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:**

		Worker	
		S	NS
Requester		$b, \quad -c$	$0, \quad 0$

$b - c > 0$

Individual vs. society goals!

Incentives needed!

Canonical (Gift-Giving) Game

- **Requester always pays the same amount q (flat-rate pricing)**
  - **The worker receives μq.**
  - **The website charges $(1-\mu)q$ as the transaction fee.**
- **Actions:**

μ and q are design choices!!

  - **Requester: no action to choose**
  - **Worker: $a \in \mathcal{A} = \{S, NS\}$**
    - **S: High level of effort**
    - **NS: Low level of effort**
- **Game:**

	Worker: S	Worker: NS
Requester	$b-q, \ \mu q - c$	$-q, \mu q$

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 ...

Good	Bad
Can get full cooperation	
Can get “anything”	Can get “anything”
	Players must be patient
	Full information

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

Design problem

$$\text{maximize} \quad \text{Designer's objective}$$

$$\text{s.t.} \quad \underset{\text{Social norm}}{\text{Consistent with self-interested behavior of entities}}$$

- Selfish agents <u>**want**</u> to follow the protocol **=** **incentive compatibility = sustainability**

~~Network Utility Maximization~~ **~~Control~~**

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

$\boxed{\text{imperfect monitoring}}$

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

public monitoring

- all players observe same s
- different players observe different s

private monitoring

We focus on imperfect public monitoring

Using the information

Example

H

L

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,"
ACM Transactions on Economics and Computation (TEAC), 2015.

Service/Resource Exchange Systems

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

System Model

(Stage-game) Model:

- A gift-giving game between a client and a server

	high effort	low effort
request	$(b, -c)$	$(0, 0)$

- A matching: $m : \mathcal{N} \to \mathcal{N}, i \text{ (server)} \mapsto m(i) \text{ (client)}$
- Set of proper matchings:
  $M = \{m : m \text{ bijective and } m(i) \neq i, \forall i \in \mathcal{N}\}$
- A matching rule: $\mu : M \to \Delta(M)$
- Focus on uniform random matching

Rating mechanisms:

- Assign each user i with a rating $\theta_i \in \Theta$
- Rating profile (Unknown to users): $\boldsymbol{\theta} \in \Theta^N$
- Rating distribution (Known to users):
  $\mathbf{s}(\boldsymbol{\theta})$

How does the rating mechanism function?

In each period t:

- The platform displays $\mathbf{s}(\theta)$, 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 $\alpha^{\mathrm{a}}(\theta_c,\theta_s)=1, \forall\theta_c,\theta_s\in\{0,1\}$
- *Selfish* plan $\alpha^{\mathrm{s}}(\theta_c,\theta_s)=0, \forall\theta_c,\theta_s\in\{0,1\}$

Cannot achieve the social optimum

- *Fair* plan $\alpha^{\mathrm{f}}(\theta_c,\theta_s)=\begin{cases}0 & \theta_s>\theta_c\\ 1 & \theta_s\le\theta_c\end{cases}$
- Erroneous report (with error probability ε)

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

- Rating update rule

Rating goes up when the reported effort level exceeds the recommended effort level

$$\tau(\theta_s'|\theta_c,\theta_s,z)=\begin{cases}\beta^+_{\theta_s}, & \theta_s'=1, z\ge\alpha_0(\theta_c,\theta_s)\\ 1-\beta^+_{\theta_s}, & \theta_s'=0, z\ge\alpha_0(\theta_c,\theta_s)\\ 1-\beta^-_{\theta_s}, & \theta_s'=1, z<\alpha_0(\theta_c,\theta_s)\\ \beta^-_{\theta_s}, & \theta_s'=0, z<\alpha_0(\theta_c,\theta_s)\end{cases}, \text{ for } \theta_s=0,1.$$

"Simple" rating mechanisms

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

Can achieve the social optimum

- Erroneous report (with error probability ε)

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

- Rating update rule

Rating goes up when the reported effort level exceeds the recommended effort level

$$\tau(\theta_s'|\theta_c,\theta_s,z)=\begin{cases}\beta^{+}_{\theta_s}, & \theta_s'=1, z\ge\alpha_0(\theta_c,\theta_s)\\ 1-\beta^{+}_{\theta_s}, & \theta_s'=0, z\ge\alpha_0(\theta_c,\theta_s)\\ 1-\beta^{-}_{\theta_s}, & \theta_s'=1, z<\alpha_0(\theta_c,\theta_s)\\ \beta^{-}_{\theta_s}, & \theta_s'=0, z<\alpha_0(\theta_c,\theta_s)\end{cases}, \text{ for } \theta_s=0,1.$$

“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 $\alpha^{f}(\theta_c, \theta_s) = \begin{cases} 0 & \theta_s > \theta_c \\ 1 & \theta_s \le \theta_c \end{cases}$

Serve everybody

Serve nobody

mum
e

Only serve users with higher or equal ratings

- Erroneous report (with error probability ε)

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

- Rating update rule

Rating goes up when the reported effort level exceeds the recommended effort level

$$\tau(\theta_s' | \theta_c, \theta_s, z) = \begin{cases} \beta^{+}_{\theta_s}, & \theta_s' = 1, z \ge \alpha_0(\theta_c, \theta_s) \\ 1 - \beta^{+}_{\theta_s}, & \theta_s' = 0, z \ge \alpha_0(\theta_c, \theta_s) \\ 1 - \beta^{-}_{\theta_s}, & \theta_s' = 1, z < \alpha_0(\theta_c, \theta_s) \\ \beta^{-}_{\theta_s}, & \theta_s' = 0, z < \alpha_0(\theta_c, \theta_s) \end{cases}, \text{ for } \theta_s = 0, 1.$$

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)

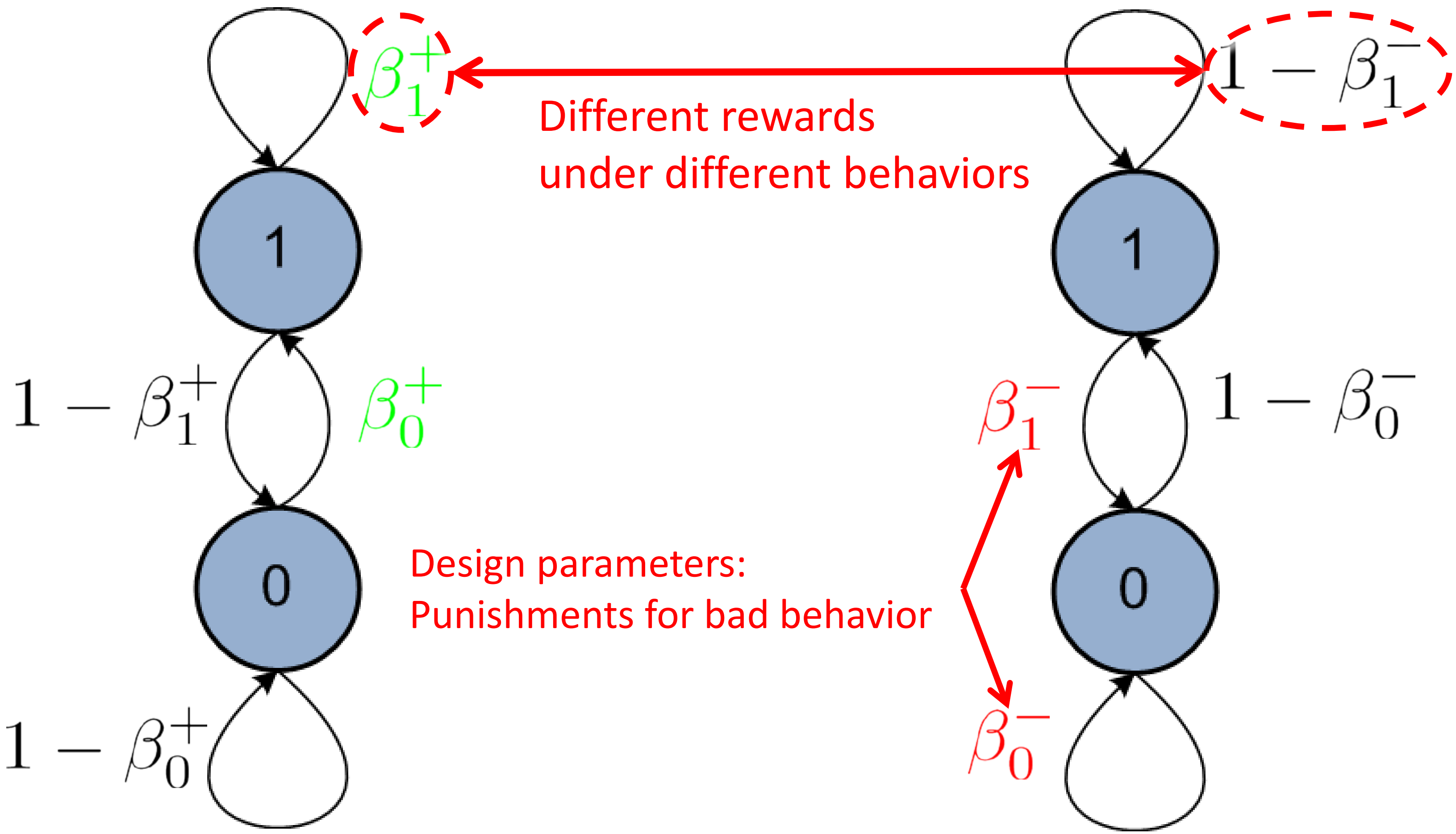
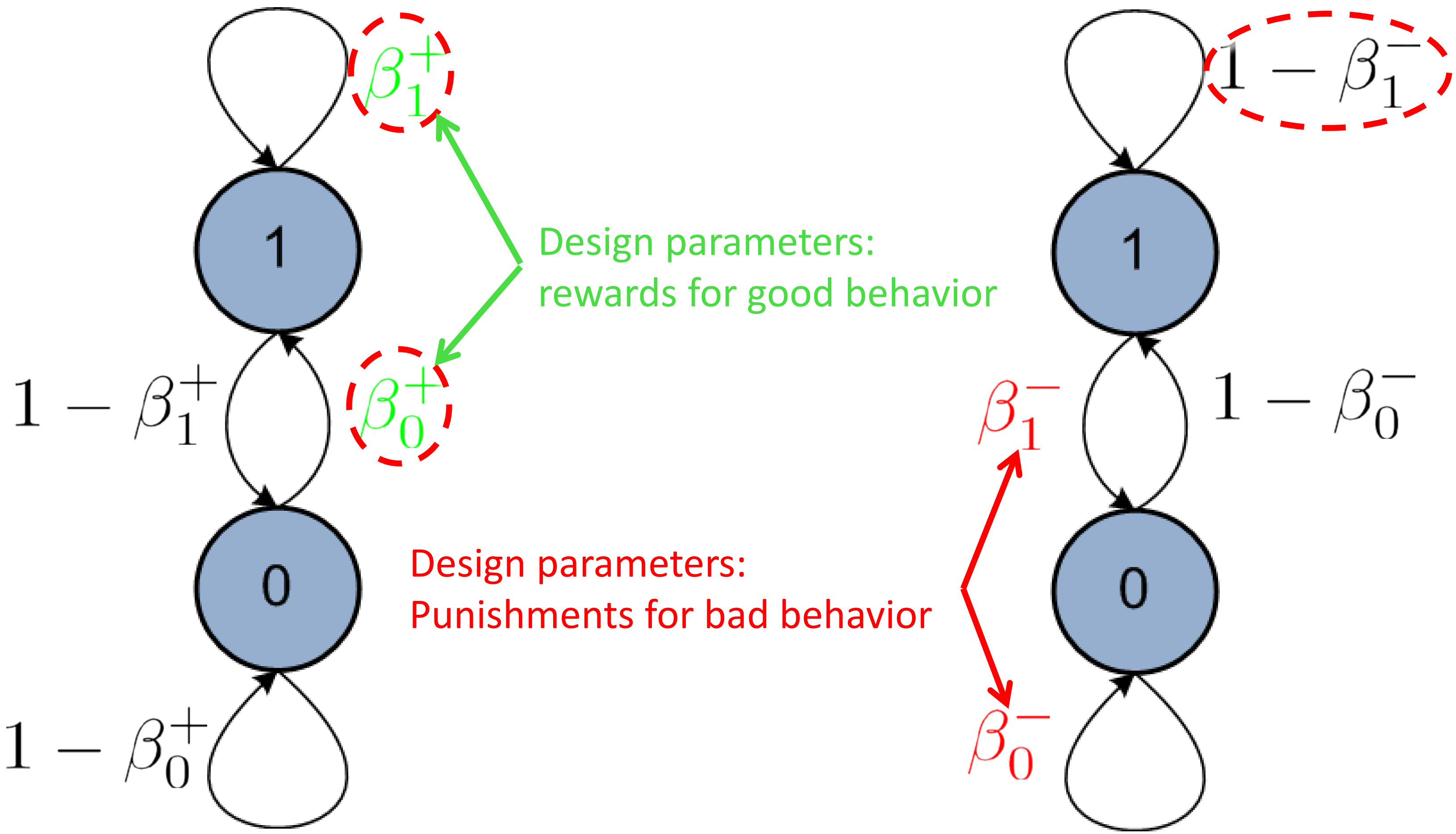


Illustration of rating update rules

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)



Stochastic game formulation

Stochastic game:

- Players: the users and the platform $\mathcal{N} \cup \{0\}$
- State: rating profile $\boldsymbol{\theta} \in \Theta^N$
- Action set (set of plans): $A \triangleq \{\alpha | \alpha : \Theta \times \Theta \rightarrow \{0,1\}\}$
- Stage-game payoff: $u_i(\boldsymbol{\theta}, \alpha_0, \boldsymbol{\alpha})$
- History at period t: $\mathbf{h}^t = (\boldsymbol{\theta}^0, \ldots, \boldsymbol{\theta}^t) \in \mathcal{H}^t$
- Strategy: $\pi_i \in \Pi : \sum_{t=0}^{\infty} \mathcal{H}^t \rightarrow A, i = 0, 1, \ldots, N$
- Strategy profile $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_N)$
- Recommended plan α_0 and Recommended strategy π_0
- Overall payoff

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

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 rating profiles
  - 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 π is a public announcement strategy, if for all $t \geq 0$ and for all $\mathbf{h}^t, \tilde{\mathbf{h}}^t \in \mathcal{H}^t$, we have

$$\pi(\mathbf{h}^t) = \pi(\tilde{\mathbf{h}}^t), \text{ if } \mathbf{s}(\theta^k) = \mathbf{s}(\tilde{\theta}^k), \; k = 0, 1, \ldots, t.$$

We write the set of all PA strategies as Π_{PA}.

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

Examples of subset B:
B={altruistic, fair, selfish},
or B={altruistic, selfish}

Which equilibrium to use?

Public Equilibrium σ

- each σ_i depends only on history of public signals
- each σ_i optimal given σ_{-i} and information

Public Perfect Equilibrium (PPE)

- each σ_i depends only on history of public signals
- each σ_i optimal given σ_{-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)

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

Definition (Equilibrium Definition)

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

$$U_i(\tilde{\theta}^t, \pi_0|_{\tilde{\mathbf{h}}^t}, \pi|_{\tilde{\mathbf{h}}^t} \cdot \mathbf{1}_N) \geq U_i(\tilde{\theta}^t, \pi_0|_{\tilde{\mathbf{h}}^t}, (\pi_i|_{\tilde{\mathbf{h}}^t}, \pi|_{\tilde{\mathbf{h}}^t} \cdot \mathbf{1}_{N-1})).$$

- 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

Rating mechanism design problem

Maximize the social welfare at the equilibrium in the worst case (with respect to different initial rating profiles)

$$\max_{\tau,(\pi_0,\pi\cdot\mathbf{1}_N)\in\Pi_{PA}\times\Pi_{PA}^N} \quad \min_{\theta^0\in\Theta^N} \quad \frac{1}{N}\sum_{i\in\mathcal{N}} U_i(\theta^0,\pi_0,\pi\cdot\mathbf{1}_N)$$

$$s.t. \qquad (\pi_0,\pi\cdot\mathbf{1}_N) \text{ is a PAE.}$$

Stationary mechanisms

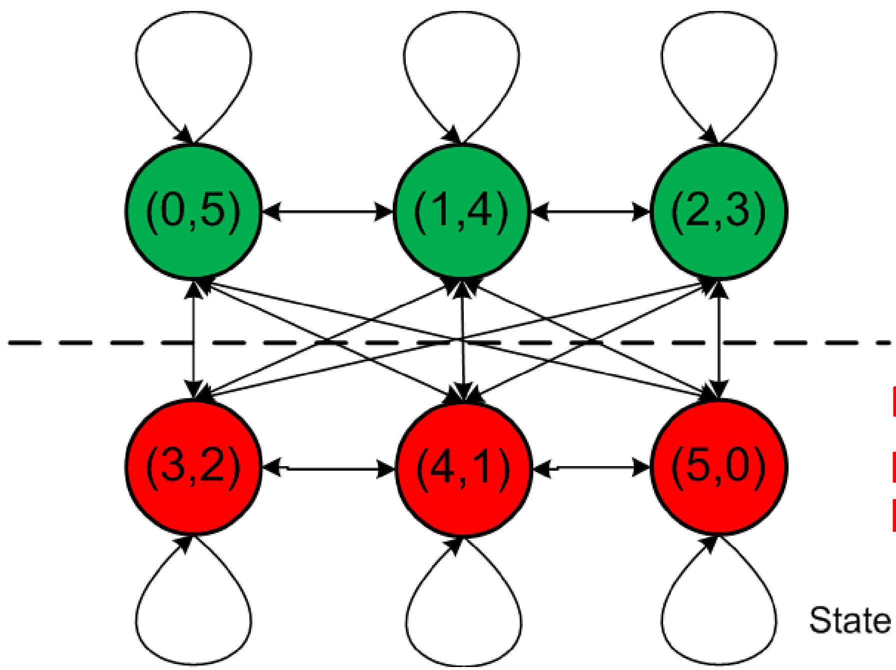
Rating mechanisms with stationary recommended strategies:

- 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:



Recommend the altruistic plan (A) when there are at least 3 good users

Recommend the selfish plan (S) when there are at least 3 bad users

State (rating distribution):

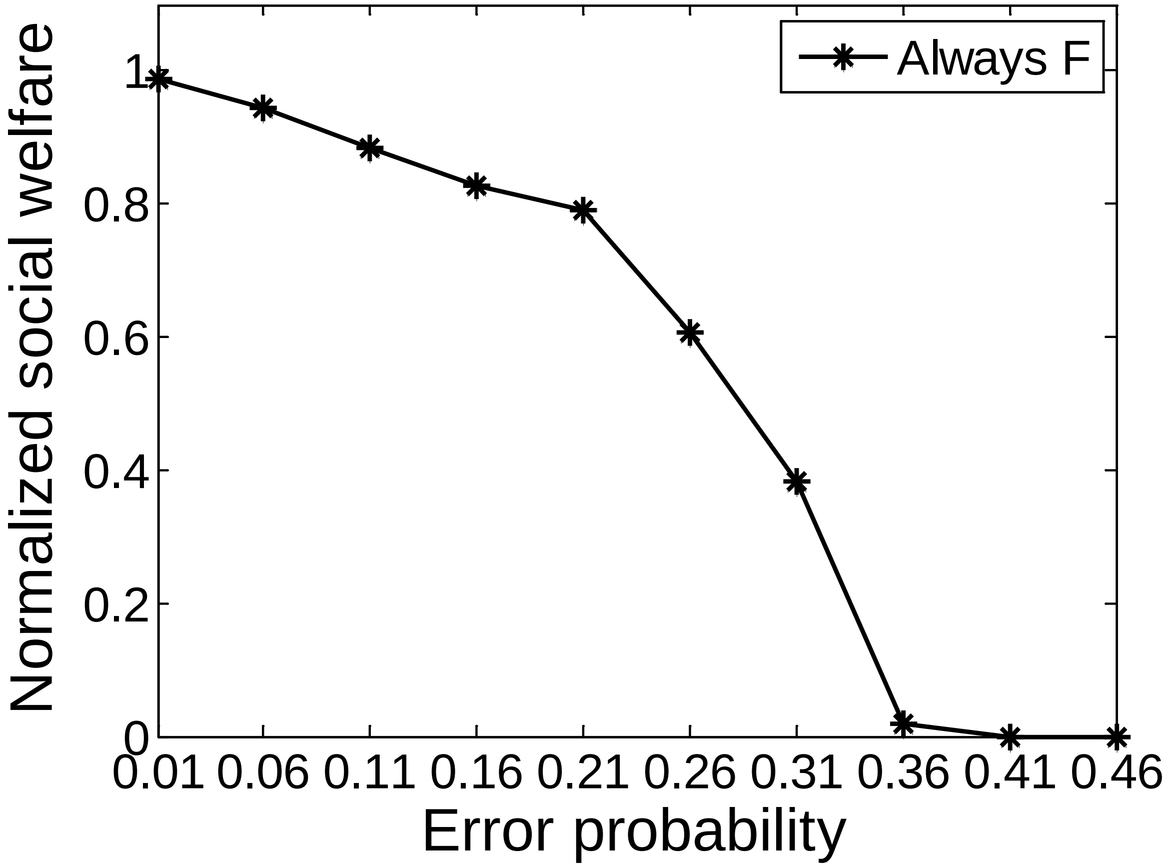
State transition:

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)

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

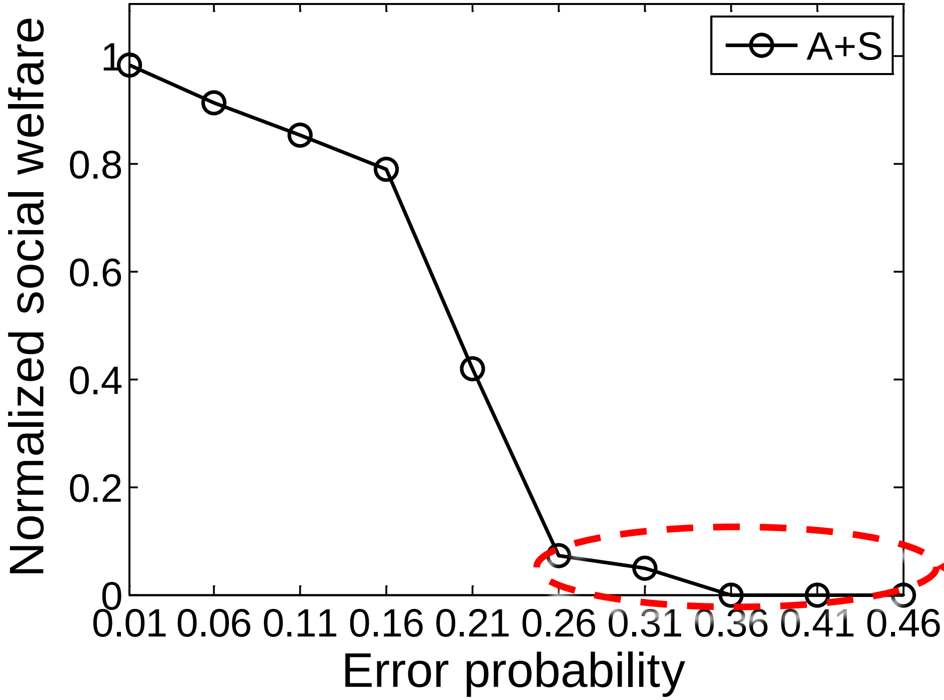
Given a discount factor, a rating update error, and the rating update rule, choose the optimal threshold:

$$\max_{\kappa} \quad \text{Social welfare}$$
$$\text{subject to} \quad \mathit{Sustainability}$$

Strategy 2: threshold-based (A+S)

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

Recommend = { Altruistic, when # of good users no smaller than a threshold
Selfish, otherwise



← 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?

Answer: No!

Intuitively:

Discount factor ↑ → Use milder punishments (by adjusting the rating update rule) → Performance loss ↓

However:

Use milder punishments (by adjusting the rating update rule) → Incentive ↓

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:

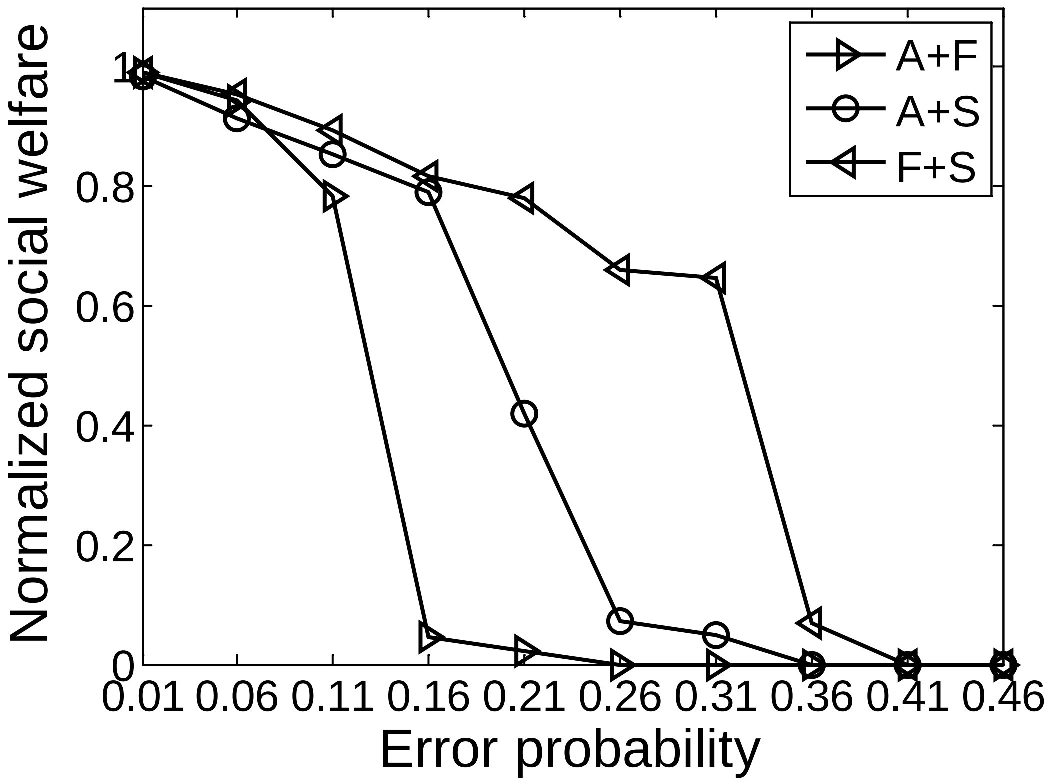
$$\text{Recommend} = \begin{cases} \text{Altruistic, when \# of good users no smaller than a threshold} \\ \text{Fair, otherwise} \end{cases}$$

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 $\boldsymbol{\theta}$, the maximum social welfare achievable by $(\pi_0, \pi \cdot \mathbf{1}_N) \in \Pi_{PA}(A^{\mathrm{as}}) \times \Pi_{PA}^{N}(A^{\mathrm{as}})$ is at most

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

where α_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 → always use the altruistic plan → not an equilibrium

In an equilibrium strategy, S_B^* nonempty → performance loss

A simple nonstationary strategy

Nonstationary policies using the altruistic and selfish actions only:

Proposition

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

$$b - c - c\,\underbrace{\rho(\boldsymbol{\theta}, \alpha_0^*, S_B^*)}_{\text{Always} > 0}\,\underbrace{\sum_{\mathrm{s}' \in S_B^*} q(\mathrm{s}' | \boldsymbol{\theta}, \alpha_0^*, \alpha^{\mathrm{a}} \cdot \mathbf{1}_N)}_{= 0 \text{ if and only if } S_B^* \text{ is empty}},$$

where α_0^*, the optimal recommended action, and S_B^*, the optimal subset of rating distributions, are the solutions to an optimization problem.

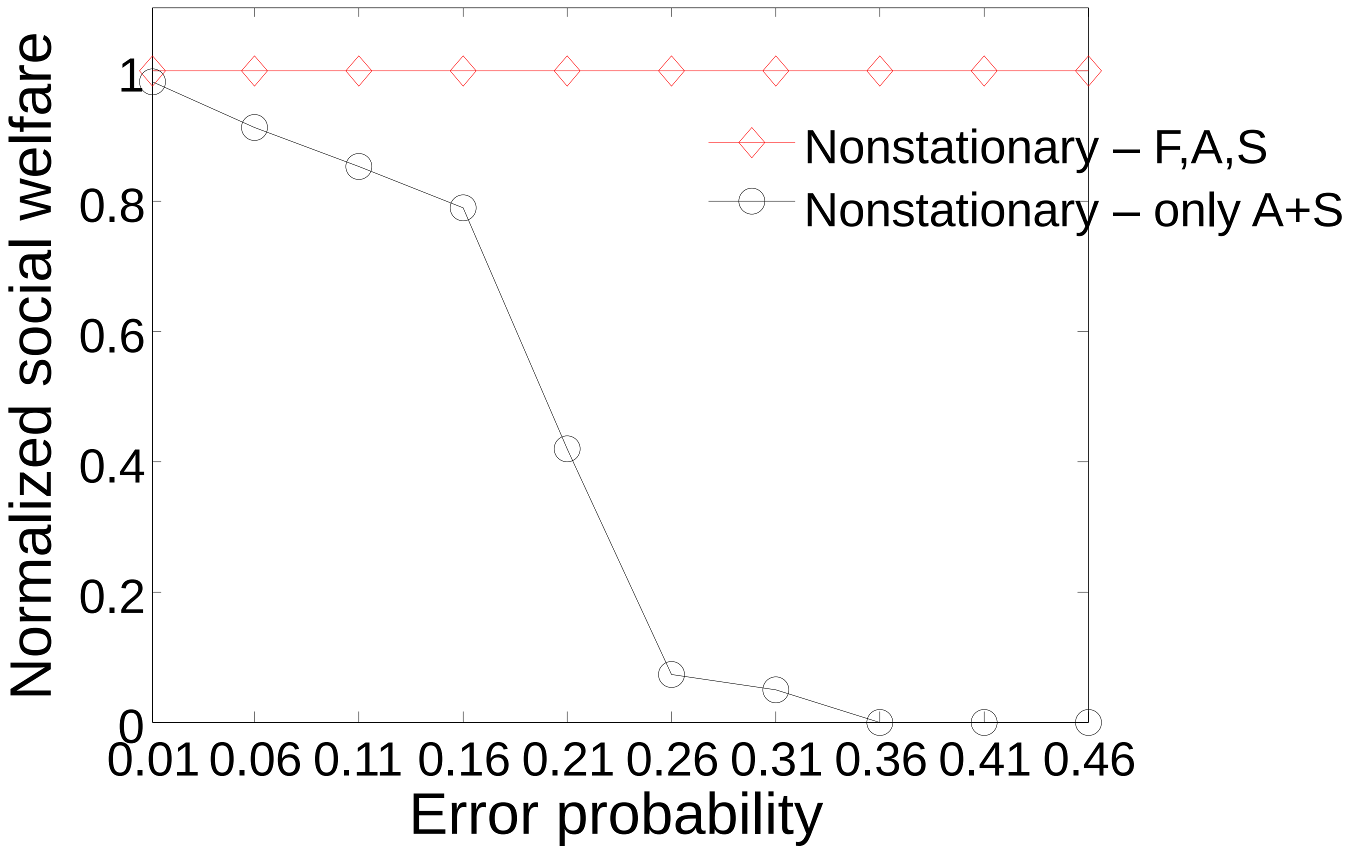
S_B^*: the set of “bad” rating distributions in which the selfish plan is recommended as punishments

S_B^* empty → always use the altruistic plan → not an equilibrium

In an equilibrium strategy, S_B^* nonempty → performance loss

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:* $x_1^+ \triangleq (1-\varepsilon)\beta_1^+ + \varepsilon(1-\beta_1^-) > \frac{1}{1+\frac{c}{(N-1)b}}$,
  - *Condition 3:* $x_0^+ \triangleq (1-\varepsilon)\beta_0^+ + \varepsilon(1-\beta_0^-) < \frac{1-\beta_1^+}{\frac{c}{(N-1)b}}$.

  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* δ *no smaller than the lower-bound discount factor* $\underline{\delta}(\varepsilon, \xi)$, *we can construct a recommended strategy* $\pi_0(\varepsilon, \xi, \delta) \in \Pi_f(A^{\mathrm{afs}})$, *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.*

Design guidelines

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:* $x_1^+ \triangleq (1-\varepsilon)\beta_1^+ + \varepsilon(1-\beta_1^-) > \frac{1}{1+\frac{c}{(N-1)b}}$,
  - *Condition 3:* $x_0^+ \triangleq (1-\varepsilon)\beta_0^+ + \varepsilon(1-\beta_0^-) < \frac{1-\beta_1^+}{\frac{c}{(N-1)b}}$.

  Design guidelines

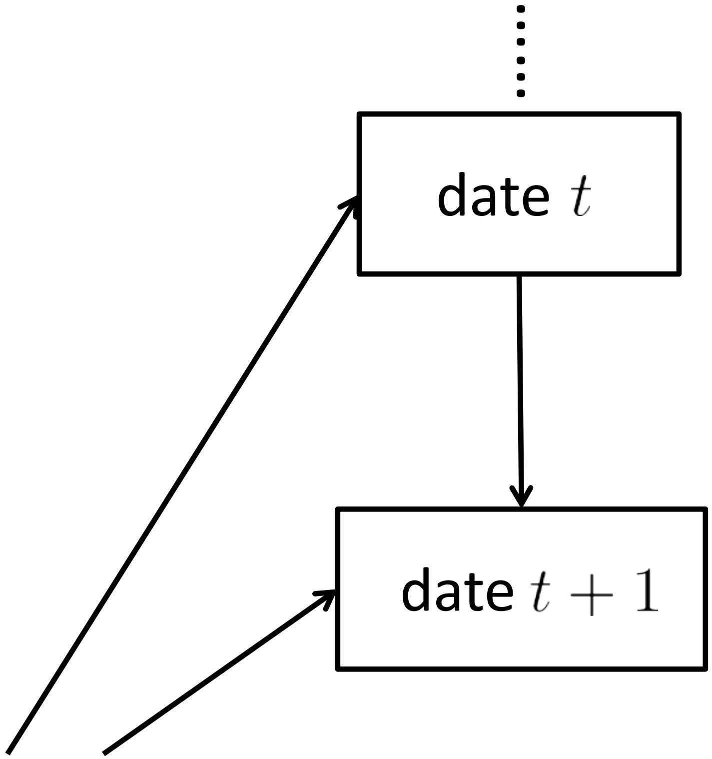
  can sustain optimal recommended strategies.
- *(Optimal recommended strategies): Given the rating update rule* $\tau(\varepsilon)$ *that satisfies the above conditions and any small performance loss* $\xi > 0$, *for any discount factor* δ *no smaller than the lower-bound discount factor* $\underline{\delta}(\varepsilon, \xi)$, *we can construct a recommended strategy* $\pi_0(\varepsilon, \xi, \delta) \in \Pi_f(A^{\mathrm{afs}})$, *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.*

Analytically determined

Short intermezzo

MDPs vs. Repeated Games

Markov Decision Process (MDP)



state $\omega_t \in \Omega$

action $a \in A$

new state ω_{t+1}

$$\pi(\omega_{t+1}|\omega_t, a)$$

stochastic transition mapping

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

Markov Decision Problem – familiar!

Bellman Principle

optimal policy $f : \Omega \to A$

$$
\begin{aligned}
Eu(f) = u(f(\omega_0))& \\
+\delta Eu(f(\omega_1))& \\
+\delta^2 Eu(f(\omega_2))& \\
\vdots&
\end{aligned}
$$

expectations taken with respect to transition probabilities

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

Bellman Principle for MDPs

If current state is ω_t,

choose action a^t to

maximize

$$\underbrace{u(a^t)}_{\text{current payoff}} \quad + \quad \underbrace{\delta EV(\omega_{t+1})}_{\text{discounted future payoff}}$$

Bellman Principle: MDPs vs Repeated Games

MDPs (Bellman)	Repeated Games (Abreu, Pearce, Stacchetti, 1990)
one agent	multi-agent
actions	action profiles
values	value profiles
optimal value functions single-valued	“optimal” value functions set-valued
optimal policy	optimal plan given plans of others (incentive compatibility)

Bellman Principle in Repeated Games 1

Generalize with respect to arbitrary $W \subset R^n$

Pair (a, V) is admissible with respect to *W* if

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

Incentive compatibility (optimality relative to others)

$$\begin{aligned} u_i(a_i, a_{-i}) + \delta E V_i(s|a_i, a_{-i}) \\ \geq u_i(\hat{a}_i, a_{-i}) + \delta E V_i(s|\hat{a}_i, a_{-i}), \ \forall \hat{a}_i \end{aligned}$$

Bellman Principle in Repeated Games 2

$$\boxed{B(W) = \{\text{utility of } (a, V) : (a, V) \text{ admissible}, V \in W\}}$$

W **self-generating:** $W \subset B(W)$

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 W self-generating and bounded; $w \in W$:

Construct PPE σ with $U(\sigma) = w$

Construction (greedy policy!):

(1) $w \in W \subset B(W)$ ➔ $w = u(a, V)$

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 $\sigma(s, \hat{s}) = a''$

⋮

Compute

- $u(\sigma | \text{any public history})$ = 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 $W_2 := W_1 \cup B(W_1)$

$\vdots$

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

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 $W \not\Rightarrow$ PPE

$$W_0 \subset W_1 \subset W_2 \subset \cdots \subset W_\infty$$

could be unbounded

Work from "top down" (not bottom up)

Alternative: "value iteration"

Start with W such that

$$\text{all PPE payoffs} \subset B(W) \subset W \ \text{compact}$$

Now iterate

$$\begin{aligned} W_0 &:= W \\ W_1 &:= B(W_0) \\ W_2 &:= B(W_1) \\ &\vdots \end{aligned}$$

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

all PPE payoffs

$\boxed{\text{But how do we find such a } W\text{?}}$

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

Socially-Optimal Design – Key ideas

- Decompose the target payoff profile $[U^0(\boldsymbol{s}), U^1(\boldsymbol{s})]^T$ by $(\alpha_0, \alpha_0 \cdot \mathbf{1}_N)$
  - decomposition:

$$\underbrace{U^\theta(\boldsymbol{s})}_{\text{Target payoff}} = (1-\delta)\cdot\left[\underbrace{u^\theta(\alpha_0,\alpha\cdot\mathbf{1}_N)}_{\text{Instantaneous payoff}} + \delta\cdot\sum_{\boldsymbol{s}',\theta'} \Pr(\boldsymbol{s}',\theta'|\boldsymbol{s},\theta,\alpha_0,\alpha\cdot\mathbf{1}_N)\underbrace{\gamma^{\theta'}(\boldsymbol{s}')}_{\text{Continuation payoff}}\right]$$

  - incentive constraints (IC): for all $\alpha' \in A$, we have ($\boldsymbol{\alpha}' \triangleq (\alpha_0, \alpha', \alpha_0\cdot\mathbf{1}_{N-1})$)

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

- *Recursive* decomposition:
  - continuation payoffs $[\gamma^0(\boldsymbol{s}'),\gamma^1(\boldsymbol{s}')]^T$ can be decomposed, $\forall\ \boldsymbol{s}'$

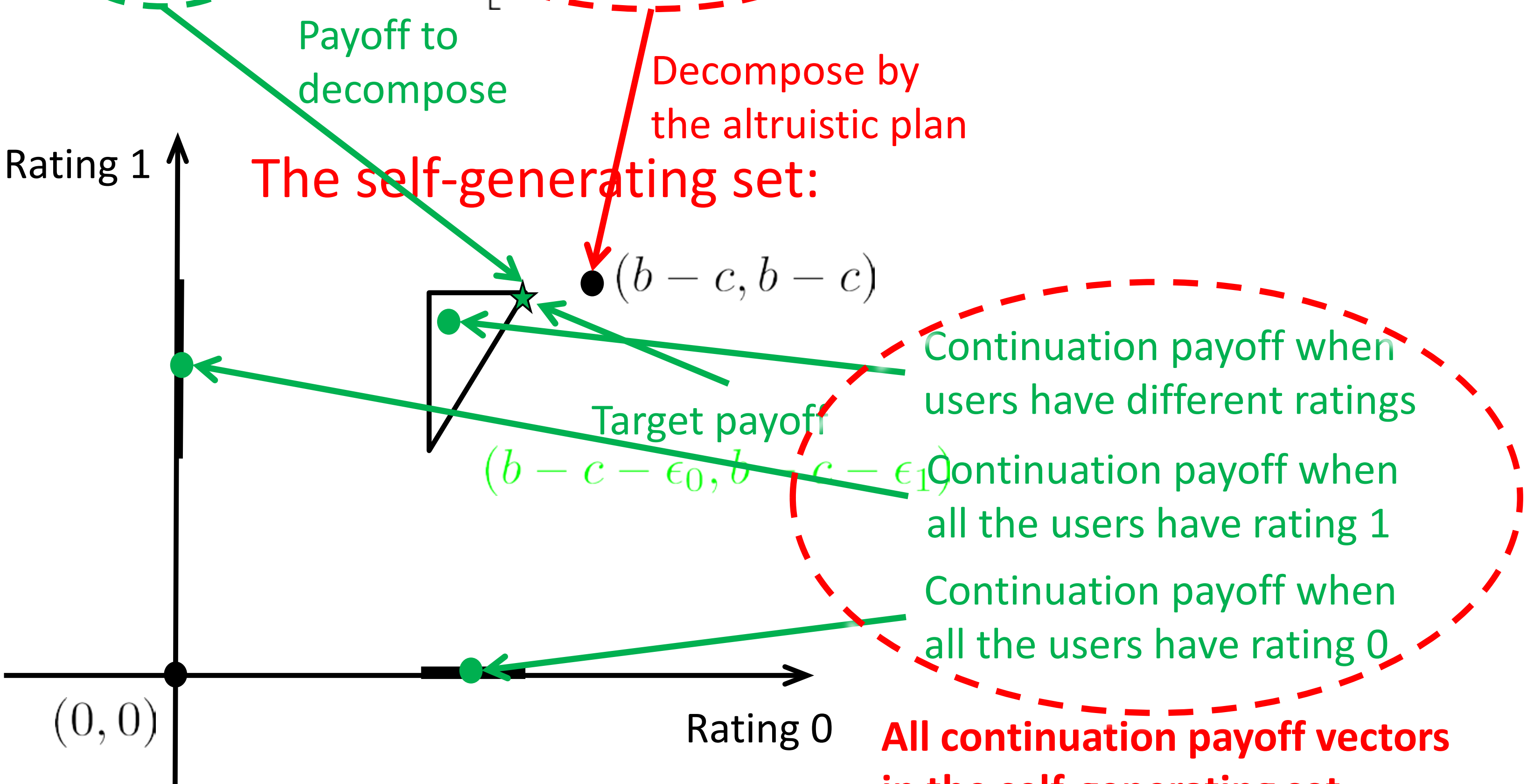
$$\gamma^{\theta'}(\boldsymbol{s}') = (1-\delta)\cdot\left[u^{\theta'}(\alpha_0,\alpha_0\cdot\mathbf{1}_N)+\delta\cdot\sum_{\boldsymbol{s}'',\theta''}\Pr(\boldsymbol{s}'',\theta''|\boldsymbol{s}',\theta',\alpha_0,\alpha_0\cdot\mathbf{1}_N)\hat{\gamma}^{\theta''}(\boldsymbol{s}'')\right]$$

  - IC: $\gamma^{\theta'}(\boldsymbol{s}') \ge (1-\delta)\cdot\left[u^{\theta'}(\boldsymbol{\alpha}')+\delta\cdot\sum_{\boldsymbol{s}'',\theta''}\Pr(\boldsymbol{s}'',\theta''|\boldsymbol{s}',\theta',\boldsymbol{\alpha}')\gamma^{\theta''}(\boldsymbol{s}'')\right]$

Illustration of self-generating sets

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

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



All continuation payoff vectors in the self-generating set.

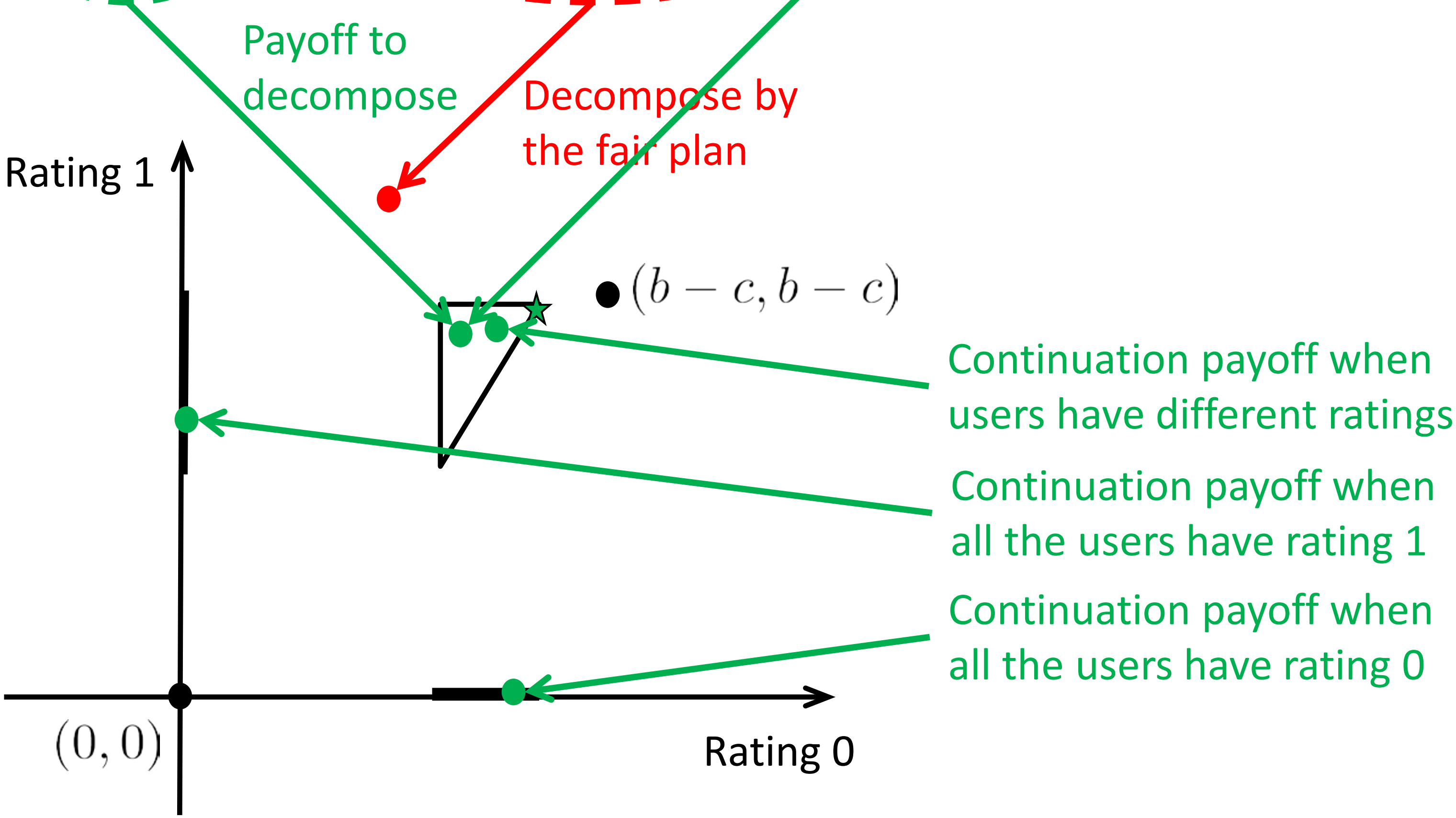
They should also be decomposable!

Recursive decomposition

Illustration of self-generating sets

For example, decompose $[\gamma^0(\boldsymbol{s}'), \gamma^1(\boldsymbol{s}')]^T$:

$$\gamma^{\theta'}(\boldsymbol{s}') = (1-\delta)\cdot[u^{\theta'}(\alpha_0, \alpha_0\cdot\mathbf{1}_N) + \delta\cdot\sum_{\boldsymbol{s}'',\theta''} \Pr(\boldsymbol{s}'', \theta''|\boldsymbol{s}', \theta', \alpha_0, \alpha_0\cdot\mathbf{1}_N)\hat{\gamma}^{\theta''}(\boldsymbol{s}'')]$$



Construct the recommended strategy

The algorithm:

Require: ξ, $\delta \geq \underline{\delta}$, s_θ

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

```
repeat
    if s_1(θ) = 0 then
        if v^0 large then
            α_0^t = α^t = α^a, update v^0 and v^1
        else
            α_0^t = α^t = α^s, update v^0 and v^1
        end
    elseif s_1(θ) = N then
        if v^1 large then
            α_0^t = α^t = α^a, update v^0 and v^1
        else
            α_0^t = α^t = α^s, update v^0 and v^1
        end
    else
        if v^1 close to v^0 then
            α_0^t = α^t = α^a, update v^0 and v^1
        else
            α_0^t = α^t = α^f, update v^0 and v^1
        end
    end
    t ← t + 1
until ∅
```

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

Stationary:

Period	0	1	2	3	4	5	6	7
Rating distribution	(0,10)	(1,9)	(3,7)	(7,3)	(5,5)	(7,3)	(4,6)	(2,8)
Recommended plan	A	A	A	F	A	F	A	A

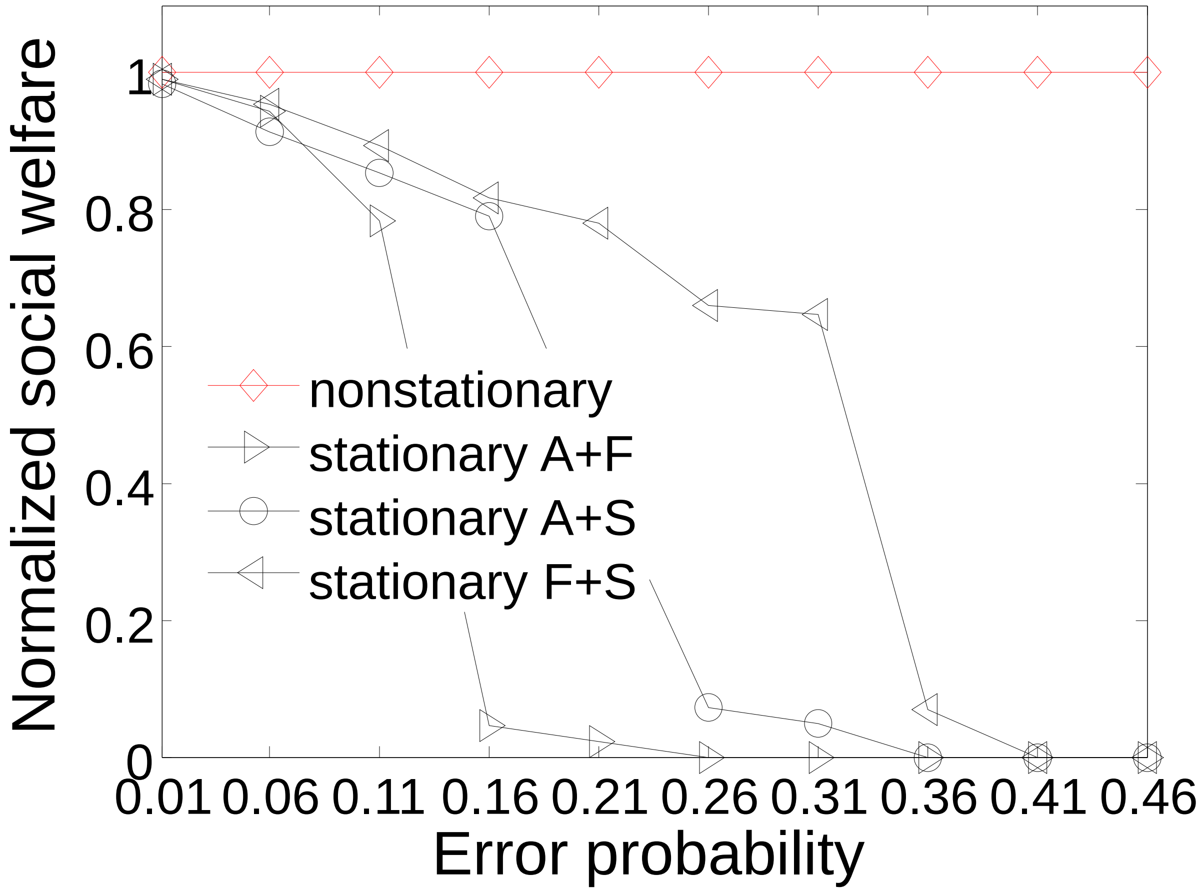
Optimal: (Non-stationary)

Period	0	1	2	3	4	5	6	7
Rating distribution	(0,10)	(1,9)	(3,7)	(7,3)	(5,5)	(7,3)	(2,8)	(1,9)
Recommended plan	A	A	A	F	A	A	A	A

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

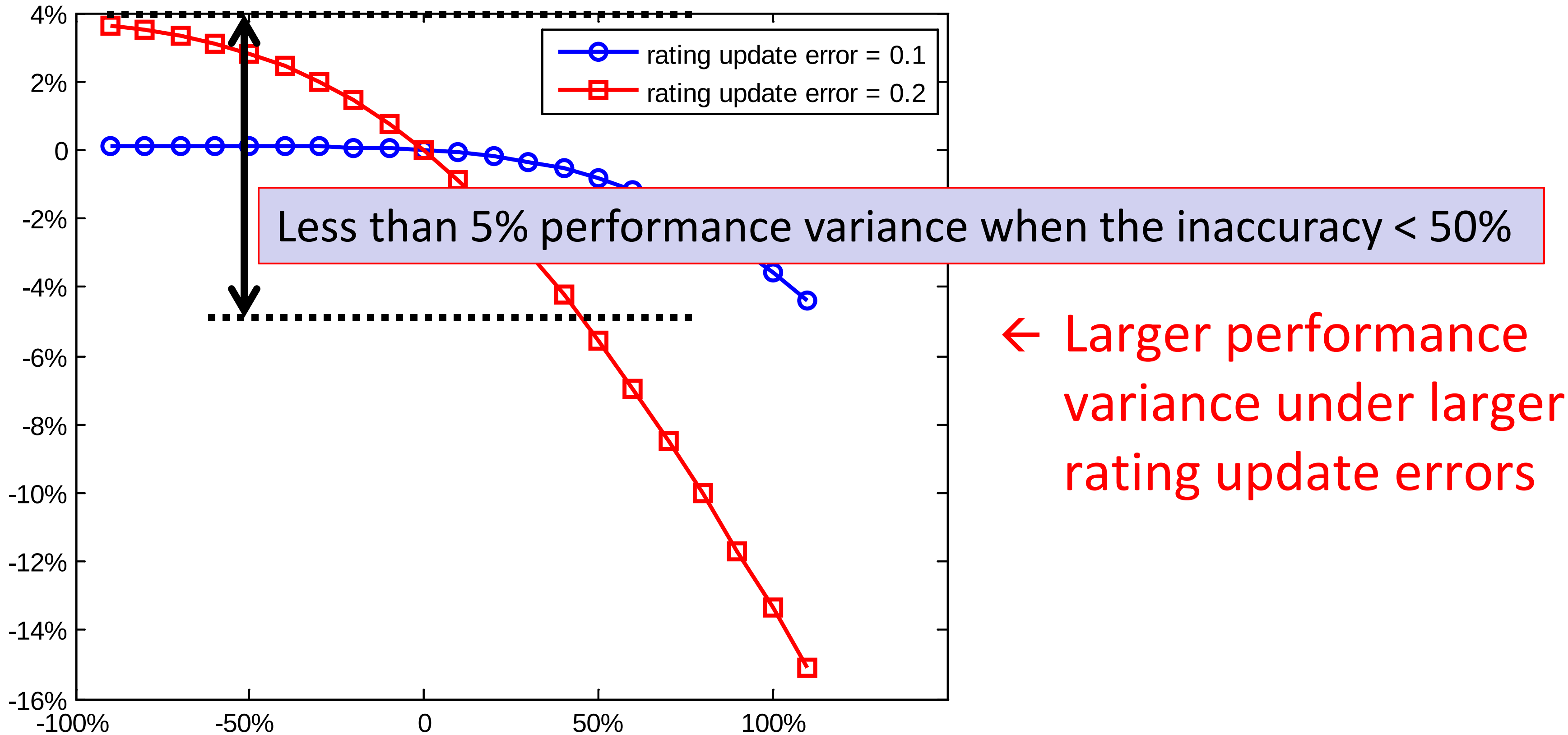
- Threshold-based stationary recommended strategies that use two plans:
  - Strategy 1: A+F, Strategy 2: A+S, Strategy 3: F+S
  - e.g., use A when the number of good users is large, use F otherwise
  - optimal threshold for each rating update error



Robustness of proposed mechanisms

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

performance gain/loss (in percentage):



← Larger performance variance under larger rating update errors

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

Related Works

	Rigorous analysis	Recommended strategy	Rating update errors	Discount factor	Performance loss due to imperfect reporting
Cohen'2003, Stoica'2006, etc.	No	Stationary	> 0	< 1	Yes
Kandori'1992, Takahashi'2010 Deb'2013, etc.	Yes	Stationary	= 0	→1	Yes
Ellison'1994	Yes	Stationary	→0	→1	Yes
Dellarocas'2005	Yes	Stationary	> 0	< 1	Yes
Design Stat. Zhang and vd Schaar'2012	**Yes**	**Stationary**	**→0**	**< 1**	**Yes**
Optimal Design Xiao and vd Schaar'2013	**Yes**	**Nonstationary**	**> 0**	**< 1**	**No**

Extensions and Beyond

- Multiple ratings

  Y. Zhang, J. Park, M. van der Schaar, “Rating Protocols for Online Communities, “ *Transaction on Economics and Computation*, 2013

- Collective ratings

  Y. Zhang, M. van der Schaar, "Incentive Provision and Job Allocation in Social Cloud Systems", *IEEE J. Sel. Areas in Commun,* 2013

- Robustness of norms

  Y. Zhang and M. van der Schaar, "Robust Reputation Protocol Design for Online Communities: A Stochastic Stability Analysis," in IEEE J. of Sel. Topics in Signal Process., vol. 7, no. 5, pp. 907-920, Oct. 2013

- Social norms on networks

  J. Xu, Y. Song, and M. van der Schaar, "Sharing in Networks of Strategic Agents," IEEE J. Sel. Topics Signal Process. - Special issue on "Signal Processing for Social Networks", Aug. 2014

  J. Xu and M. van der Schaar, "Efficient Working and Shirking in Networks," IEEE JSAC Bonus Issue for Emerging Technologies, April 2015.

Extensions and Beyond

- Tokens vs. Ratings

  M. van der Schaar, J. Xu, W. Zame, "Efficient online exchange via fiat money", *Economic Theory,* Feb. 2013

- Adverse selection

  M. van der Schaar and S. Zhang, "A Dynamic Model of Certification and Reputation," Economic Theory, vol. 58, no. 3, pp. 509-541, Oct. 2014.

- 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.