can sustain optimal recommended strategies.

- **(Optimal recommended strategies):** Given the rating update rule $\tau(\varepsilon)$ that satisfies the above conditions and any small performance loss $\xi > 0$, for any discount factor $\delta$ no smaller than the lower-bound discount factor $\underline{\delta}(\varepsilon, \xi)$, we can construct a recommended strategy $\pi_0(\varepsilon, \xi, \delta) \in \Pi_f(\mathcal{A}_{afs})$, such that $(\pi_0(\varepsilon, \xi, \delta), \pi_0(\varepsilon, \xi, \delta) \cdot 1_N)$ is a PAE and achieves social welfare $b - c - \xi$, starting from any initial rating profile.
Theorem

Given any rating update error $\varepsilon \in [0, 0.5)$,

- **(Design rating update rules):** A rating update rule $\tau(\varepsilon)$ that satisfies
  - **Condition 1:** $\beta_1^+ > 1 - \beta_1^-$ and $\beta_0^+ > 1 - \beta_0^-$,
  - **Condition 2:** $x_1^+ \triangleq (1 - \varepsilon)\beta_1^+ + \varepsilon(1 - \beta_1^-) > \frac{1}{1 + (N-1)b}$,
  - **Condition 3:** $x_0^+ \triangleq (1 - \varepsilon)\beta_0^+ + \varepsilon(1 - \beta_0^-) < \frac{1 - \beta_1^+}{(N-1)b}$

  can sustain optimal recommended strategies.

- **(Optimal recommended strategies):** Given the rating update rule $\tau(\varepsilon)$ that satisfies the above conditions and any small performance loss $\xi > 0$, for any discount factor $\delta$ no smaller than the lower-bound discount factor $\delta(\varepsilon, \xi)$, we can construct a recommended strategy $\pi_0(\varepsilon, \xi, \delta) \in \Pi_f(A_{afs})$, such that $(\pi_0(\varepsilon, \xi, \delta), \pi_0(\varepsilon, \xi, \delta) \cdot 1_N)$ is a PAE and achieves social welfare $b - c - \xi$, starting from any initial rating profile.
Short intermezzo

MDPs vs. Repeated Games
Markov Decision Process (MDP)

Decision problem looks same at date $t, t + 1$

- State: $\omega_t \in \Omega$
- Action: $a \in A$
- New state: $\omega_{t+1}$

Stochastic transition mapping:

$\pi(\omega_{t+1} | \omega_t, a)$
Markov Decision Problem – familiar!

Bellman Principle

optimal policy \( f : \Omega \rightarrow A \)

\[
Eu(f) = u(f(\omega_0)) + \delta Eu(f(\omega_1)) + \delta^2 Eu(f(\omega_2)) + \ldots
\]

expectations taken with respect to transition probabilities

Optimal Value \( V(\omega_0) = \sup_{f \in \text{plans}} Eu(f) \)

If current state is \( \omega_t \),
choose action \( a^t \) to
maximize

\[
 u(a^t) + \delta EV(\omega_{t+1})
\]

- current payoff
- discounted future payoff
<table>
<thead>
<tr>
<th>MDPs (Bellman)</th>
<th>Repeated Games (Abreu, Pearce, Stacchetti, 1990)</th>
</tr>
</thead>
<tbody>
<tr>
<td>one agent</td>
<td>multi-agent</td>
</tr>
<tr>
<td>actions</td>
<td>action profiles</td>
</tr>
<tr>
<td>values</td>
<td>value profiles</td>
</tr>
<tr>
<td>optimal value functions</td>
<td>“optimal” value functions</td>
</tr>
<tr>
<td>single-valued</td>
<td>set-valued</td>
</tr>
<tr>
<td>optimal policy</td>
<td>optimal plan given plans of others (incentive compatibility)</td>
</tr>
</tbody>
</table>
Generalize with respect to arbitrary $W \subset \mathbb{R}^n$

Pair $(\alpha, V)$ is admissible with respect to $W$ if

- $\alpha \in A$ (action profile)
- $V : S \rightarrow W$ (continuation payoff function)
- incentive compatibility holds

Incentive compatibility (optimality relative to others)

$$u_i(\alpha, a_{-i}) + \delta EV_i(s|\alpha, a_{-i}) \geq u_i(\hat{\alpha}, a_{-i}) + \delta EV_i(s|\hat{\alpha}, a_{-i}), \ \forall \hat{\alpha}$$
Self-generating set: a set of payoff vectors in which every payoff vector can be decomposed by a plan profile, and the continuation payoff vector lies in the set. All payoffs in the self-generating set are equilibrium payoffs!
From self-generating set to PPE

Given $\mathcal{W}$ self-generating and bounded; $w \in \mathcal{W}$:

Construct PPE $\sigma$ with $U(\sigma) = w$

Construction (greedy policy!):

1. $w \in \mathcal{W} \subset B(\mathcal{W}) \Rightarrow w = u(a, V)$
   
   Set $\sigma(\emptyset) = a$

2. each $s \in S, V(s) = u(a', V')$
   
   Set $\sigma(s) = a'$

3. each $s \in S, \hat{s} \in S, V'(\hat{s}) = u(a'', V'')$
   
   Set $\sigma(s, \hat{s}) = a''$

\[ \vdots \]

Compute

- $u(\sigma | \text{any public history}) = \text{continuation value}$
- implies PPE
- implies $u(\sigma) = w$
Finding a self-generating set

But how do we get started?

Need $W$ self-generating Bootstrap?

any $W_0$

if $W_0 \subset B(W_0)$ done

if not

set $W_1 := W_0 \cup B(W_0)$

if $W_1 \subset B(W_1)$ done

if not

set $W_2 := W_1 \cup B(W_1)$

Define $W_\infty = \bigcup_{t=0}^{\infty} W_t$

actions, signals finite $\Rightarrow B(W_\infty) = W_\infty$

self-generating set
This doesn’t work

Problem:

• construction uses discounted payoffs
• these are finite if \( W \) bounded
• otherwise infinite
• unbounded \( W \nRightarrow \) PPE

\[ W_0 \subset W_1 \subset W_2 \subset \cdots \subset W_\infty \]

\[ \text{could be unbounded} \]
Work from “top down” (not bottom up)

Alternative: “value iteration”

Start with \( W \) such that all PPE payoffs \( \subseteq B(W) \subseteq W \) compact

Now iterate

\[
W_0 := W \\
W_1 := B(W_0) \\
W_2 := B(W_1) \\
\vdots
\]

Define \( W_\infty = \bigcap_{t=0}^{\infty} W_t \)

all PPE payoffs

But how do we find such a \( W \)?
Cautions

- How to find a self-generating set???
- APS powerful but not easy to apply
  - computation hard
- PPE strategies not unique
  - hard to find constructive algorithm
End intermezzo
Decompose the target payoff profile \( [U^0(s), U^1(s)]^T \) by \((\alpha_0, \alpha_0 \cdot 1_N)\)

- **decomposition:**
  \[
  U^\theta(s) = (1 - \delta) \cdot \left[ u^\theta(\alpha_0, \alpha \cdot 1_N) + \delta \cdot \sum_{s', \theta'} \Pr(s', \theta'| s, \theta, \alpha_0, \alpha \cdot 1_N) \gamma^\theta'(s') \right]
  \]

- **incentive constraints (IC):** for all \( \alpha' \in A \), we have \((\alpha' \triangleq (\alpha_0, \alpha', \alpha_0 \cdot 1_{N-1}))\)
  \[
  U^\theta(s) \geq (1 - \delta) \cdot \left[ u^\theta(\alpha') + \delta \cdot \sum_{s', \theta'} \Pr(s', \theta'| s, \theta, \alpha') \gamma^\theta'(s') \right]
  \]

**Recursive decomposition:**

- **continuation payoffs:** \( [\gamma^0(s'), \gamma^1(s')]^T \) can be decomposed, \( \forall s' \)
  \[
  \gamma^\theta'(s') = (1 - \delta) \cdot \left[ u^\theta'(\alpha_0, \alpha_0 \cdot 1_N) + \delta \cdot \sum_{s'', \theta''} \Pr(s'', \theta''| s', \theta', \alpha_0, \alpha_0 \cdot 1_N) \gamma^{\theta''}(s'') \right]
  \]

- **IC:** \( \gamma^\theta'(s') \geq (1 - \delta) \cdot \left[ u^\theta'(\alpha') + \delta \cdot \sum_{s'', \theta''} \Pr(s'', \theta''| s', \theta', \alpha') \gamma^{\theta''}(s'') \right] \)
Illustration of self-generating sets

Decompose the target payoff profile $[U^0(s), U^1(s)]^T$:

$$
U^\theta(s) = (1 - \delta) \cdot \left[ U^\theta(\alpha_0, \alpha \cdot 1_N) + \delta \cdot \sum_{s', \theta'} \Pr(s', \theta'|s, \theta, \alpha_0, \alpha \cdot 1_N) \gamma^{\theta'}(s') \right]
$$

The self-generating set:

- Payoff to decompose
- Decompose by the altruistic plan

Target payoff:

$$
(b - c, b - c)
$$

Continuation payoff when users have different ratings:

$$
(b - c - \epsilon_0, b - c - \epsilon_1)
$$

Continuation payoff when all the users have rating 1:

Continuation payoff when all the users have rating 0:

All continuation payoff vectors in the self-generating set. They should also be decomposable!

Recursive decomposition
Illustration of self-generating sets

For example, decompose $\begin{bmatrix} \gamma^0(s') \end{bmatrix}$.

$$
\gamma^\theta(s') = (1 - \delta) \cdot \left[ u^\theta(\alpha_0, \alpha_0 \cdot 1_N) + \sum_{s'', \theta''} \Pr(s'', \theta'' \mid s', \theta', \alpha_0, \alpha_0 \cdot 1_N) \gamma^\theta''(s'') \right]
$$

Payoff to decompose
Decompose by the fair plan

Rating 1

Rating 0

Continuation payoff when users have different ratings
Continuation payoff when all the users have rating 1
Continuation payoff when all the users have rating 0
Construct the recommended strategy

The algorithm:

Require: $\xi, \delta \geq \delta_0, s_\theta$

Initialization: $t = 0, v^0 = b - c - \epsilon_\theta$

repeat

if $s_1(\theta) = 0$ then
    if $v^0$ large then
        $\alpha^f_0 = \alpha^f = \alpha^a$, update $v^0$ and $v^1$
    else
        $\alpha^f_0 = \alpha^f = \alpha^s$, update $v^0$ and $v^1$
    end
elseif $s_1(\theta) = N$ then
    if $v^1$ large then
        $\alpha^f_0 = \alpha^f = \alpha^a$, update $v^0$ and $v^1$
    else
        $\alpha^f_0 = \alpha^f = \alpha^s$, update $v^0$ and $v^1$
    end
else
    if $v^1$ close to $v^0$ then
        $\alpha^f_0 = \alpha^f = \alpha^a$, update $v^0$ and $v^1$
    else
        $\alpha^f_0 = \alpha^f = \alpha^f$, update $v^0$ and $v^1$
    end
end
$t \leftarrow t + 1$
until $\emptyset$

Input: performance loss tolerance, a feasible discount factor, the initial state

Set the target payoff

1. Determine the recommended plan based on the current rating distribution and the continuation payoff
2. Update the continuation payoff analytically

May choose different plans under the same rating distribution

Nonstationary
Comparisons with Stationary policies

- benefit = 3, cost = 1, # of users = 10
- A stationary recommended strategy:
  - the altruistic plan (A) when at least half of the users have good rating
  - the fair plan (F) when less than half of the users have good rating
- Comparison of a sample path in the first few periods

<table>
<thead>
<tr>
<th>Period</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating distribution</td>
<td>(0,10)</td>
<td>(1,9)</td>
<td>(3,7)</td>
<td>(7,3)</td>
<td>(5,5)</td>
<td>(7,3)</td>
<td>(4,6)</td>
<td>(2,8)</td>
</tr>
<tr>
<td>Recommended plan</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>F</td>
<td>A</td>
<td>F</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

Stationary:

<table>
<thead>
<tr>
<th>Period</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating distribution</td>
<td>(0,10)</td>
<td>(1,9)</td>
<td>(3,7)</td>
<td>(7,3)</td>
<td>(5,5)</td>
<td>(7,3)</td>
<td>(4,6)</td>
<td>(2,8)</td>
</tr>
<tr>
<td>Recommended plan</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>F</td>
<td>A</td>
<td>F</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

Optimal: (Non-stationary)

<table>
<thead>
<tr>
<th>Period</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating distribution</td>
<td>(0,10)</td>
<td>(1,9)</td>
<td>(3,7)</td>
<td>(7,3)</td>
<td>(5,5)</td>
<td>(7,3)</td>
<td>(2,8)</td>
<td>(1,9)</td>
</tr>
<tr>
<td>Recommended plan</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>F</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

Optimal policy does not punish because the continuation payoffs are low. Intuition: users have been punished in the past (F in period 3)
Performance improvements

- Threshold-based stationary recommended strategies that use two plans:
  - e.g., use A when the number of good users is large, use F otherwise
  - optimal threshold for each rating update error
Robustness of proposed mechanisms

What if we have an inaccurate estimation of the rating update error?

performance gain/loss (in percentage):

- Less than 5% performance variance when the inaccuracy < 50%
- Larger performance variance under larger rating update errors

Inaccuracy of the estimation (in percentage): $\frac{\hat{e} - \varepsilon}{\varepsilon} \times 100\%$
## Related Works

<table>
<thead>
<tr>
<th></th>
<th>Rigorous analysis</th>
<th>Recommended strategy</th>
<th>Rating update errors</th>
<th>Discount factor</th>
<th>Performance loss due to imperfect reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohen’2003, Stoica’2006, etc.</td>
<td>No</td>
<td>Stationary</td>
<td>&gt; 0</td>
<td>&lt; 1</td>
<td>Yes</td>
</tr>
<tr>
<td>Kandori’1992, Takahashi’2010 Deb’2013, etc.</td>
<td>Yes</td>
<td>Stationary</td>
<td>= 0</td>
<td>→ 1</td>
<td>Yes</td>
</tr>
<tr>
<td>Ellison’1994</td>
<td>Yes</td>
<td>Stationary</td>
<td>→ 0</td>
<td>→ 1</td>
<td>Yes</td>
</tr>
<tr>
<td>Dellarocas’2005</td>
<td>Yes</td>
<td>Stationary</td>
<td>&gt; 0</td>
<td>&lt; 1</td>
<td>Yes</td>
</tr>
<tr>
<td>Design Stat. Zhang and vd Schaar’2012</td>
<td>Yes</td>
<td>Stationary</td>
<td>→ 0</td>
<td>&lt; 1</td>
<td>Yes</td>
</tr>
<tr>
<td>Optimal Design Xiao and vd Schaar’2013</td>
<td>Yes</td>
<td>Nonstationary</td>
<td>&gt; 0</td>
<td>&lt; 1</td>
<td>No</td>
</tr>
</tbody>
</table>
Extensions and Beyond

• **Multiple ratings**

• **Collective ratings**

• **Robustness of norms**

• **Social norms on networks**
Extensions and Beyond

• **Tokens vs. Ratings**

• **Adverse selection**

• **Ratings + adverse selection + moral hazard + endogenous matching**
  Y. Xiao, F. Dörfler and M. van der Schaar, "Incentive Design in Peer Review: Rating and Repeated Endogenous Matching"

With William Zame – see talk on Friday!
Social norms – broader impact

Two recent columns in the NYTimes have highlighted our work on ratings and reputations in social/societal networks.

Cyber-security:
http://op-talk.blogs.nytimes.com/2014/09/21/can-we-build-a-safer-internet/ discusses how the work can be used to promote a safer Internet and in particular to discourage malicious users.

Control “social” behavior on the Internet:
http://op-talk.blogs.nytimes.com/2015/03/03/an-easier-way-to-fight-bullying/ discusses how the work can be used to control social bullying behavior in societies.