Notes and Comments
Overcoming Incentive Constraints by Linking decisions

Matthew Jackson and Hugo Sonnenschein

Econometrica 2007

April 28, 2009

Introduction: Mechanism Design

"The mechanism design problem is to implement an optimal system-wide solution to a decentralized optimization problem with self-interested agents with private information about their preferences for different outcomes".

Two ways to think of it:

- How to get rugged individualists to work together.
- How to design the payoff matrix so that agents pick the strategies that you want them to pick.

Introduction : Banana Splitting

Suppose we have a banana b and two agents $1, 2$. We can give the banana to the first agent, $b1$, or to the second $b2$.

Agent 1	$h_1(b1) = 3,\ h_1(b2) = 0$	$f_1(b1) = 1,\ f_1(b2) = 0$
Agent 2	$h_2(b2) = 3,\ h_2(b1) = 0$	$f_2(b2) = 1,\ f_2(b1) = 0$

Table: Banana Splitting

Agent 1	$P_1(h_1) = P_1(f_1) = 0.5$
Agent 2	$P_2(h_2) = 0.3,\ P_2(f_2) = 0.7$

Table: Banana Splitting

Suppose we want to give the banana to the agent that is hungry, if both or neither are, we flip a coin.
This social choice function is not implementable: agents have no incentive to truthfully reveal their types.

Problem and Idea

- Incentive constraints impose limitations on the attainments of socially efficient outcomes.

- Idea: link decisions and 'budget' an agents moves/messages depending on the probability distribution of his types. goto

Decision problems

An n-agent decision problem is a triple $\mathcal{D} = (D, U, P)$ where

- D, a finite set of possible alternative decisions,
- $U = (U_1 \times \ldots \times U_n)$, a finite set of possible utility functions $(u_1, \ldots, u_n)$ where $u_i : D \to \mathbb{R}$,
- $P = (P_1, \ldots, P_n)$ a profile of probability distributions, where P_i is a distribution over U_i.

We assume u_i's are drawn independently.

Social Choice Function

A *social choice function* on a decision problem $\mathcal{D} = (D, U, P)$ is a function

$$f : U \to \Delta(D),$$

where $\Delta(D)$ is the set of probability distributions on D. f is the *target* outcome function.

Social Choice Function

A *social choice function* on a decision problem $\mathcal{D} = (D, U, P)$ is a function

$$f : U \to \Delta(D),$$

where $\Delta(D)$ is the set of probability distributions on D. f is the *target* outcome function.

f is *ex ante Pareto efficient* if there does not exist a f' on $\mathcal{D}$ such that

$$\Sigma_u[P(u)\Sigma_d(f'_d(u)u_i(d))] \geq \Sigma_u[P(u)\Sigma_d(f_d(u)u_i(d))],$$

for all i with $>$ for some i.

Given a decision problem $\mathcal{D} = (D, U, P)$ and K linkings.

- A *linking mechanism* (M, g) is a message space $M = (M_1 \times \ldots \times M_n)$ and an outcome function $g : M \to \Delta(D^k)$.

Given a decision problem $\mathcal{D} = (D, U, P)$ and K linkings.

- A *linking mechanism* (M, g) is a message space $M = (M_1 \times \ldots \times M_n)$ and an outcome function $g : M \rightarrow \Delta(D^k)$.
- A *strategy* for agent i in (M, g) on K copies of the decision problem is a mapping

$$\sigma_i^K : U_i^K \rightarrow \Delta(M_i)$$

Linking Mechanisms

Given a decision problem $\mathcal{D} = (D, U, P)$ and K linkings.

- A *linking mechanism* (M, g) is a message space $M = (M_1 \times \ldots \times M_n)$ and an outcome function $g : M \to \Delta(D^k)$.
- A *strategy* for agent i in (M, g) on K copies of the decision problem is a mapping

$$\sigma_i^K : U_i^K \to \Delta(M_i)$$

- We consider Bayesian Equilibria, σ's, of such mechanisms.

Linking Mechanisms, continued

- A Bayesian Nash equilibrium is a strategy profile that maximizes the expected payoff for each player given their beliefs about the other players' types and given the strategies played by the other players.

Linking Mechanisms, continued

- A Bayesian Nash equilibrium is a strategy profile that maximizes the expected payoff for each player given their beliefs about the other players' types and given the strategies played by the other players.

- Given a social choice function f defined on $\mathcal{D}$, we say that a *sequence of linking mechanisms* $\{(M^1, g^1), (M^2, g^2), \ldots, (M^k, g^k), \ldots\}$ and a corresponding sequence of Bayesian equilibria, $\{\sigma^K\}$, *approximate* f if

$$lim_k[maxProb_{k \leq K}\{g_k^K(\sigma^K(u)) \neq f(u^k)\}] = 0.$$

Their general linking mechanism

- M_i consists of announcements of utility functions for each decision problem.
- The intuitive idea: each agent announces utility functions for the K problems such that the agent's announcements across the K problems match the expected frequency distribution. That is, the number of times i can announce u_i is $K \times P_i(u_i)$. (Sometimes we need to approximate)

Their general linking mechanism

- M_i consists of announcements of utility functions for each decision problem.
- The intuitive idea: each agent announces utility functions for the K problems such that the agent's announcements across the K problems match the expected frequency distribution. That is, the number of times i can announce u_i is $K \times P_i(u_i)$. (Sometimes we need to approximate)
- Formally: agent i's strategy set is

$$M_i^k = \{\hat{u}_i \in U_i^k \mid \#\{k \mid \hat{u}_i^k = v_i\} = P_i^k(v_i)K \text{ for each } v_i \in U_i\}.$$

- The decision of g^k for the problem k is simply $g^k(\hat{u}) = f(\hat{u}^k)$.

Their general linking mechanism

- Sometimes, an agent is forced to lie.
- A strategy is *approximately truthful* if the agent's announcements always involve *as few lies as possible*.
- Formally, $\sigma_i^K : U_i^K \to M_i^K$ is approximately truthful if
$$\#\{K \mid [\sigma_i^K(u_i^1, \dots, u_i^K)]^k \neq u_i^k\} \leq \#\{k \mid m_i^k \neq u_i^k\}$$
for all $m_i \in M_i^K$.

Back to the bananas

Remember:

Agent 1	$h_1(b1) = 3,\ h_1(b2) = 0$	$f_1(b1) = 1,\ f_1(b2) = 0$
Agent 2	$h_2(b2) = 3,\ h_2(b1) = 0$	$f_2(b2) = 1,\ f_2(b1) = 0$

Table: Banana Splitting

Agent 1	$P_1(h_1) = P_1(f_1) = 0.5$
Agent 2	$P_2(h_2) = 0.3,\ P_2(f_2) = 0.7$

Table: Banana Splitting

- The social choice function $f =$ 'Give the Banana to the hungry agent, if both or neither are, flip a coin.'
- If $K = 10$, then, following the mechanism, player 1 gets 5 tokens to say 'I am hungry', and 5 for 'I am not hungry'. Similarly, player 2 gets 3 tokens for 'I am hungry', and 7 for 'I am not hungry'. back

The main result

Theorem

Given a 2-player decision problem $\mathcal{D}$ and an ex ante pareto efficient social choice function f. The sequence of the linking mechanisms, $\{(M^1, g^1), (M^2, g^2), \ldots, (M^k, g^k), \ldots\}$ just defined satisfies the following:

(i) There exists a corresponding sequence of Bayesian equilibria that are approximately truthful.

(ii) The sequence of linking mechanisms together with these equilibria approximate f.

(iii) Any sequence of approximately truthful strategies for an agent i secures a sequence of utility levels that converge to the ex ante target level $\bar{u}_i$, the expected outcome of f for player i.

The main result

Theorem, continued

(iv) All sequences of Bayesian equilibria of the linking mechanisms result in expected utilities that converge to the ex ante efficient profile of target utilities of $\bar{u}$ per problem.

(v) For any sequence of Bayesian equilibria and any sequence of deviating coalitions, the maximal gain by any agent in the deviating coalitions vanishes along the sequence.

The theorem also holds in a general setting of n players but requires some modification of the linking mechanism.

To conclude

Some crucial assumptions:

- Reports after K linkings.
- Independence:
  - Of utilities among players,
  - Of utilities across instances of the game.

If any of these assumptions are dropped, then, nobody knows...