Hotelling Lectures: Evidence in Games and Mechanisms
Part 1: Games

Barton L. Lipman
Boston University

Large parts based on joint work with
Elchanan Ben-Porath (Hebrew University)
Eddie Dekel (Northwestern/Tel Aviv University)

November 2023

Introduction

Standard models of strategic communication based on Spence/Crawford–Sobel.

Key idea is to use variation in preferences across types to induce different choices by different types.

Traditional work in mechanism design similarly introduces monetary transfers and differences across types in willingness to pay.

Focus of these lectures: The role of evidence — hard information that establishes key facts regardless of incentives.

Evidence plays an important role in many contexts:

- Lawyers trying to persuade a judge to rule in their favor
- A buyer observing evidence about the product of a seller
- An investor learning about an entrepreneur's project
- An employer considering a potential employee
- Voters studying politicians' behavior
- A manager discovering which units in an organization to reward

Part 1: Games

1. Classical Single–Agent Model: Unraveling.
2. Single–Agent Models without Unraveling: Theory and applications.
3. Multiple Agents.

Part 2: Mechanisms

1. Revelation Principle.
2. Single–Agent Results on Value to Commitment.
3. Multi–Agent Value to Commitment and Robustness.

Part 3: Other Directions

1. Costly Verification.
2. Acquisition of Evidence: Games.
3. Acquisition of Evidence: Mechanisms.

Modeling Evidence

Types $t \in T$ differ in the evidence they can present and in other standard aspects.

Two (Equivalent) Ways to Model Evidence.

1. $\mathcal{M}(t)$ is set of *messages* t can send.

Sending m *proves* t is a type with $m \in \mathcal{M}(t)$.

Example. Type n cannot play the piano, type p can. Then

$$\mathcal{M}(n) = \{r\}$$

$$\mathcal{M}(p) = \{c, r\}.$$

Playing c proves that the agent is type p.

Impossible for agent to prove she can't play the piano.

2. $\mathcal{E}(t)$ is set of subsets of T that type t can prove.

In effect, replace m as message with proving the event $\{t \mid m \in \mathcal{M}(t)\}$.

Continuing example:

$$\mathcal{E}(n) = \{T\}$$

$$\mathcal{E}(p) = \{\{p\}, T\}$$

Require evidence to be **true**: $E \in \mathcal{E}(t)$ implies $t \in E$

and **consistent**: $s \in E$ for some E feasible for some type implies $E \in \mathcal{E}(s)$.

With either approach, we assume agent can only present *one* piece of evidence (one message/one event).

Without loss of generality: If agent could present K, rewrite evidence sets.

A common assumption: *Normality.*

Informally: No costs of/time constraints on evidence presentation — as if could present unlimited number of messages.

Formally: For all t,

$$\bigcap_{E \in \mathcal{E}(t)} E \in \mathcal{E}(t).$$

This version, due to Lipman–Seppi (*JET*, 1995), is called *full reports condition*. Equivalent definition, name due to Bull–Watson (*GEB*, 2007).

Will use $M(t)$ to denote LHS — *maximal evidence*.

Continuing example. Satisfies normality.

$$\mathcal{E}(n) = \{T\}$$

$$\mathcal{E}(p) = \{\{p\}, T\}$$

$$M(n) = T \in \mathcal{E}(n), \quad M(p) = \{p\} \in \mathcal{E}(p)$$

Suppose there are two types of music. Type n cannot play the piano, type c can play classical music, type b blues, type a all. If there's only time to play one piece of music:

$$\mathcal{E}(n) = \{T\}, \quad \mathcal{E}(c) = \{\{c, a\}, T\}$$

$$\mathcal{E}(b) = \{\{b, a\}, T\}, \quad \mathcal{E}(a) = \{\{c, a\}, \{b, a\}, T\}.$$

$M(a) = \{a\} \notin \mathcal{E}(a)$, so normality is violated.

Games: Classical Single–Agent Model

Basic assumptions:

- Sender/agent learns her type $t \in T$.
- Sender sends some evidence message to receiver/principal.
- Receiver chooses some action $a \in A$.
- Payoff $u(a)$ for sender, $v(a, t)$ for receiver.

Unless stated otherwise, “equilibrium” means PBE.

Sender’s preferences independent of t, so can’t use differences in preferences across types to get communication/induce truth–telling.

Games: Classical Single–Agent Model

Seminal model of evidence: Grossman (*Jour of Law and Econ*, 1981); Milgrom (*RAND Journal*, 1981).

Payoff structure: Square–error loss.

$T \subseteq \mathbf{R}_+$
$A = \mathbf{R}_+$
$u(a) = a$
$v(a, t) = -(a - t)^2$

Receiver is trying to estimate t (action = conditional expectation) and the sender wants receiver's estimate to be high.

This is not how Grossman and Milgrom wrote model but captures same features.

Examples:

- Receiver is employer wanting to pay sender her productivity t and a is the wage.
- Receiver is "the market" wanting to price stock at its true value t, sender is firm manager, a is the stock price.
- Receiver is investor who wants to invest "correctly" where t is optimal investment; sender is entrepreneur wanting the investment, a is the funds invested.

Key assumption: Sender has access to evidence that can prove any true fact — *complete provability*.

$\mathcal{E}(t) = \{E \subseteq T \mid t \in E\}$.

Comment. Interpretation: Verifiability versus proof.

Equilibrium: Highest type cannot be pooled with any lower types.

Reason: She could prove her type and be strictly better off.

But second–highest type cannot pool with any lower types, etc., leading to *unraveling*.

No type can pool with lower types — all private information is eliminated.

This says there is no equilibrium without revelation.

Construct an equilibrium with *skeptical beliefs*: whatever is proven, receiver believes worst thing for sender consistent with this.

Easy to see that sender's best reply is to prove everything and this makes receiver's belief correct on path.

Implication: The unique equilibrium outcome is the same as under full information.

Multiple equilibria since we don't know exactly what types prove, but unique outcome.

Generalization: Can generalize in various ways. However, *existence* of equilibrium with full–information outcome is more general than *uniqueness*.

Fix *any* utility functions $u(a)$ and $v(a, t)$.

Let $a^*(t)$ be the worst action for the sender among those maximizing $v(a, t)$.

Then with complete provability, there is an equilibrium in which the receiver chooses action $a^*(t)$ when the sender's type is t.

Proof: Skeptical beliefs. Receiver infers worst t for sender given any evidence. Sender will prove true t to make this least bad.

Uniqueness result is easy to generalize to models where sender's payoff *strictly* increases if we improve receiver's belief about her.

Key to unraveling is that sender with "best" type *strictly* prefers revealing this to pooling with any lower type, so sender with "second–best" strictly prefers revealing, etc.

Don't need *complete* provability — $\mathcal{E}(t) = \{\{t\}, T\}$ enough. Sometimes referred to as *disclosure game*.

Models without Unraveling: Different Preferences

Alternative payoff structure: Accept/reject.

$T \subseteq \mathbf{R}_+$
$A = \{0(\text{reject}), 1(\text{accept})\}$
$u(a) = a$
$v(a, t) = a(t - \hat{t})$

So receiver wants to accept types with high enough t and reject others; sender wants to be accepted.

Examples:

- Receiver is employer who can't affect wage $\hat{t}$; wants to hire if productivity t is above wage; sender is applicant.
- Receiver is investor who can't affect the amount needed for investment and wants to invest only if the return is above the investment required; sender is entrepreneur.
- Receiver is buyer who considers purchasing a good at a fixed price $\hat{t}$ and wants to buy only if value of good exceeds this; sender is seller.

The two payoff structures seem broadly similar.

In both, sender wants receiver to think t is large and receiver wants to figure out true t.

Assume $\mathrm{E}(t) > \hat{t}$.

Assume complete provability.

From reasoning above, there is an equilibrium in which sender always proves her type and receiver accepts only types above $\hat{t}$.

Use skeptical beliefs: if sender doesn't prove $t \geq \hat{t}$, believe it's below.

Another equilibrium: Sender proves nothing and receiver accepts.

Recall that $\mathrm{E}(t) > \hat{t}$, so receiver's strategy best reply.

Obvious that sender's strategy is optimal.

Why doesn't unraveling occur?

Highest type could prove her type and improve receiver's belief about her, but it doesn't increase her payoff.

If $\mathrm{E}(t) < \hat{t}$, sender's best equilibrium still corresponds to Bayesian persuasion outcome of Kamenica–Gentzkow (*AER*, 2011).

See Titova (2022) and Zhang (2022) for more general characterizations.

See Callander, Lambert, and Matouschek (*JPE*, 2021) and Ali, Lewis, and Vasserman (*RES*, 2023) for economic applications.

Models without Unraveling: Incomplete Provability

Complete provability is a natural case to study but hardly the most realistic.

Even in square–error loss setting, can't get unraveling without a lot of provability.

I'll primarily focus on cases without separation.

Models without Unraveling: Dye Evidence

Dye (*Jour of Accounting Research*, 1985) (Jung and Kwon, 1988): The sender has perfect evidence with probability $q \in (0, 1)$ and otherwise has no evidence.

$\mathcal{E}(t) = \{\{t\}, T\}$ or $\mathcal{E}(t) = \{T\}$.

This makes it useful to distinguish between t and "value" associated with t.

So change receiver's utility function to $-(a - v(t))^2$.

Let v^* be receiver's expectation of v if sender presents no evidence.

Sender won't present evidence if either (a) she can't present or (b) $v(t) \leq v^*$.

So

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

v^* is unique.

Equilibrium:

- If sender has evidence and $v(t) > v^*$, sender proves this.
- Sender types with proof that $v(t) \leq v^*$ pool with senders who have no evidence.
- Expectation in response to nondisclosure is v^*.

This is a workhorse model in economics, finance, and accounting.

Dye Evidence: Applications

Shin (Econometrica, 2003) uses this model of disclosure to understand stock price responses to disclosure.

Firm has N projects, each of which succeeds with probability $r \in (0,1)$ and fails otherwise.

If s projects succeed and $N-s$ fail, value of firm at time $T+1$ is $u^s d^{N-s}$ where $0 < d < u$.

At each $t = 1, \ldots, T$, manager has probability q of receiving evidence proving realization for any given project. If evidence received, manager chooses whether to disclose.

Market sets price at each t equal to expected value of firm given observations up through times t.

Manager wants highest possible stock price at each date.

Equilibrium: Manager discloses any success as soon as she can, will never disclose any failure.

Empirical implications:
Contrast stock prices with strategic disclosure versus exogenous (mandatory) disclosure.

Exogenous disclosure: Nondisclosure reveals nothing and so has no effect on stock price.

Strategic disclosure: Nondisclosure is bad news and so reduces stock price.

Exogenous disclosure: Effect of disclosing a success is independent of period of announcement.

Strategic disclosure: The effect of a late disclosure of success is larger than the effect of an early disclosure.

Exogenous disclosure: Uncertainty about future stock price not monotonic in current price.

Reason: Price is highest with good news, in the middle with no disclosure, and lowest with bad news. Uncertainty is highest in middle case.

Strategic disclosure: Uncertainty decreasing in stock price.

Reason: Price is higher with a lot of disclosure of successes, which is also when uncertainty is lower.

Shin shows data is more consistent with strategic disclosure.

Disclosure and Choice: BDL (*RES*, 2018)

Consider effect of these incentives on manager's choice of projects.

Period 0. Manager chooses project, choice not observed.

Period 1. She may get evidence proving outcome. If so, can disclose. Stock price = market's expectation of firm value given observations.

Period 2. Market observes realization = stock price.

Manager's utility is $\alpha\times$ "long–run" price $+(1-\alpha)\times$ "short–run" price.

Result: Strategic disclosure can lead to significant efficiency loss.

Idea: Manager discloses good information and suppresses bad.

Hence she has an incentive to take actions *ex ante* to influence this information revelation stage.

Such incentives are inefficient: manager has incentive to improve appearances even if they don't help (or even harm).

Seems wrong: Manager maximizes α times x plus $(1-\alpha)$ times short–run price.

Market can't be wrong in equilibrium, so expectation of short–run stock price is correct expectation of x.

So manager's equilibrium payoff is the expectation of x, so she should choose projects to maximize this, right?

Example:

Manager cares only about the short–run stock price ($\alpha = 0$).

Manager can disclose true value of firm at $t = 1$ with probability $q_1 \in (0, 1)$.

Two projects available:

- F_1 gives $x = 4$ with probability 1.
- F_2 gives $x = 6$ with probability $1/2$, 0 otherwise.

F_1 maximizes the expected value of the firm. But is it an equilibrium?

If so, short–run stock price if no evidence is disclosed is 4.

But then manager's payoff to deviating to F_2 is

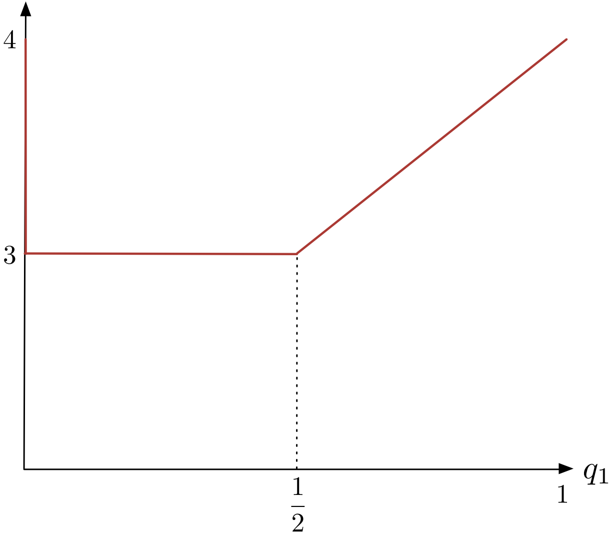
$$(1 - q_1)(4) + q_1 \left[\frac{1}{2}(4) + \frac{1}{2}(6) \right] > 4.$$

What happened?

Market cannot be fooled *in equilibrium.*

Out of equilibrium, market can be fooled and this might be better for the manager, as in the example.

This eliminates some equilibria, potentially (as in the example) making the manager and firm worse off.


4
3
$\frac{1}{2}$
1
q_1

Payoffs can be as low as 50% of the first best, but no lower (ruling out degenerate case where $\alpha = q_1 = 0$).

"First–best payoff:" Maximum expected value of firm over feasible projects.

Theorem 1: Fix any α, q_1, set of feasible projects, and any equilibrium. If Π^* is first–best payoff and Π equilibrium payoff, then unless $\alpha = q_1 = 0$,

$$\Pi \geq \frac{1}{2}\,\Pi^*.$$

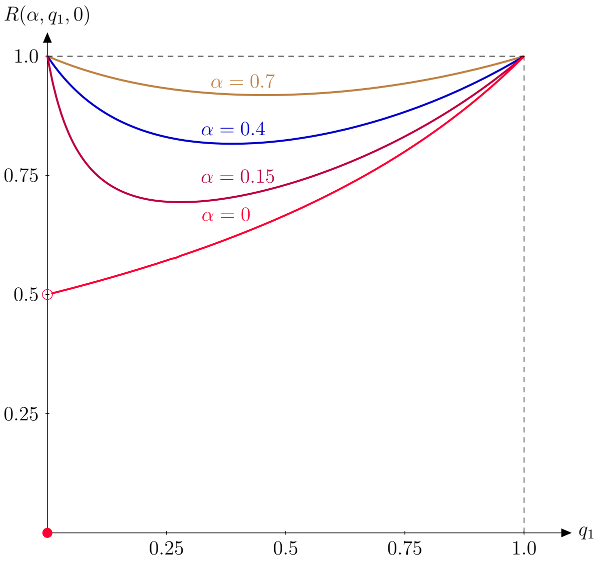
Lower bound is essentially attainable.

A more general result:

Theorem 2: Fix any α, q_1, set of projects, and any equilibrium. If Π^* is first–best payoff and Π equilibrium payoff, then unless $\alpha = q_1 = 0$,

$$\Pi \geq \left[\frac{\alpha + (1-\alpha)q_1}{\alpha + (1-\alpha)q_1(2-q_1)}\right] \Pi^* \geq \left[\frac{1+\sqrt{\alpha}}{2}\right] \Pi^*.$$

Bounds are essentially attainable.


$R(\alpha, q_1, 0)$
1.0
0.75
0.5
0.25
$\alpha = 0.7$
$\alpha = 0.4$
$\alpha = 0.15$
$\alpha = 0$
0.25
0.5
0.75
1.0
q_1

Other interesting applications of Dye evidence:. Archarya, DeMarzo, and Kremer (2011), Guttman, Kremer, and Skrzypacz (2013).

Also, more below under mechanism design.

Models without Unraveling: Time/Attention Constraints

Fishman and Hagerty (QJE, 1990): time/attention constraints prevent unraveling.

Good has N attributes, each of which could be *high* or *low*. Value of good v is number of h's.

For each attribute, seller has evidence which would prove whether it's h or ℓ. But she can only show evidence for one attribute.

So evidence structure not normal.

As before, price is buyer's expectation of v given information and seller wants highest possible price.

Equilibrium 1: Seller randomly picks one h attribute to show if she has one, shows an ℓ otherwise.

Buyer's response: Seeing an h means at least one attribute is h, seeing ℓ means none do.

So seller's strategy is optimal.

Equilibrium 2: Seller shows the "first" attribute for which she has an h.

If seller shows h for kth attribute, buyer learns first $k-1$ are ℓ.

Buyer gets a lot more information in second equilibrium.

See also Milgrom (*RAND*, 1981) for a related example; Glazer and Rubinstein (*GEB*, 2001; *Econometrica*, 2004).

Other approaches:

1. Verrecchia (*Jour of Accounting and Econ*, 1983).

Assume disclosure costly. So low types won't disclose.

2. Truthful lower bound. (Okuno–Fujiwara, Postlewaite, and Suzumura, *RES*, 1990; Dziuda, *JET*, 2011.)

$T = [\underline{t}, \overline{t}] \subset \mathbf{R}$.
$\mathcal{E}(t) = \{[s, \overline{t}] \mid s \leq t\}$.

t can prove type is *at least* s for any s for which the statement is true.

Example: If I have \$20, I can prove I have at least, say, \$10 by showing this amount.

But I could never prove I don't have more hidden.

Multi–Agent Games

Multiple senders with conflicting preferences can lead to separation under much weaker assumptions.

Milgrom and Roberts (RAND, 1986): With complete provability but limited rationality by the receiver and general sender preferences, full separation still possible with conflicting interests among senders.

Lipman and Seppi (JET, 1995) show separation under conflicting interests with much weaker evidence structures.

Example.

Lawyer 1 wants judge to rule that damages done to her client are large, lawyer 2 wants to prove damages small.

Only lawyer 2 has evidence. Can only pick one value of d and prove damages are *not* equal to d (if this is true).

Violates normality.

Sequential game:

1. Lawyer 1 makes a claim about d, say d_1.
2. Lawyer 2 provides one piece of evidence and makes her own claim about d, say d_2.
3. Judge rules on value of d.

Equilibrium with full separation:

- If lawyer 2 doesn't refute lawyer 1's claim, d_1, judge concludes $d = d_1$.
- If lawyer 2 refutes d_1, judge concludes $d = d_2$.

Clearly, given conficting interests, lawyer 1 will tell truth and lawyer 2 will be unable to refute.

Judge infers correctly, even if he doesn't know preferences of lawyers or range of possible d's.

Not possible in a simultaneous move game between lawyers.

See Hagenbach, Koessler, and Perez–Richet (*Econometrica*, 2014) for characterization of PBE with separation with multiple agents and partial provability.

Different effect of multiple agents: Onuchic and Ramos (2023).

Suppose agents are a team and jointly control disclosure decisions.

Surprising effects.

Example. 2 agents. $t_i \in T_i = \{0, 1, \ldots, K\}$, independent, full support priors, not necessarily identical.

Disclosure environment, meaning that true $t = (t_1, t_2)$ can be disclosed or not.

$\mathcal{E}(t_1, t_2) = \{\{(t_1, t_2)\}, T_1 \times T_2\}$.

Consider square–error loss model. Agent i wants receiver to infer highest possible t_i, $i = 1, 2$.

Case 1: Either agent can unilaterally disclose.

As in the one–agent case, we get unraveling. Best types of each agent would disclose, therefore next best will, etc.

Case 2: Disclosure occurs only if *both* agents agree.

Note: If no disclosure, receiver does not see who blocked it.

An equilibrium: Let t_i^* be receiver's expectation of t_i given no disclosure.

Then we get disclosure iff $t_i \geq t_i^*$ for *both* i.

$$t_1^* = \mathrm{E}\left[t_1 \mid t_1 \leq t_1^* \text{ or } t_2 \leq t_2^*\right].$$

$$t_2^* = \mathrm{E}\left[t_2 \mid t_1 \leq t_1^* \text{ or } t_2 \leq t_2^*\right].$$

Like endogenous Dye, where i's ability to disclose is determined by j's incentive to do so.

So what's the best rule for disclosure?

Suppose agents privately take effort at a cost and consider disclosure rules maximizing effort.

Suppose types are determined by these efforts.

Case 1: Agent *i*'s effort shifts up the distribution only of *i*'s type.

Best disclosure rule is that either agent can disclose unilaterally. Gets unraveling and efficient effort, as in "Disclosure and Choice."

Case 2. Agent i's effort shifts up the distribution only of j's type.

Now requiring unanimity is better.

Intuition: i wants *option* to show t_i. Maximizes probability of having this option by making j want to disclose.

So i takes effort to push up t_j to induce j to want to disclose.

End of Part 1.