Hotelling Lectures: Evidence in Games and Mechanisms
Part 2: Mechanism Design

Barton L. Lipman
Boston University

November 2023

Revelation Principle

Literature contains various versions of the Revelation Principle for evidence.

Basic result: With normal evidence, can restrict attention to *simple, truth–telling, maximal evidence* mechanisms.

More formal: Any outcome that can be implemented with normal evidence can be implemented in a simple mechanism that induces truth–telling and presentation of maximal evidence.

Simple: Agents simultaneously report types and present evidence, followed by mechanism giving outcome.

Normality clearly necessary for this as Lipman–Seppi example shows.

Truth telling and maximal evidence: Mechanism induces t_i to report t_i and to provide maximal evidence.

Normality clearly necessary for this too: maximal evidence not feasible otherwise.

Intuition clear: Simplicity for usual reasons plus normality; truth telling for usual reasons; maximal evidence because why not?

See Green and Laffont (*RES*, 1986), Bull and Watson (*GEB*, 2007), Deneckere and Severinov (*GEB*, 2008), and Forges and Koessler (*JMathEcon*, 2005).

More general model treated in Ben-Porath, Dekel, and Lipman (2021).

Implication: The mechanism design problem to solve is

$$\max_{\gamma} \mathrm{E}_t \left[\sum_{a \in A} \gamma(a \mid t, M(t)) v(a, t) \right]$$

subject to

$$\sum_{a \in A} \gamma(a \mid t, M(t)) u(a, t) \geq \sum_{a \in A} \gamma(a \mid t', E) u(a, t), \ \forall t, t', \ \forall E \in \mathcal{E}(t).$$

Explanation: $\gamma : T \times \mathcal{E} \to \Delta(A)$.

Treating A as finite and ignoring IR for simplicity.

Often, principal can punish "obvious deviations" — e.g., claim of type t' but evidence $\neq M(t')$.

In such models, these constraints don't bind so IC becomes

$$\sum_{a \in A} \gamma(a \mid t, M(t)) u(a, t) \geq \sum_{a \in A} \gamma(a \mid t', M(t')) u(a, t),$$
$$\forall t, \ \forall t' \text{ with } M(t') \in \mathcal{E}(t).$$

Here evidence determines which types t can imitate. This is how Green and Laffont (*RES*, 1986) defined the evidence model.

See Sher and Vohra (*TE*, 2015) for an application with this structure.

Value of Commitment

In usual mechanism design, rare to see cases where commitment isn't helpful.

Typically, principal commits to some ex post inefficient response to better induce truth–telling.

With evidence, many settings with no value to commitment.

Evidence is not magic, but it makes us think about different models.

Example. $T = \{1, \ldots, 1000\}$.

Prior: $\mu(1000) = 2/1001$, $\mu(t) = 1/1001$ for $t \neq 1000$.

Evidence: $\mathcal{E}(1000) = \{T\}$, $\mathcal{E}(t) = \{\{t\}, T\}$ for $t \neq 1000$.

Actions: $A = \mathbf{R}_+$.

Utility for agent/sender: a.

Case 1:

$$v(a,t) = \begin{cases} 1, & \text{if } a = t \\ 0, & \text{otherwise.} \end{cases}$$

Claim: In every equilibrium without commitment, after observing no evidence, principal chooses $a = 1000$.

So no type presents evidence and the principal's expected payoff is $\mu(1000) = 2/1001$.

If principal can commit, she can do better.

Suppose principal commits to choosing $a = 0$ if no evidence is presented.

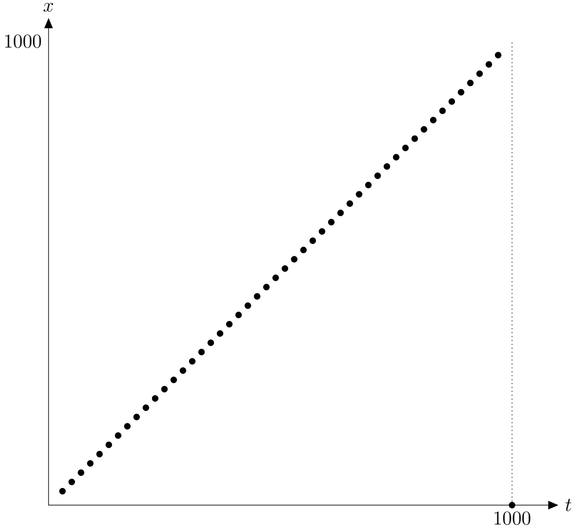
Every type with evidence presents it, so principal sets $a = t$ for all $t \neq 1000$. Expected payoff is $999/1001$.

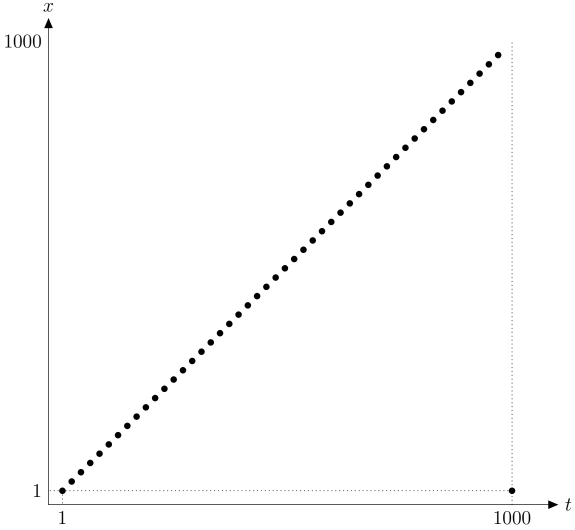
Case 2:

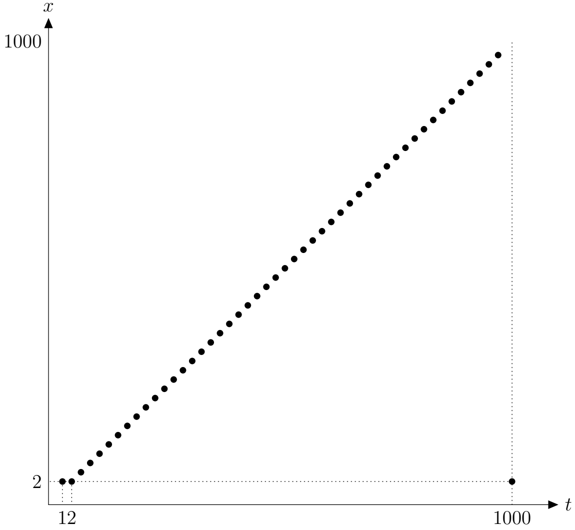
$$v(a, t) = -(t - a)^2.$$

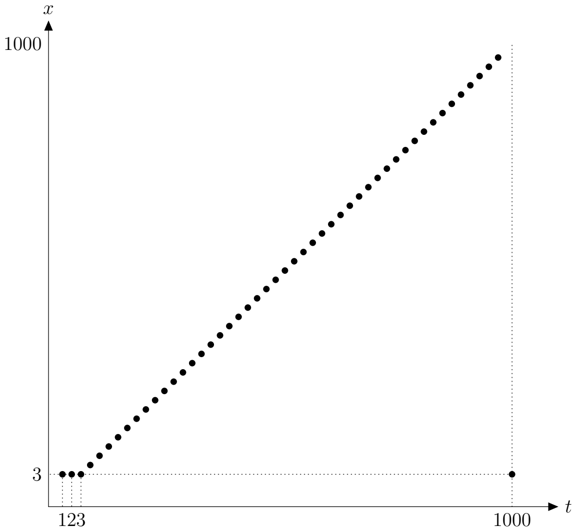
Now commitment has no value.

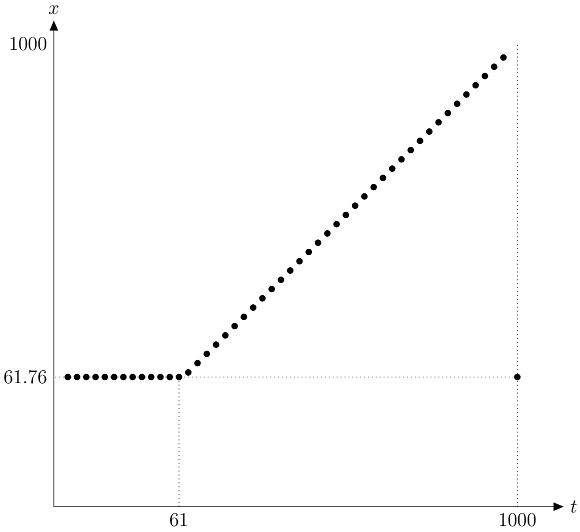
Why doesn't the same commitment as above help?


x
1000
t
1000


x
1000
1
1
1000
t


x
1000
2
1
2
1000
t


x
1000
3
123
1000
t


x
1000
61.76
61
1000
t

Continues to action equal to pooled group's mean and below the lowest type outside the pool.

This is the Dye equilibrium.

A General But Weird Result

From Ben-Porath, Dekel, and Lipman (2021).

T = set of types of agent.

A = set of actions for principal.

$u(a, t)$ = utility function of agent, $v(a, t)$ = utility function of principal.

Fix any *communication protocol*. I.e., stages of cheap talk, evidence presentation, ultimately choice of a.

B = pure strategies for agent, $\Delta(B)$ mixed strategies, β typical mixed strategy.

G = pure strategies for principal, $\Delta(G)$ mixed strategies, γ typical mixed strategy.

$U(\beta, \gamma)$ = expected utility for agent, $V(\beta, \gamma)$ = expected utility for principal.

Let

$$BR_u(\gamma) = \text{Best replies for agent to } \gamma.$$

Payoff to principal under commitment:

$$V^* \equiv \max_{\gamma \in \Delta(G)} \left[\max_{\beta \in BR_u(\gamma)} V(\beta, \gamma) \right].$$

Assumptions:

1. Set of pure strategies for agent and for principal finite.
2. There exists γ^* such that for *every* $\beta \in BR_u(\gamma^*)$,

$$V(\beta, \gamma^*) = V^*.$$

Changes in agent's best response don't affect principal's payoff.

Weird because it's not directly a condition on primitives of the model.

Theorem. [No value to commitment.] Under assumptions 1 and 2, there is a Nash equilibrium in game without commitment giving the principal payoff V^*.

Can extend to perfect Bayesian given additional structure on protocol.

Rough intuition: Need for commitment in usual models is that γ^* is not a best reply to β^*.

Instead, it's chosen to "push" the agent in a certain way.

The "weird condition" limits the effects of changes in the agent's strategy and hence in principal's incentive to move away from best reply.

▸ Jump

Partial proof. Fix γ^* and $\beta^* \in BR_u(\gamma^*)$ with $V(\beta^*, \gamma^*) = V^*$.

By assumption $\beta^* \in BR_u(\gamma^*)$, so if γ^* is best reply to β^*, we're done.

By our assumption, we can change β^* to any other best reply to γ^* without changing principal's payoff.

So if we can find any $\hat{\beta} \in BR_u(\gamma^*)$ with the property that γ^* is best reply to $\hat{\beta}$, we're done.

Consider *restricted game* where agent can only pick $\beta \in BR_u(\gamma^*)$ and principal can pick any γ.

Let $(\hat{\beta}, \hat{\gamma})$ be NE of this restricted game.

As noted, if γ^* is a best reply to $\hat{\beta}$, we're done, so suppose not.

Then $V(\hat{\beta}, \hat{\gamma}) > V(\hat{\beta}, \gamma^*)$.

Consider the mixed strategy $\gamma_\varepsilon = \varepsilon\hat{\gamma} + (1-\varepsilon)\gamma^*$.

Claim: For some very small $\varepsilon > 0$, $\hat{\beta}$ is a best reply to γ_ε.

Why? For ε small enough, it's like a lexicographic problem where first you maximize against γ^* and then break indifferences by maximizing against $\hat{\gamma}$. (Uses finiteness.)

By construction, within set of best replies to γ^*, $\hat{\beta}$ is best reply to $\hat{\gamma}$.

$\gamma_\varepsilon = \varepsilon\hat{\gamma} + (1-\varepsilon)\gamma^*$.

So if principal commits to γ_ε and picks $\hat{\beta}$ as the agent's best reply, her payoff is

$$\begin{aligned} V(\hat{\beta}, \gamma_\varepsilon) &= \varepsilon V(\hat{\beta}, \hat{\gamma}) + (1-\varepsilon) V(\hat{\beta}, \gamma^*) \\ &> V(\hat{\beta}, \gamma^*) = V^* \end{aligned}$$

This contradicts V^* being principal's payoff under best commitment.

Special Cases

1. $A = \{a_1, a_2\}$.

Reason: Best strategy for agent must give highest possible probability of a_1 for types who prefer it and lowest for types who prefer a_2.

So any other best reply must have *same* probability over actions for every type.

But then the principal's payoff same.

Includes all accept/reject problems; generalizes Glazer–Rubinstein (*Econometrica*, 2004, *TE*, 2006).

2. Optimal mechanism deterministic. Agent's utility either strictly increasing or strictly decreasing in a, can vary by t

Best strategy for agent (given no randomization) has each type t getting largest/smallest a she can, say a_t.

So any other best reply has to also give type t action a_t for all t.

So principal indifferent.

Includes square–error loss settings. Generalizes Sher (*GEB*, 2011) and Hart, Kremer, and Perry (*AER*, 2017).

Multi–Agent Mechanisms

Ben-Porath, Dekel, and Lipman (*Econometrica*, 2019) has the only multi–agent result on commitment that I know.

Has other interesting features as well.

Problem 1: Simple allocation problem.

Principal has one indivisible good to allocate to one of N agents.

Each agent cares only about getting good and prefers getting it to not getting it.

Value to principal of giving good to i is $v_i(t_i)$ where t_i is known only to agent i.

No monetary transfers possible.

Examples:

1. Dean has one job slot for a department in College; departments know who they'd hire and how this would contribute to dean's payoff.
2. Divisions of an organization all want to head up a prestigious project for the firm; they know more about their capabilities than senior management.
3. Regional government will build hospital in one city in the region; cities all want the hospital and have private information about health issues in their cities.

Variations:

1. Multiple identical units.
2. Task assignment: No agent wants the good.
3. Task assignment: Some agents want the good and this is private information which may be correlated with ability.

Problem 2: Public good problem.

Principal must decide whether to provide public good or not.

If so, each of N agents pays $1/N$ of cost.

Net value to i of public good is $v_i(t_i)$ where t_i is known only to agent i.

Principal's objective function is sum of agents' utilities.

Results

For a broad class of mechanism design problems, including these:

- Randomization has no value for the principal. Not exciting but plays a role in the commitment discussion.
- Requiring robust IC (instead of Bayesian IC) has no cost for the principal.
- Commitment has no value for the principal. There is an equilibrium in game without commitment giving principal same payoff as optimal mechanism.
- This equilibrium has a simple form and is also robust.

Model.

N agents. T_i = set of types of agent i. Assumed finite.

A = set of actions for principal. Assumed finite.

Agents' utility: *Simple type dependence.*

$$u_i(a, t_i) = \begin{cases} u_i(a); & \text{if } t_i \in T_i^+; \\ -u_i(a); & \text{otherwise.} \end{cases}$$

T_i^+ = positive types.
$T_i^- = T_i \setminus T_i^+$ = negative types.

Utility can depend on type, but indifference curves same for every type.

Principal's utility:

$$v(a,t) = u_0(a) + \sum_{i=1}^{N} u_i(a)v_i(t_i).$$

If t_i positive type and $v_i(t_i) > 0$, principal is putting positive weight on t_i's utility. Similar if t_i is negative type and $v_i(t_i) < 0$.

Can think of $v_i(t_i)$ as measuring how much principal "cares" about t_i's utility or as how much what principal likes lines up with t_i's utility.

All examples given earlier are special cases.

Model evidence through proving events. t_i can prove something about her type, not types of others.

So $\mathcal{E}_i(t_i)$ is collection of subsets of T_i that t_i can prove.

Assume normality.

Mechanism is $P : T \times \mathcal{E} \to \Delta(A)$. Objective function is

$$\mathrm{E}_t \left[\sum_{a \in A} P(a \mid t, M(t)) \sum_{i=0}^{N} u_i(a) v_i(t_i) \right].$$

Let

$$\mathcal{U}_i(s_i, E_i, t_{-i}, E_{-i} \mid P, t_i) = \sum_{a \in A} P(a \mid s_i, t_{-i}, E) u_i(a, t_i).$$

Type t_i's expected payoff in P given she reports (s_i, E_i) and others say (t_{-i}, E_{-i}).

Robust IC

Usual incentive compatibility constraint:

$$\mathrm{E}_{t_{-i}} \mathcal{U}_i(t_i, M_i(t_i), t_{-i}, M_{-i}(t_{-i}) \mid P, t_i) \\ \geq \mathrm{E}_{t_{-i}} \mathcal{U}_i(s_i, E_i, t_{-i}, M_{-i}(t_{-i}) \mid P, t_i),$$

for all $t_i, s_i \in T_i$, all $E_i \in \mathcal{E}_i(t_i)$, for all i.

Honesty is optimal in expectation given honesty by others.

Robust incentive compatibility:

$$\mathcal{U}_i(t_i, M_i(t_i), t_{-i}, E_{-i} \mid P, t_i) \geq \mathcal{U}_i(s_i, E_i, t_{-i}, E_{-i} \mid P, t_i),$$

for all $t_i, s_i \in T_i$, all $E_i \in \mathcal{E}_i(t_i)$, all $t_{-i} \in T_{-i}$, all $E_{-i} \in \mathcal{E}_{-i}$, for all i.

Honesty is optimal no matter what other agents' type reports and evidence are. Same idea as dominant strategy incentive compatibility, but stronger in this setting.

Robust Equilibrium

Analog to robust IC: Each agent's strategy is optimal given any actions by the other agents.

Important point: Robustness is only wrt the other agents, not wrt the principal's strategy.

Simple Equilibrium

For remainder of explanation, focus on simple allocation problem.

"Simplicity": Can construct the equilibrium without commitment via equilibria of a family of *auxiliary games*.

For each i, define auxiliary game between i and the principal where

- i sends cheap talk report and evidence
- principal responds with $\hat{v} \in \mathbf{R}$
- principal's payoff is $-(v_i(t_i) - \hat{v})^2$
- i's payoff is $\hat{v}$

To get equilibrium for game without commitment:

For each agent i, i's strategy = her strategy in the auxiliary game.

Principal's strategy is best reply to these strategies for agents.

So result is that this will be an equilibrium and give principal same payoff as in optimal mechanism.

Intuition

In auxiliary game, i just wants to make the principal believe v_i is large.

In game without commitment, principal allocates good to agent for whom his expectation of v_i is largest.

i still just wants to make the principal believe v_i is large. So she follows the same strategy.

PBE robust: i's strategy optimal regardless of what other agents are doing.

This gives robust incentive compatibility.

This doesn't explain why principal can't do better with commitment.

Very rough intuition: The robustness property makes it like N separate one–agent problems.

Then a version of one–agent proof works.

Illustrate with Dye evidence.

Equilibrium of the auxiliary game: There exists unique v_i^* such that

$$v_i^* = \mathrm{E}\left[v_i(t_i) \mid t_i \text{ has no evidence or } v_i(t_i) \leq v_i^*\right].$$

- If agent i has evidence and $v_i(t_i) > v_i^*$, i discloses.
- Evidence types with $v_i(t_i) \leq v_i^*$ pool with no–evidence types.
- Principal's expectation of v_i given nondisclosure is v_i^*.

Our result says these strategies are used in the equilibrium of the game without commitment and give optimal mechanism.

Assume $v_1^* > v_2^*, \ldots, v_N^*$.

1 either proves she's better than v_1^* or proves nothing and principal's expectation $= v_1^*$.

If no agent $i \neq 1$ proves $v_i(t_i) > v_1^*$, principal gives good to 1.

If some agent $i \neq 1$ proves $v_i(t_i) > v_1^*$, good goes to agent who proves highest value.

Mechanism: 1 is *favored agent* and v_1^* is *threshold*. If no non–favored agent proves value above threshold, favored agent gets good. Otherwise, good goes to whoever proves highest value.

Can get optimal mechanisms in other examples similarly.

End of Part 2.