# Bayesian Persuasion and Information Design

Anton Kolotilin

Sydney, November, 2019

## Motivation

- What can be achieved purely by means of persuasion?

- An interested party wants to influence a decision-maker by disclosing decision-relevant information

- Full control over the decision-maker's information

## Examples

- Grading policy chosen by a university
- Clinical trial designed by a pharmaceutical company
- Media channel controlled by a government

## Today

- Tractable problems
- Various approaches
- Optimal information structures

# Kamenica and Gentzkow (2011)

## Model

- Two players: Sender and Receiver

- Receiver chooses action $a \in A$

- Prior belief $\mu_0$ over the state $\omega \in \Omega$

- Sender designs signal (experiment, information structure, etc.)

- Receiver's utility is $u(a, \omega)$ and Sender's utility is $v(a, \omega)$

## Signal

- Signal realization space $S$

- Stochastic matrix $\pi(s|\omega)$ for each $s \in S$ and $\omega \in \Omega$

- $\Pi$ is the set of all signals

## Timing

1. Sender designs a signal $\pi$

2. State $\omega$ is drawn from $\mu_0(\cdot)$

3. Signal realization $s$ is drawn from $\pi(\cdot|\omega)$

4. Receiver observes design $\pi$ and realization $s$

5. Receiver takes action $a$

## Bayesian persuasion: Interpretation

- Commitment
  - Sender is informed
  - Ex-ante can commit to the way of disclosing information
- Test
  - Sender is uninformed
  - Can design a test whose results are publicly revealed

## Grading policy

- School chooses a grading policy $\pi$
  - Student ability: $\omega$
  - Student's transcript: $s$
- School reveals transcript $s$
- Employer makes hiring decision $a$

## Drug approval process

- Pharma discovers a new drug and designs clinical trials $\pi$
  - True drug quality: $\omega$
  - Results of clinical trial: $s$
- Pharma reveals all results $s$
- FDA makes approval decision $a$

## Commitment to signal before state is realized

- Law requires to register tests and reveal results
- Cheap talk in large population implements optimal signal
- Sender can build reputation to choose optimal signal
- Optimal signal gives upper bound on influence

## Action-recommendation approach

- By the revelation principle, we can restrict attention to *direct signals* whose realizations are *action recommendations* obeyed by the receiver

- Proof: “merge” signal realizations that induce the same action and “label” signal realizations by actions they induce

- A signal is a decision rule: $\pi(a|\omega)$

## Two-step procedure

1. Characterize the set of obedient decision rules:

$$\mathbb{E}_{\mu_0}[u(a,\omega)\pi(a|\omega)] \geq \mathbb{E}_{\mu_0}[u(a',\omega)\pi(a|\omega)],\ \forall a, a' \in A$$

2. Choose an obedient decision rule that maximizes the sender's expected utility:

$$\mathbb{E}_{\mu_0,\pi}[v(a,\omega)]$$

## Binary example

- Sender wants to persuade Receiver to accept proposal
- Proposal is equally likely to be good ($\omega = 1$) or bad ($\omega = 0$)
- Receiver decides whether to accept ($a = 1$) or reject ($a = 0$)
- Sender's utility is $a$
- Receiver's utility is $(\omega - r)a$, where $r \in (1/2, 1)$

## Benchmark signals

- Completely uninformative signal:

$$\pi(a|1) = \pi(a|0) = 1 \text{ for } a = 0$$

- Fully informative signal:

$$\pi(1|1) = 1 \text{ and } \pi(1|0) = 0$$

- Partially informative signal:

$$\pi(1|1) = 1 \text{ and } \pi(1|0) = q$$

where obedience requires

$$(1 - r)\frac{1}{2} + (0 - r)q\frac{1}{2} \geq 0 \iff q \leq \frac{1 - r}{r}$$

## Optimal signal

- Probability of acceptance $\frac{1}{2}(1+q)$ is maximized when

$$\pi^*(1|1) = 1 \text{ and } \pi^*(1|0) = \frac{1-r}{r}$$

- Receiver is indifferent when recommended to accept

- Receiver is certain that $\omega = 0$ when recommended to reject

# Kolotilin (2015)

## Continuous example

- Sender wants to persuade Receiver to accept proposal
- Proposal's value $\omega$ is distributed with density $f$ on $[0, 1]$
- Receiver decides whether to accept ($a = 1$) or reject ($a = 0$)
- Sender's utility is $a$
- Receiver's utility is $(\omega - r)a$, where $r \in (\mathbb{E}[\omega], 1)$

## Optimal signal

**Lemma.** The optimal signal is given by

$$\pi^*(1|\omega) = \begin{cases} 1, & \text{if } \omega \geq \omega^* \\ 0, & \text{otherwise} \end{cases}$$

where $\omega^* \in (0, r)$ is the unique solution to

$$\mathbb{E}[\omega|\omega \geq \omega^*] = r.$$

- Receiver is indifferent when recommended to accept
- Receiver learns whether $\omega$ is above or below the cutoff $\omega^*$

## Proof

The optimal signal $\pi^*(1|\omega)$ solves

$$\max_{\pi(1|\omega)\in[0,1]} \mathbb{E}[\pi(1|\omega)]$$

$$\text{s.t. } \mathbb{E}[(\omega - r)\pi(1|\omega)] \geq 0$$

The Lagrangian for this problem is

$$L = \mathbb{E}[(1 + \lambda(\omega - r))\pi(1|\omega)]$$

so

$$\pi^*(1|\omega) = \begin{cases} 1, & \text{if } \omega \geq r - \frac{1}{\lambda} \\ 0, & \text{otherwise} \end{cases}$$

## Comparative Statics

$F_2$ is higher than $F_1$ in the increasing convex order ($F_2 \geq_{icx} F_1$) if there exists $F$ such that $F_2 \geq_{FOSD} F \geq_{MPS} F_1$

**Theorem.** Sender's expected utility is higher under $F_2$ than under $F_1$ for all $r$ iff $F_2 \geq_{icx} F_1$.

**Corollary.** Let $F_1$ and $F_2$ be such that $\mathbb{E}_{F_1}[\omega] = \mathbb{E}_{F_2}[\omega]$. Sender's expected utility is higher under $F_2$ than under $F_1$ iff $F_2 \geq_{MPS} F_1$.

# Kamenica and Gentzkow (2011)

## Belief approach

- After observing $s$, Receiver uses Bayes' rule to update his belief from the prior $\mu_0 \in \Delta(\Omega)$ to the *posterior* $\mu \in \Delta(\Omega)$

- Given posterior $\mu$, Receiver takes a *best-response* action

$$a^*(\mu) \in \arg\max \mathbb{E}_\mu[u(a, \omega)],$$

breaking possible ties in favor of Sender

- Sender's *indirect utility* from posterior $\mu$ is

$$V(\mu) = \mathbb{E}_\mu[v(a^*(\mu), \omega)]$$

- A signal $\pi$ induces a distribution $\tau$ over posteriors $\mu$, so Sender's expected utility is $\mathbb{E}_\tau[V(\mu)]$

## Splitting lemma

**Lemma.** There exists a signal $\pi \in \Pi$ that induces a distribution of posteriors $\tau \in \Delta(\Delta(\Omega))$ iff

$$\mathbb{E}_\tau[\mu] = \mu_0.$$

*Proof.* The only if part follows from the law of iterated expectations. The if part is shown by construction:

$$\pi(\mu|\omega) = \frac{\mu(\omega)\tau(\mu)}{\mu_0(\omega)}, \ \forall \mu \in \text{supp}(\tau).$$

## Concavification

Sender's problem is to find distribution of posteriors $\tau$ to

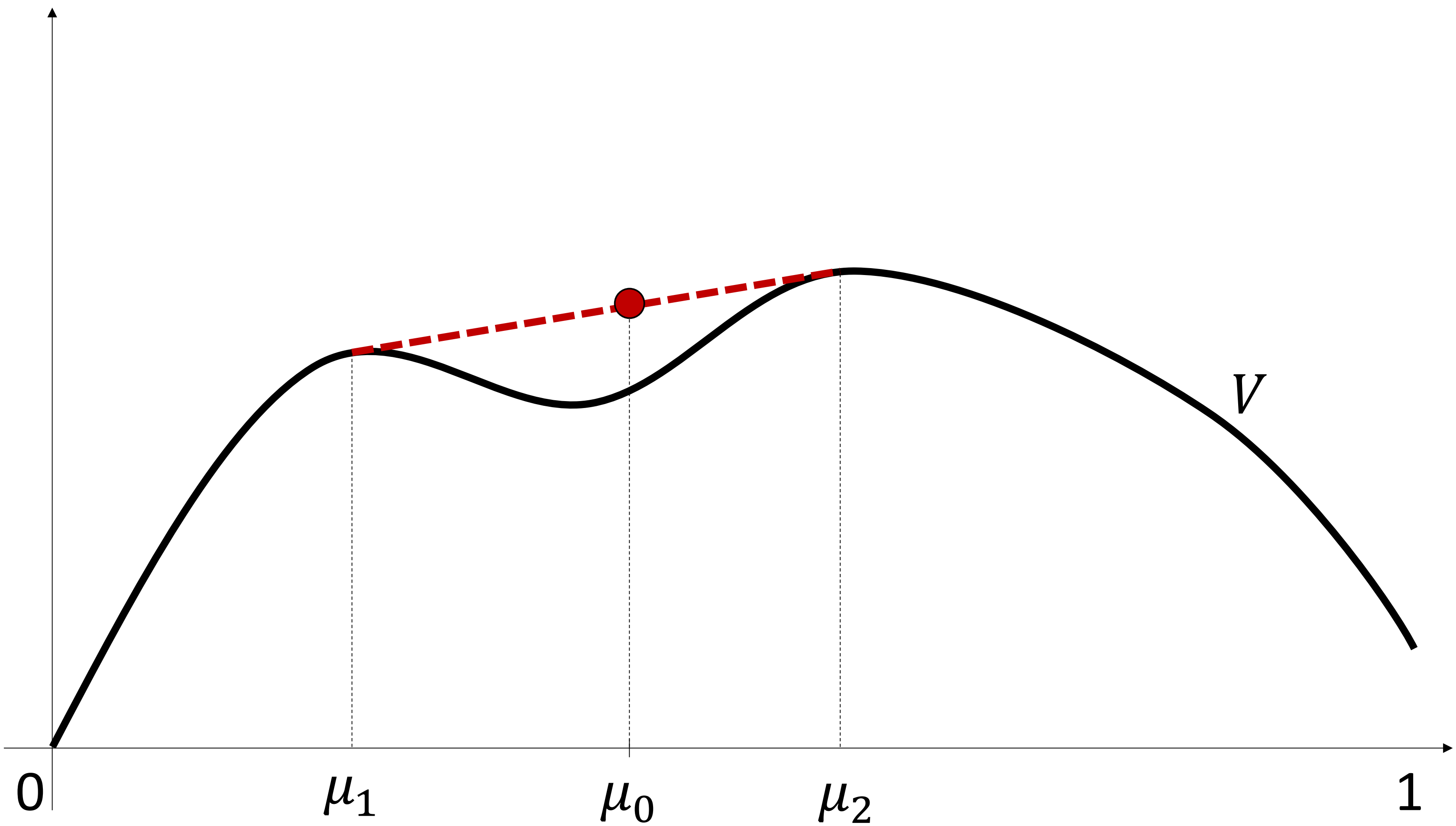
$$\text{maximize } \mathbb{E}_\tau[V(\mu)]$$

$$\text{subject to } \mathbb{E}_\tau[\mu] = \mu_0$$

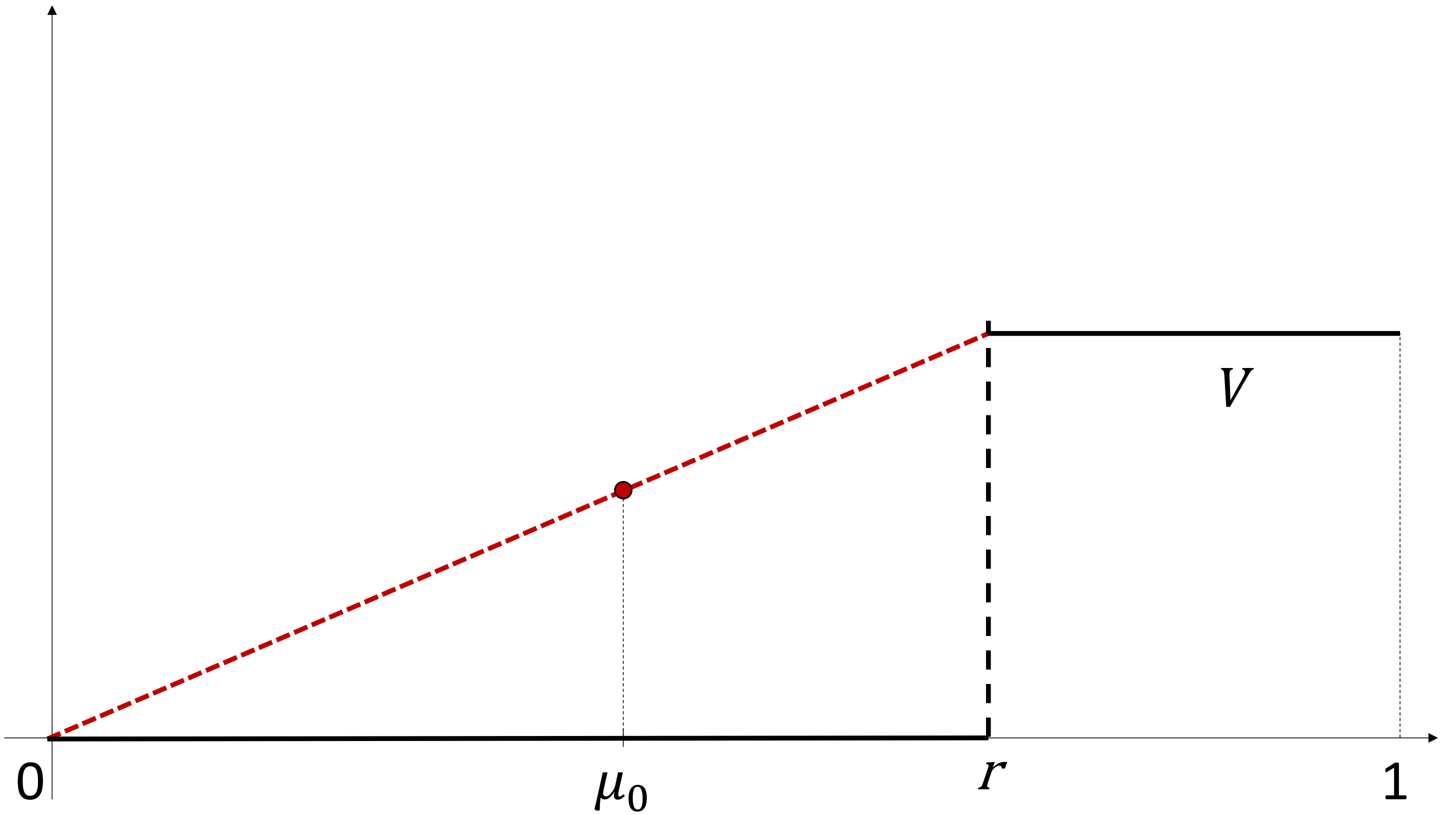
Smallest concave function that is everywhere greater than $V$ is called *concavification* of $V$ and is denoted by $\widehat{V}$

**Theorem.** The value of Sender's problem is $\widehat{V}(\mu_0)$.

## Illustration of concavification



# Binary example




Receiver takes $a = 1$ iff $(1-r)\mu - r(1-\mu) \geq 0$:

$$V(\mu) = \begin{cases} 1, & \text{if } \mu \geq r \\ 0, & \text{otherwise} \end{cases}$$

## Implications of concavification

**Corollary 1.** Persuasion is valuable iff $\widehat{V}(\mu_0) > V(\mu_0)$.

**Corollary 2.** If $V$ is concave, then no disclosure is optimal.

**Corollary 3.** If $V$ is convex, then full disclosure is optimal.

**Corollary 4.** There exists an optimal signal with $|S| \leq |\Omega|$.

# Dworczak and Kolotilin (2019)

## Primal problem

Primal problem is to find distribution of posteriors $\tau$ to

$$\text{maximize } \mathbb{E}_\tau[V(\mu)]$$

$$\text{subject to } \mathbb{E}_\tau[\mu] = \mu_0$$

Analogy to a linear production problem:

- $\mu_0$ – producer's endowment with resources $\omega$
- $\Delta(\Omega)$ – available linear production processes $\mu$
- $V(\mu)$ – income from operating process $\mu \in \Delta(\Omega)$ at unit level
- $\tau$ – production plan specifies operation level $\tau(\mu)$ of process $\mu$

## Dual problem

Dual problem is to find *price function* $P : \Omega \to \mathbb{R}$ to

$$\text{minimize } \mathbb{E}_{\mu_0}[P(\omega)]$$

$$\text{subject to } \mathbb{E}_{\mu}[P(\omega)] \geq V(\mu),\ \forall \mu \in \Delta(\Omega)$$

Analogy to a linear production problem:

- Wholesaler wants to buy out producer
- $P(\omega)$ – unit price for each resource $\omega$
- Prices make producer willing to sell all resources
- Wholesaler minimizes total cost of buying up all resources

## Weak duality

**Theorem.** Feasible $\tau$ and $P$ satisfy

$$\mathbb{E}_\tau[V(\mu)] \leq \mathbb{E}_{\mu_0}[P(\omega)].$$

Moreover, if $\mathbb{E}_\tau[V(\mu)] = \mathbb{E}_{\mu_0}[P(\omega)]$, then $\tau$ and $P$ are optimal.

*Interpretation:*

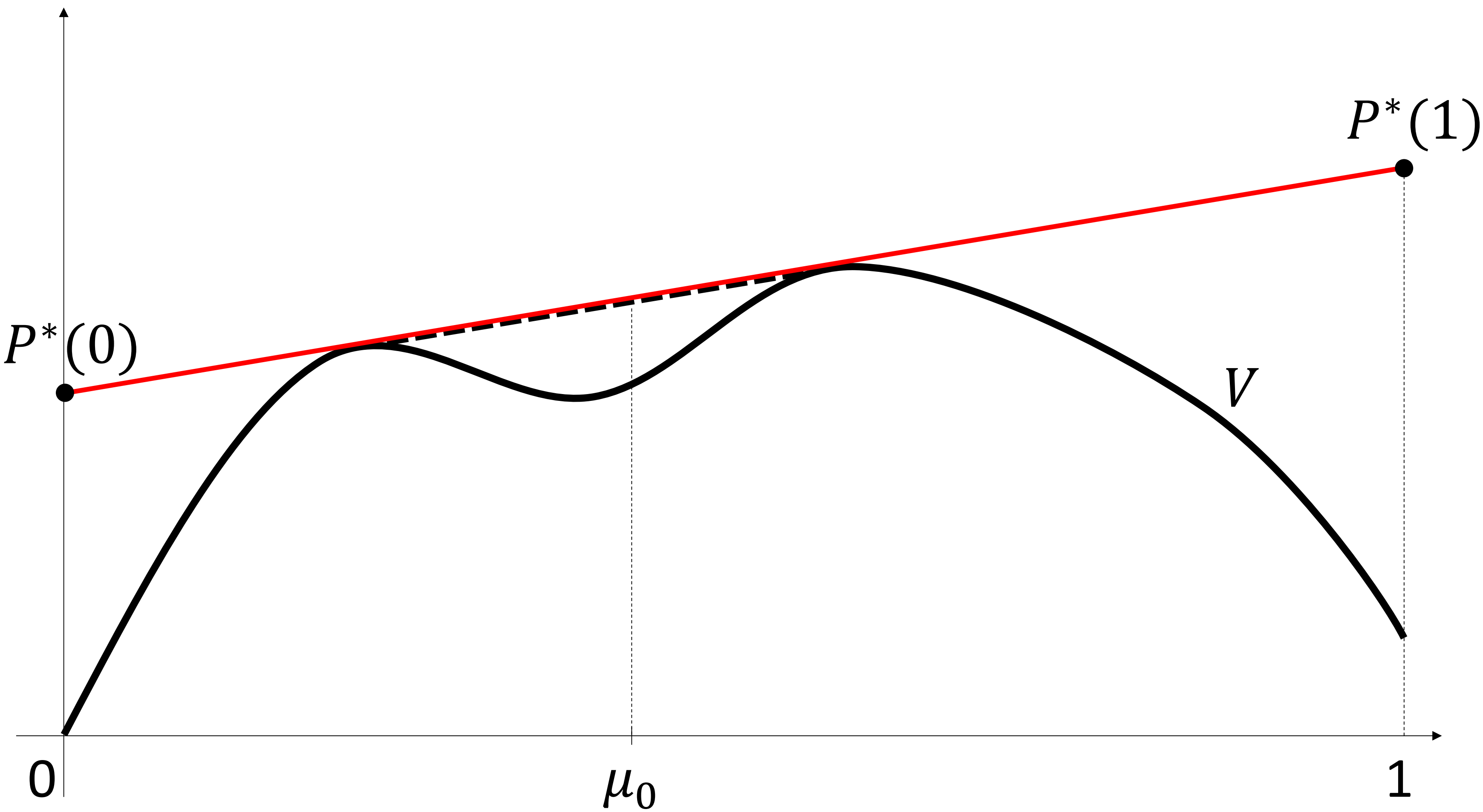
- Total income of producer cannot exceed total cost of resources under prices that make producer willing to sell

- If there exists production plan and prices that equalize total income and total cost, then they are optimal

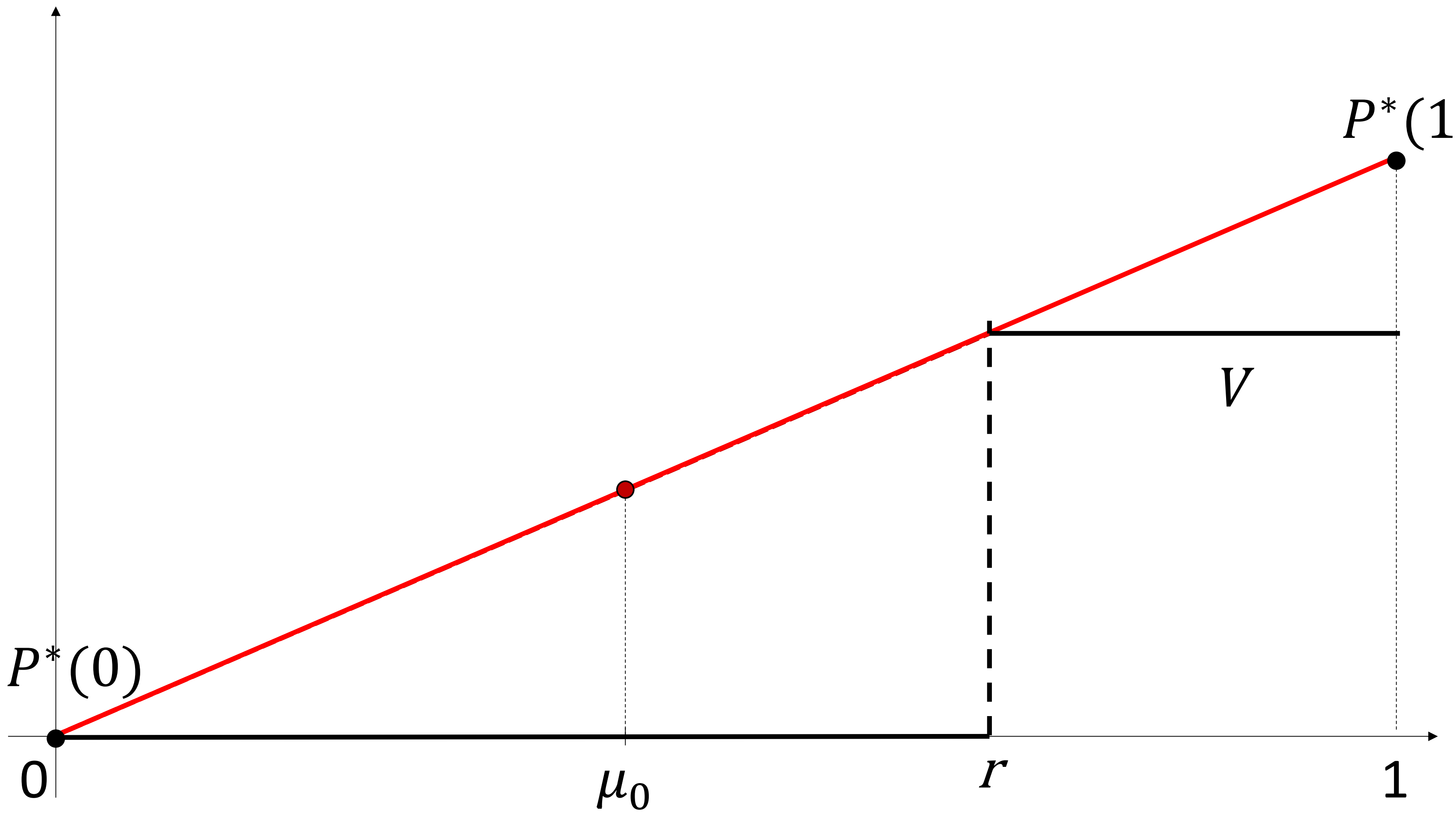
## Strong duality

**Theorem.** Optimal $\tau$ and $P$ that satisfy $\mathbb{E}_\tau[V(\mu)] = \mathbb{E}_{\mu_0}[P(\omega)]$ exist iff concave closure $\widehat{V}$ is superdifferentiable at $\mu_0$.

Illustration



# Binary example



## Complementary slackness

**Corollary.** Feasible $\tau$ and $P$ are optimal iff

$$\text{supp}(\tau) \subset \{\mu \in \Delta(\Omega) : \mathbb{E}_\mu[P(\omega)] = V(\mu)\}.$$

*Interpretation:* If operation cost of process $\mu$ exceeds its income, then it is not operated

## Full disclosure

- Under full disclosure, $\text{supp}(\tau)$ is the set of all $\delta_\omega$

- Complementary slackness requires $P(\omega) = V(\delta_\omega)$ for all $\omega$

- Full disclosure is optimal iff $P(\omega) = V(\delta_\omega)$ is feasible:

$$\mathbb{E}_\mu[V(\delta_\omega)] \geq V(\mu),\ \forall \mu \in \Delta(\Omega)$$

## Linear persuasion problem

Gentzkow and Kamenica (2016), Kolotilin et al. (2017), Kolotilin (2018), and Dworczak and Martini (2019)

**Assumption.** $\Omega = [\underline{\omega}, \overline{\omega}]$ and $V(\mu) = v(\mathbb{E}_\mu[\omega])$ for some $v$

- Binary and continuous examples above are special cases

- Prior $\mu_0$ is represented by distribution $F$

## Majorization

Key simplification: distribution $\tau$ of posteriors $\mu$ matters only through induced distribution $G$ of posterior means $\mathbb{E}_\mu[\omega]$

**Lemma.** There exists $\tau$ that induces distribution $G$ iff

$$F \geq_{MPS} G.$$

*Proof:* The only if part holds because any signal is a garbling of the fully informative signal. The if part follows from Strassen's theorem: there exists a joint distribution $J$ of $a$ and $\omega$ such that the marginal distribution of $\omega$ is $F$, the marginal distribution of $a$ is $G$, and $\mathbb{E}_J[\omega|a] = a$.

## Primal problem

Primal problem is to find distribution of posterior means $G$ to

$$\text{maximize } \mathbb{E}_G[v(x)]$$

$$\text{subject to } G \leq_{MPS} F$$

*Interpretation:*

- $\Omega$ describes both resources and production processes
- Process $x \in \Omega$ generates income $v(x)$
- Producer can transform measure $\mu$ of resources into one unit of resource $\mathbb{E}_\mu[\omega]$

## Dual problem

Dual problem is to find *price function* $p(x)$ to

$$\text{minimize } \mathbb{E}_F[p(x)]$$

$$\text{subject to } p \geq v$$

$$p \text{ is convex}$$

*Interpretation:* If price function failed to be convex, producer could sell at effectively higher prices by using transformations

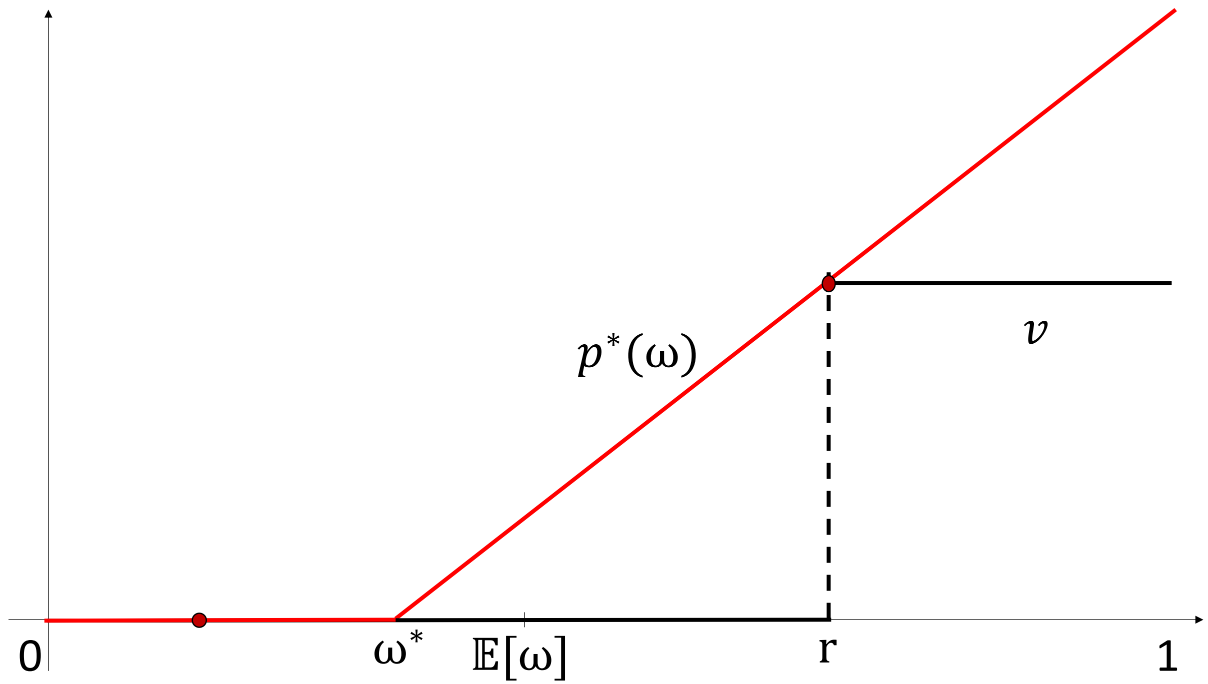
## Duality theorem

**Theorem.** Feasible $G$ and $p$ are optimal iff $\mathbb{E}_G[v(x)] = \mathbb{E}_F[p(x)]$.

*Key idea:* Prices of posterior means $p(x)$ can be derived from prices of states $P(\omega)$ by finding "cheapest" way in which given posterior mean can be generated by pooling states:

$$p(x) = \min \{\mathbb{E}_\mu[P(\omega)] : \mu \in \Delta(\Omega) \text{ and } \mathbb{E}_\mu[\omega] = x\}$$

# Continuous example




Receiver takes action $a = 1$ iff $\mathbb{E}_\mu[\omega] - r \geq 0$:

$$v(x) = \begin{cases} 1, & \text{if } x \geq r \\ 0, & \text{otherwise} \end{cases}$$

Optimal signal reveals $\omega \geq \omega^*$ or $\omega < \omega^*$, s.t. $\mathbb{E}[\omega | \omega \geq \omega^*] = r$

# Kolotilin (2018)

## Duality for action-recommendation approach

Linear persuasion problem:

- $A = \Omega = [0, 1]$
- Sender's utility is $v(a)$ and receiver's utility is $-(a - \omega)^2$
- Given posterior $\mu$, Receiver's best-response action is $\mathbb{E}_{\mu}[\omega]$
- Sender's indirect utility from posterior $\mu$ is $v(\mathbb{E}_{\mu}[\omega])$

## Primal and dual problems

Primal problem is to find joint distribution $J$ of recommended action $a$ and state $\omega$ to

$$\text{maximize } \mathbb{E}_J[v(a)]$$

$$\text{subject to } J(1, \omega) = F(\omega),\ \forall \omega \in \Omega$$

$$\mathbb{E}_J[\omega | a] = a,\ \forall a \in A$$

Dual problem is to find functions $q(\omega)$ and $r(a)$ to

$$\text{minimize } \mathbb{E}_F[q(\omega)]$$

$$\text{subject to } q(\omega) + r(a)(a - \omega) \geq v(a),\ \forall (a, \omega) \in A \times \Omega$$

## Equivalence of dual problems

If $p^*$ is optimal price function, then $(q^*, r^*) = (p^*, \partial p^*)$ is optimal,

$$p^*(\omega) + \partial p^*(a)(a - \omega) \geq p^*(a) \geq v(a)$$

because $p^*$ is convex and $p^* \geq v$

## Upper-censorship

- A signal is upper-censorship with cutoff $\omega^* \in [0, 1]$ if
  - states $\omega < \omega^*$ are separated and states $\omega \geq \omega^*$ are pooled

- Induced action is $\omega$ for $\omega < \omega^*$ and $a^* = \mathbb{E}[\omega | \omega \geq \omega^*]$ for $\omega \geq \omega^*$

- Full disclosure is degenerate upper-censorship with $\omega^* = 1$

- No disclosure is degenerate upper-censorship with $\omega^* = 0$

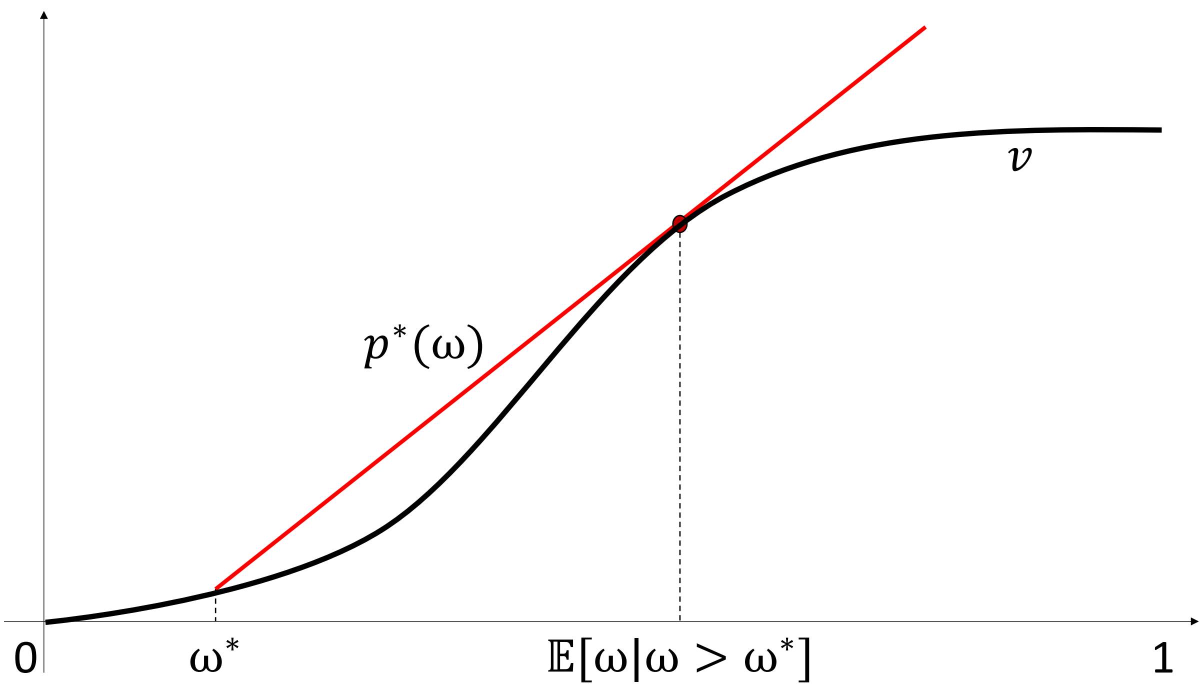
## Upper-censorship

**Theorem.** Upper-censorship with cutoff $\omega^*$ is optimal iff

$$v(a) \text{ is convex on } [0, \omega^*],$$

$$v(a) \leq v(a^*) + v'(a^*)(a - a^*),\ \forall a \in [\omega^*, 1],$$

with equality at $a = \omega^*$.



# Kolotilin et al. (2017)

## Model

- Receiver decides whether to accept ($a = 1$) or reject ($a = 0$)
- Prior distribution $F$ of state $\omega \in [0, 1]$
- Prior distribution $G$ of receiver's private type $r \in [0, 1]$
- Random variables $r$ and $\omega$ are independent
- Receiver's utility is $(\omega - r)a$ and Sender's utility is $a$

## Receiver's Private Information: Interpretation

- One Receiver with a private type
- A continuum of heterogeneous Receivers

## Equivalence to linear persuasion

- Given posterior $\mu$, receiver accepts ($a = 1$) iff $r \leq \mathbb{E}_\mu[\omega]$

- Sender's indirect utility from posterior $\mu$ is

$$V(\mu) = \mathsf{Pr}(r \leq \mathbb{E}_\mu[\omega]) = G(\mathbb{E}_\mu[\omega])$$

## Timing

1. Sender designs a menu of signals

2. State and receiver's type are drawn

3. Receiver observes his type

4. Receiver chooses a signal from the menu

5. Receiver observes the signal realization

6. Receiver takes an action

## Persuasion mechanism

- A direct mechanism $\pi : R \times \Omega \to \Delta A$
  - asks Receiver to report $\widehat{r}$
  - recommends $\widehat{a} = 1$ w/pr $\pi(\widehat{r}, \omega)$ and $\widehat{a} = 0$ o/w
- WLOG restrict attention to mechanisms in which
  - Receiver is honest and obedient
- What can persuasion mechanisms gain over signals?

# Incentive-compatible mechanisms

## Notation

- For a mechanism $\pi$, define
  - probability that $r$ accepts ($a = 1$)

$$q_\pi(r) = \int \pi(r, \omega)\, dF(\omega)$$

  - expected value of $\omega$ given $r$ and $a = 1$

$$p_\pi(r) = \frac{1}{q_\pi(r)} \int \omega\pi(r, \omega)\, dF(\omega)$$

  - expected utility of type $r$

$$U_\pi(r) = \int (\omega - r)\pi(r, \omega)\, dF(\omega) = q_\pi(r)(p_\pi(r) - r)$$

    * analogy: standard linear trading environment

## Incentive compatible mechanisms

**Lemma.** A (feasible) mechanism $\pi$ is incentive compatible iff

$$\begin{aligned} q_\pi \quad & \text{is} \quad \text{non-increasing} \\ U_\pi(r) &= \int_r^1 q_\pi(s)ds, \\ U_\pi(0) &= \int_0^1 \omega dF(\omega) = \mathbb{E}[\omega], \end{aligned}$$

- there are obedience instead of individual rationality constraints

- there are also no transfers

## Incentive compatible mechanisms

**Lemma.** A mechanism $\pi$ is incentive compatible if and only if

$$\begin{aligned} q_\pi \quad & \text{is} \quad \text{non-increasing} && (1) \\ U_\pi(r) &= \int_r^1 q_\pi(s)ds, && (2) \\ U_\pi(0) &= \int_0^1 q_\pi(s)ds = \mathbb{E}[\omega], && (3) \end{aligned}$$

Differences from Mirrlees:

- There are two boundary conditions, for $r = 1$ and $r = 0$
- Not all $q_\pi$'s that satisfy (1) – (3) are implementable
  - The set of implementable $q_\pi$'s depends on $F$ not only through $\mathbb{E}_F[\omega]$

Lemma is not too useful:

- What is the set of implementable $(q_\pi, U_\pi, V_\pi)$?

- What can persuasion mechanisms gain over signals?

# Implementable utility profiles

## Utility schedules

- A pair ($q_\pi$,$V_\pi$) is pinned down by $U_\pi$:

$$V_\pi(r) = q_\pi(r) = -U'_\pi(r)$$

- We can focus on implementable utility profiles $U_\pi$

## Necessary condition 1

- No information

$$\underline{U}(r) = \max\{\mathbb{E}[\omega] - r, 0\}$$

- Full information

$$\overline{U}(r) = \int_r^1 (\omega - r)\, dF(\omega)$$

- Thus, under any mechanism

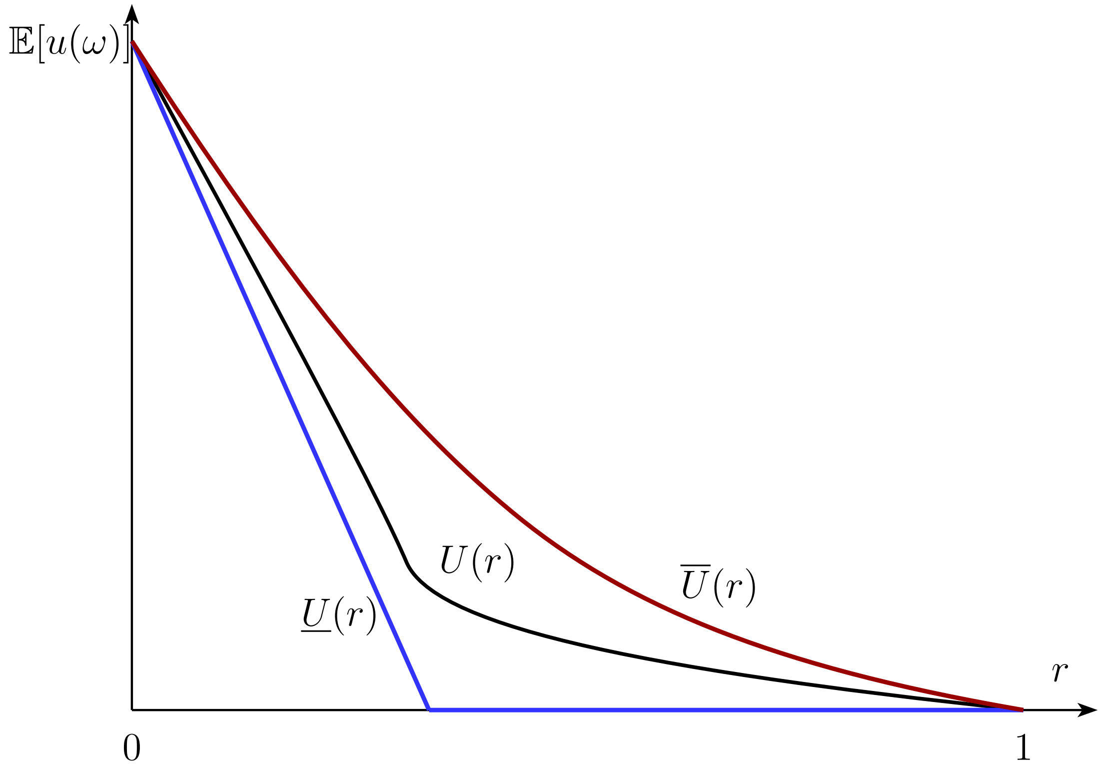
$$\underline{U}(r) \leq U_\pi(r) \leq \overline{U}(r)$$

## Necessary condition 2

Incentive compatibility $\Rightarrow$ $U_\pi$ is convex

(the standard Envelope Theorem argument)

# Necessary conditions




$$\underline{U} \leq U \leq \overline{U}$$

$U$ is convex

## Main Result

**Theorem.** The following statements are equivalent:

(a) $U$ is a convex function such that $\underline{U} \leq U \leq \overline{U}$

(b) $U$ is implementable by a persuasion mechanism

(c) $U$ is implementable by a signal

**Sketch of proof:** (c) $\Rightarrow$ (b) trivially

(b) $\Rightarrow$ (a) by monotonicity of $q$

(a) $\Rightarrow$ (c) Mirrlees meets Blackwell

## Sketch of Proof: (a) $\Rightarrow$ (c)

- Every signal can be equivalently described by distribution $H$ of posterior means $\mathbb{E}_\mu[\omega]$

- Fix convex $U$ such that $\underline{U} \leq U \leq \overline{U}$.

- Construction: Let $H(r) = 1 + U'(r)$

- $H$ is a distribution function:
  - $H(0) = 0$ because $\underline{U}'(0) = \overline{U}'(0) = -1$
  - $H(1) = 1$ because $\underline{U}'(1) = \overline{U}'(1) = 0$
  - $H(r)$ is increasing because $U$ is convex

## Sketch of Proof: (a) $\Rightarrow$ (c)

- Observe that $H(r) = 1 + U'(r)$ satisfies

$$\int_r^1 (1 - H(s))\mathrm{d}s = U(r) \leq \overline{U}(r) = \int_r^1 (1 - F(s))\mathrm{d}s.$$

  where $1 - F(r)$ is probability that fully informed Receiver $r$ chooses to accept ($a = 1$).

- Therefore, $F$ is a mean-preserving spread of $H$

- Thus, there exists a required signal that induces the distribution $H$ of posterior means

$QED$

# Optimal signals

## Sender's Problem

**Lemma.** For every incentive-compatible mechanism $\pi$,

$$\int_R V_\pi(r)\mathrm{d}G(r) = g(0)\mathbb{E}[\omega] + \int_R U_\pi(r)g'(r)\mathrm{d}r,$$

*Proof.* Integration by parts.

## Sender's Problem

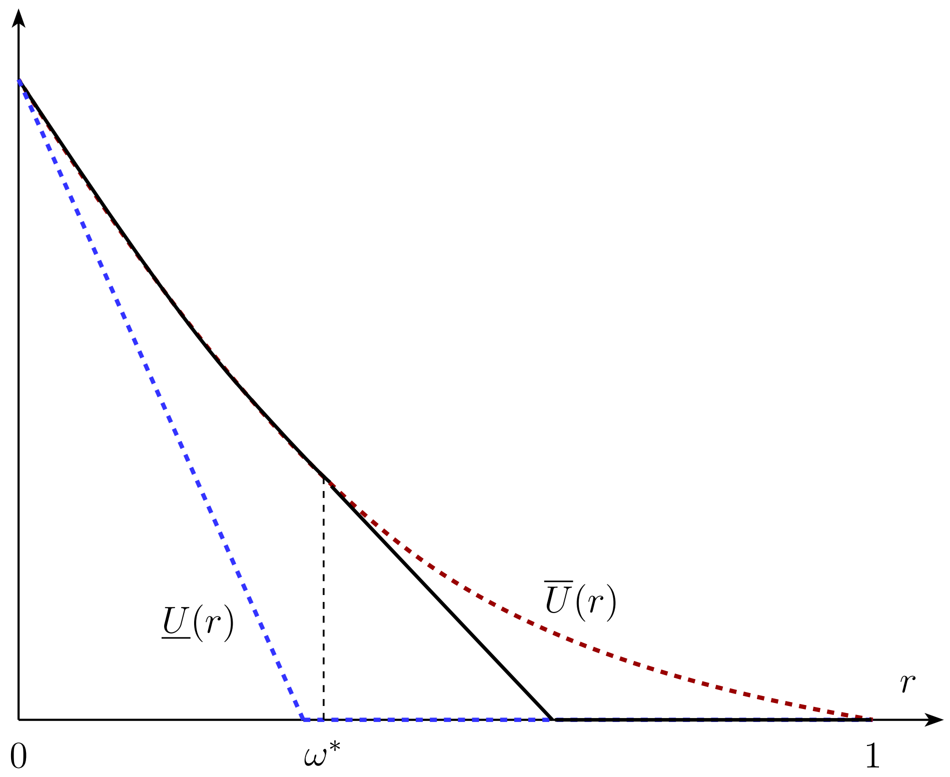
Sender's problem is to find $U$ to

$$\text{maximize} \int_R U(r) g'(r) \mathrm{d}r$$

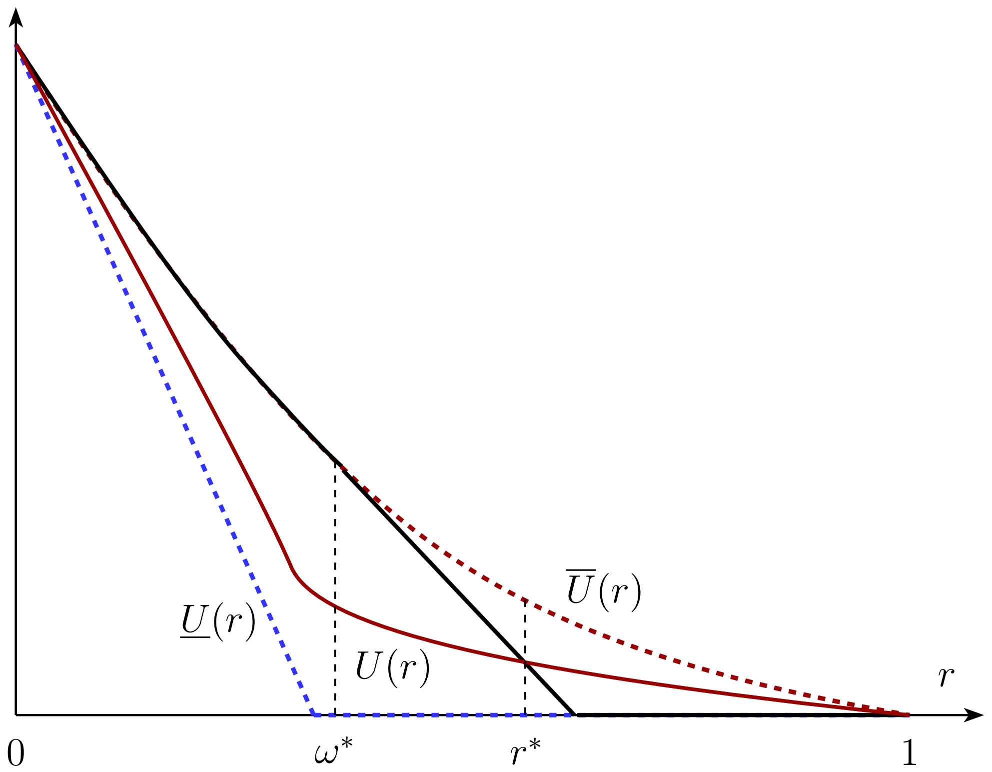
$$\text{subject to } U \text{ is convex and } \underline{U} \leq U \leq \overline{U}$$

Recall this is a linear persuasion problem with $V(\mu) = G(\mathbb{E}_\mu[\omega])$

# Upper-censorship



# Optimality of upper-censorship




$$G \text{ is S-shaped:} \quad g'(r) \begin{cases} \leq 0, & \text{if } r > r^* \\ \geq 0, & \text{otherwise} \end{cases}$$

## Extensions

- Multiple Receivers
- Multiple Senders
- Dynamics