Implementation and mind control

Hitoshi Matsushima

University of Tokyo, Economics Department

July 8, 2009
Nash implementation

- Uninformed principal (planner) delegates decision to informed agents.
- Adverse Selection, Mechanism design, unique NE

Standard approach

- Agent’s incentive is based on material interest
  
  cf. Glazer and Rosenthal (92): Salience
- Mechanism design: Punish and Reward

  Maskin (77/99), Abreu + Matsushima (92)
A behavioral approach

- Agents’ incentive is based, not only on material interest, but also, on social psychology (obedience, conformity)
  
  Ash (55), Milgram (74), Zimbardo (77)

- psychological cost of lying
  
  \[ W_i(s_i) = 0 \quad \text{if strategy} \quad s_i \quad \text{implies honest} \]
  
  \[ W_i(s_i) > 0 \quad \text{otherwise} \]

- Tiny cost functions in implementation
  
  Matsushima (02, 08a, 08b), Dutta + Sen (09), Kartik + Tercieux (09)
Eichmann test

Prison experiments
Ash experiment
Present paper

- Psychological cost depend on ‘expectation’

  ex. Psychological game  Geanakoplos et al (89)
  Charness et al (06)

\[ W_i(s_i, s_{-i}) \]

- \( S_i \): Agent i’s strategy
- \( S_{-i} \): Agent i’s expectation on others’ strategies
• **Expectation-based obedience (EBO)**

  Psychological cost is greater if he expects others have kept honest.
  Psychological cost is smaller if he expects others have lied.

• **Principal designs mechanism that makes it easy to control agents’ mind.**

  ⇒ Agents expect others to keep honest for a short while.
  ⇒ They want to keep honest longer.
  ⇒ They expect others to keep honest longer.
  ⇒ ⋯ ⋯ ⇒ ‘Honest ever’ becomes unique NE
How to design mechanism?

\[ A, (M, g, r), M \equiv \times_{i \in \mathbb{N}} M_i, \quad g : M \rightarrow A, \quad r \in [0,1), \quad s_i : [0, \infty) \rightarrow M \]

- **Continuous time horizon** \([0, \infty)\)
- Agents make announcements at initial time 0, \(s_i(0) \in M_i\).
- Agents can change announcements any time, many times.
- Principal determines terminal time \(\tilde{t}\) **randomly** with hazard rate \(r\).
- Principal follows **final** announcements.
- Principal **prohibits** mutual monitoring and communication.
- **No** ‘punish and reward’ scheme is used.
Probability Density \( = r \cdot \exp(-rt) \)

Terminate at time \( \tilde{t} \)

Agent \( i \) makes Announcement \( s_i(t) \in M_i \)

Principal selects \( g(s(\tilde{t})) \in A \)
Principal prohibits mutual monitoring and communication

⇒ Strategy is path-independent, $s_i : [0, \infty) \rightarrow M_i$

cf. Montgomery Bus Boycott in 1955
Utility (expectation-based)

\[ U_i(s) = V_i(s) - W_i(s) \]

- \( s_i \in S_i \): His strategy
- \( s_{-i} \in S_{-i} \): His expectation on other’s strategies

**Material Payoff**: Expected value of his intrinsic utility \( v_i(\alpha) \)

\[ V_i(s) = \int_{-\infty}^{\infty} v_i(g(s(t)))d[1 - \exp(-rt)] \]
Utility satisfies EBO!

Agent $i = 3'$s expectation

He expects Agent 1 to keep honest

He expects Agent 2 to keep honest

He keeps honest

He keeps honest longer

Lie! More guilty

Lie! Less guilty

Small $\varepsilon > 0$
**Expectation-based obedience (EBO): Definition**

\[ m_i^* \in M_i, \quad m^* = (m_i^*)_{i=1}^n \]  
Truthful message

\[ s_i^* \in S_i \]  
Truthful strategy,  
\[ s_i^*(t) = m_i^* \quad \text{for all} \quad t \geq 0 \]

\[ t_i(s_i) \in [0, \infty) \]  
First time for agent \( i \) to tell a lie

\[ s_i(t_i(s_i)) \neq m_i^*, \quad s_i(\tilde{t}) = m_i^* \quad \text{for all} \quad \tilde{t} < t_i(s_i) \]

\[ s_{i,t} \in S_i \]  
Agent \( i \) keeps honest before \( t \), follows \( s_i \) afterwards

\[ s_{i,t}(\tilde{t}) = m_i^* \quad \text{for all} \quad \tilde{t} \in [0,t) \]

\[ s_{i,t}(\tilde{t}) = s_i(\tilde{t}) \quad \text{for all} \quad \tilde{t} \geq t \]
Utility Function satisfies EBO if

For every $i \in N$, $j \in N \setminus \{i\}$, and $s \in S \setminus \{s^*\}$,

$$[t_i(s_i) \leq t_j(s_j) \leq t_h(s_h) \text{ for all } h \in N \setminus \{i, j\}]$$

$$\downarrow$$

$$\lim_{\varepsilon \downarrow 0} \frac{W_i(s) - W_i(s/s_{i,j}(s_j) + \varepsilon)}{\varepsilon} > r \max_{(a, a') \in A^2} |v_i(a) - v_i(a')| \exp(-r t_j(s_j))$$

- Lie after someone else has lied saves psychological cost.
- Marginal decrease in psychological cost is greater than marginal decrease in intrinsic (material) utility.
Incentive compatibility in terms of intrinsic utility (IC)

\[ \nu_i(g(m^*)) \geq \nu_i(g(m^*/m_i)) \]

for all \( i \in N \) and all \( m_i \in M_i \)

Main Theorem

With \( n \geq 3 \), EBO, and IC, truthful strategy profile \( s^* \) is unique Nash equilibrium
Agent 3 is honest

Agent 2

Lie! Feel more guilty

Tail-chasing competition (a la AM)

Time 0

t₃
t₂

t₂ + ε

Agent 3

Probability of Termination is small

εr \exp(-rt₂) \approx 0

∴ Expected decrease in material payoff is at most

\max_{(a,a') \in A^2} \left| v_i(a) - v_i(a') \right| \varepsilon r \exp(-rt₂) \approx 0

∴ Less than psychological cost saving (EBO)

Expects others to be honest

∴ Payoff increases (IC)
Tail-chasing competition: difference from AM

AM mechanism controls material interest by fining first deviant explicitly. Mechanism in present paper control mind to dislike being first deviant.

Psychological cost can be negligible compared to material payoff

$$\max_{(s,s') \in S^2} |W_i(s) - W_i(s')| \approx 0 \text{ for all } i \in N$$