Screening for Experiments

Daehong Min
Korea Information Society Development Institute

Seoul National University
Tae-Sung Kim Memorial Seminar

April 25, 2024

Motivation

Information Design Problems:

$$ID \rightarrow Information\ Structure \rightarrow DM$$

- The FDA (DM) vs. a drug company (an Information Designer)
  - Conflict of interests $\Rightarrow$ misaligned preferences over experiments
  - The FDA might want to control the quality of drug experiments
  - But the FDA might not know what experiments are feasible for the drug company

Q What is an optimal decision rule for DM when DM does not know what experiments an information designer can conduct?

Motivation

Information Design Problems:

$$DM \rightarrow Decision\ Rule \rightarrow ID \rightarrow Information\ Structure \rightarrow DM$$

e.g., The FDA (DM) vs. a drug company (an Information Designer)

- Conflict of interests $\Rightarrow$ misaligned preferences over experiments
- The FDA might want to control the quality of drug experiments
- But the FDA might not know what experiments are feasible for the drug company

Q What is an optimal decision rule for DM when DM does not know what experiments an information designer can conduct?

Literature Review

- Information Design
  - Kamenica and Gentzkow (2011), Bergemann and Morris (2016)
- Mechanism Design
  - Myerson (1979), Green and Laffont (1986)
- Information Design + Mechanism Design
  - Kolotilin et al. (2017), Yoder (2022)

The Model

Basic Setup

- Two players: Agent (a drug company) and Principal (the FDA)
- Two states of the world: $\{G, B\}$ with $P(G) = p$.
- Preferences:

	Good	*Bad*
*A*pprove	1,1	1,0
*D*isapprove	0,0	0,1

The Model

AGENT'S SET OF ACTIONS

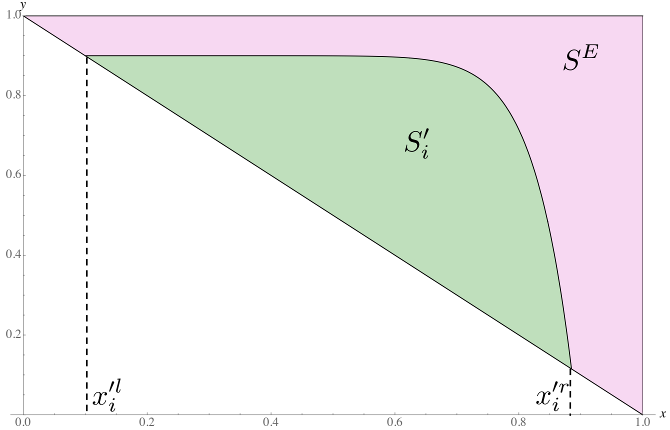
- Agent's type space: $\{\theta_1, \theta_2\}$ with $P(\theta_1) = t$
- S_i: Type θ_i's set of feasible experiments (convex & closed)
- A typical element in S_i: a binary experiment, π_i^k,

	Good state	Bad state
g (positive)	$\pi_i^k(g\|G) = y_i^k$	$\pi_i^k(g\|B) = 1 - x_i^k$
b (negative)	$\pi_i^k(b\|G) = 1 - y_i^k$	$\pi_i^k(b\|B) = x_i^k$

Assumptions on S_i

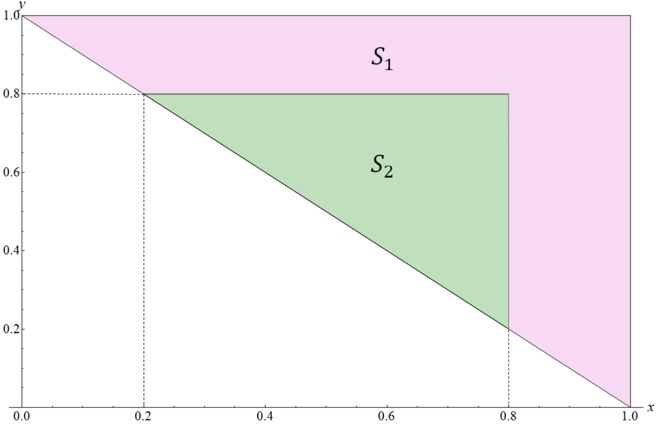
(a) S_i is closed and convex.

(b) The "northeast boundary" of S_i is *non-increasing* and *concave*.



Assumptions on S_i

(c) $S_2 \subset S_1$



The Model

Timing of the Game & Principal's Decision Rule

- Principal $\rightarrow$ Decision Rule $\rightarrow$ Agent $\rightarrow$ Experiment $\rightarrow$ Decision
  - Decision Rule: contingent on observables

  i.e., $d : S_1 \times \{g, b\} \rightarrow \Delta\{A, D\}$, c.f. $S_1 = S_2 \cup S_1$

- By the revelation principle, the principal's decision rule is equivalent to an incentive compatible menu offer
  - $M = \{(\pi_1, A_g^1, A_b^1), (\pi_2, A_g^2, A_b^2)\}$,

    A_g^i: the probability of action A associated with π_i and g,

    A_b^i: the probability of action A associated with π_i and b.

Principal's Problem

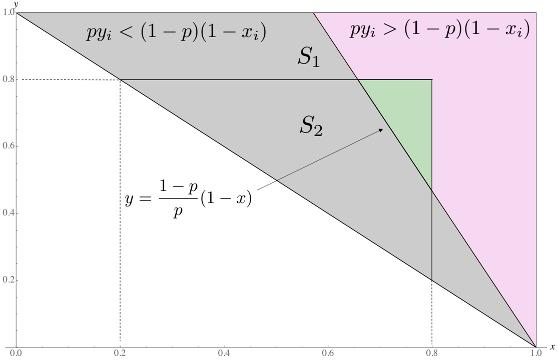
- Given $M = \{(\pi_1, A_g^1, A_b^1), (\pi_2, A_g^2, A_b^2)\}$, denote (π_i, A_g^i, A_b^i) by m_i for $i = 1, 2$

$$\begin{aligned}
& \max_{m_1, m_2} t \cdot EU^P(m_1) + (1-t) \cdot EU^P(m_2) \\
s.t. \quad & EU^A(m_1|\theta_1) \geq EU^A(m_2|\theta_1) \quad \text{(IC for type } \theta_1) \\
& EU^A(m_2|\theta_2) \geq EU^A(m_1|\theta_2) \quad \text{(IC for type } \theta_2) \\
& \pi_1 \in S_1 \text{ and } \pi_2 \in S_2 \quad \text{(Feasibility Constraints)}
\end{aligned}$$

$$\begin{aligned}
EU^P(m_i) &= (1-p) - (1-2p)A_b^i \\
&\quad + (A_g^i - A_b^i)(\underbrace{py_i}_{true\ positive} - \underbrace{(1-p)(1-x_i)}_{false\ positive}) \\
EU^A(m_i|\theta_i) &= A_b^i + (A_g^i - A_b^i)\underbrace{(py_i + (1-p)(1-x_i))}_{Postive\ outcome}
\end{aligned}$$

Ex post Optimality

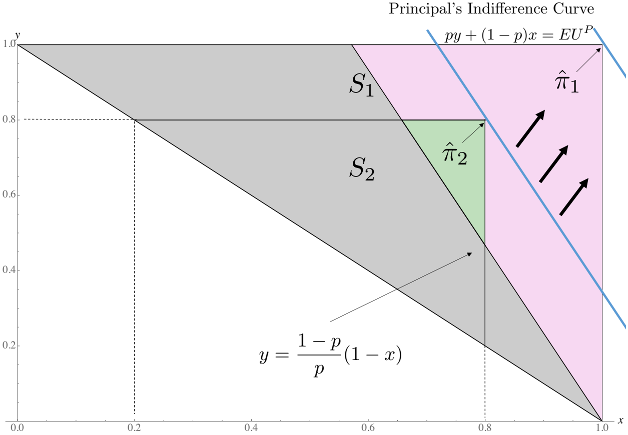
$$EU^P(m_i) = (1-p) - (1-2p)A_b^i + (A_g^i - A_b^i)(\underbrace{py_i}_{true\ positive} - \underbrace{(1-p)(1-x_i)}_{false\ positive})$$



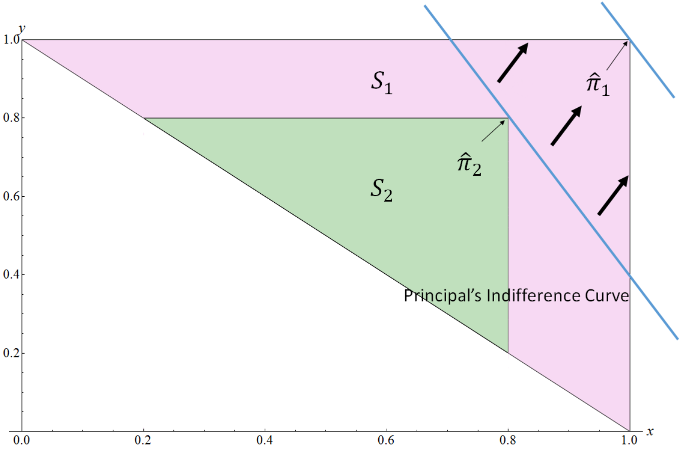
- Ex post Optimality: $(A_g^i, A_b^i) = (1, 0)$

Favorite Experiment in S_i: $\hat{\pi}_i$

$$\begin{aligned} EU^P(m_i) &= (1-p) - (1-2p)A_b^i + (A_g^i - A_b^i)(py_i - (1-p)(1-x_i)), \\ &= py_i + (1-p)x_i \ \textit{if} \ (A_g^i, A_b^i) = (1,0). \end{aligned}$$



The First-best Outcome: Observable Types



- The (ex ante) First-best Outcome: $[(\hat{\pi}_1, 1, 0), (\hat{\pi}_2, 1, 0)]$,

  Favorite Experiments and *Ex post Optimality*

Summary of Results

Depending on the properties of S_1 and S_2,

- Cases in which the first-best outcome is achievable
  - $\{(\hat{\pi}_1, 1, 0), (\hat{\pi}_2, 1, 0)\}$

Cases in which the first-best outcome is not achievable:

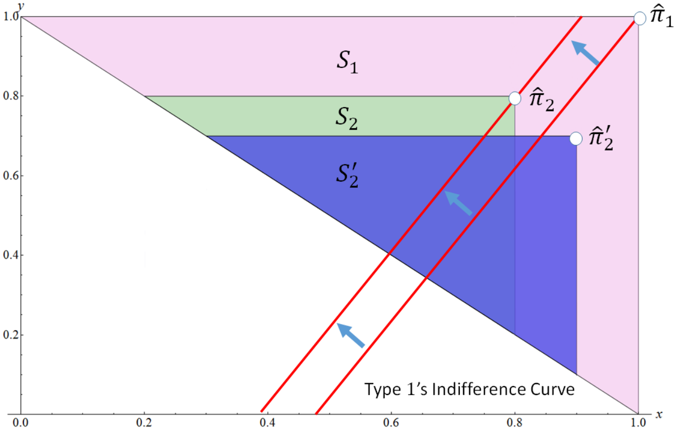
Favorite experiments vs. Ex post Optimality

- Favorite Experiments with distortion in ex post decisions
  - $\{(\hat{\pi}_1, A_g^1, A_b^1), (\hat{\pi}_2, A_g^2, A_b^2)\}$
  - optimal if $\hat{\pi}_i$ meets *quality* requirements
- Ex post optimal decisions with distortion in assigning experiments
  - $\{(\pi_1, 1, 0), (\pi_2, 1, 0)\}$
  - optimal if $\hat{\pi}_i$ fails to meet *quality* requirements

Cases when the FB outcome is achievable

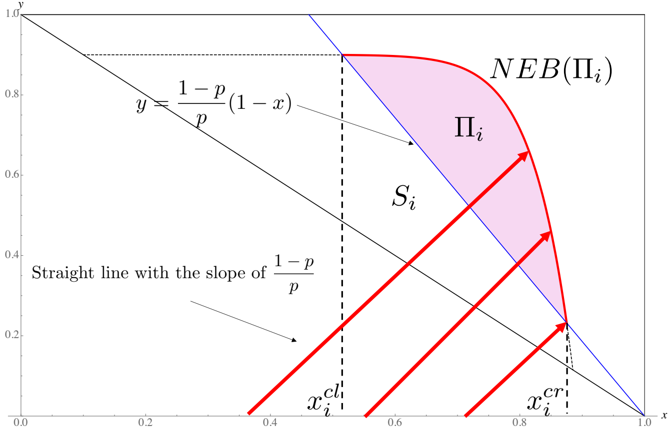
Remark

The principal can achieve the first-best outcome if $\hat{\pi}_1$ generates the positive outcome more frequently than $\hat{\pi}_2$



$$M = \{(\hat{\pi}_1, 1, 0), (\hat{\pi}_2, 1, 0)\} \text{ vs. } M' = \{(\hat{\pi}_1, 1, 0), (\hat{\pi}_2', 1, 0)\}$$

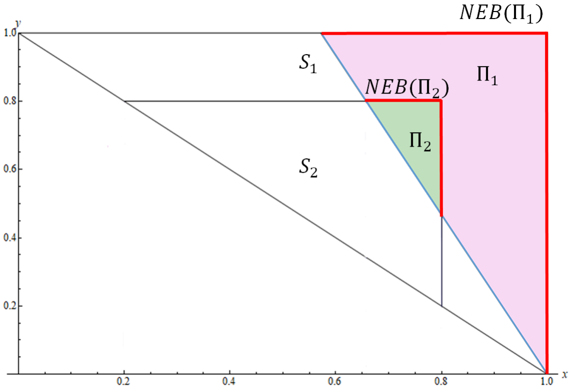
Simplifying the Problem: Experiments



Lemma

An optimal decision rule must assign π_i on the $NEB(\Pi_i)$ for $i = 1, 2$.

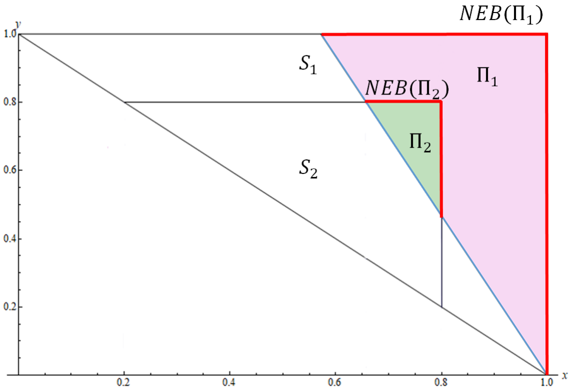
Simplifying the Problem: IC Constraints



$$\max_{m_1, m_2} t \cdot EU^P(m_1) + (1-t) \cdot EU^P(m_2)$$

$$\begin{aligned} s.t. \quad & EU^A(m_1|\theta_1) \geq EU^A(m_2|\theta_1) \quad (\text{IC for type } \theta_1) \\ & EU^A(m_2|\theta_2) \geq EU^A(m_1|\theta_2) \quad (\text{IC for type } \theta_2) \\ & \pi_1 \in NEB(\Pi_1) \text{ and } \pi_2 \in NEB(\Pi_2) \end{aligned}$$

Simplifying the Problem: IC Constraints



$$\max_{m_1, m_2} t \cdot EU^P(m_1) + (1-t) \cdot EU^P(m_2)$$

$$s.t. \quad EU^A(m_1|\theta_1) \geq EU^A(m_2|\theta_1) \quad \text{(IC for type } \theta_1\text{)}$$

$$\pi_1 \in NEB(\Pi_1) \text{ and } \pi_2 \in NEB(\Pi_2)$$

Two "Best" Experiments

according to other quality measures

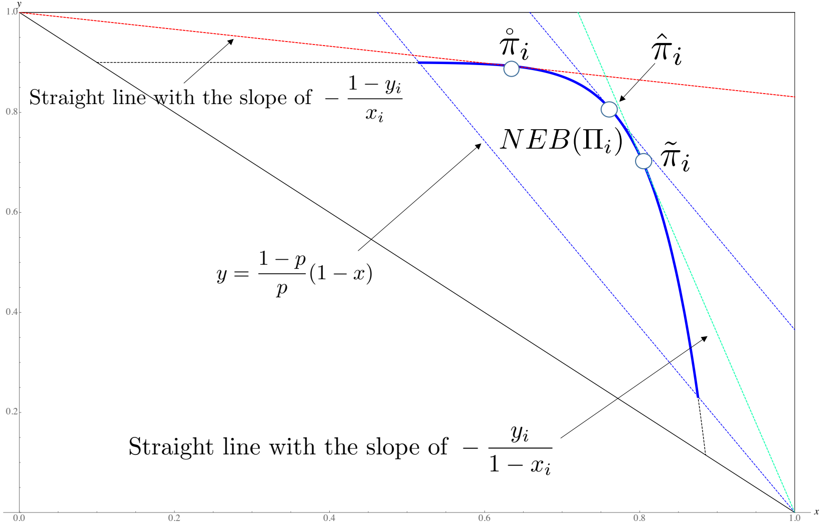
Definition

(1) $\tilde{\pi}_i$ is the experiment which maximizes the *positive likelihood ratio*, $\frac{y_i}{1-x_i}$.

(2) $\mathring{\pi}_i$ is the experiment which minimizes the *negative likelihood ratio*, $\frac{1-y_i}{x_i}$.

- The positive and negative likelihood ratios measure the quality of experiments

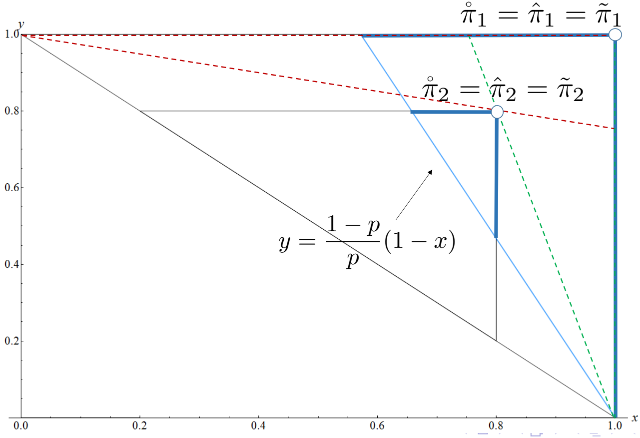
Two "Best" Experiments



Optimal Decision Rule 1

Corollary

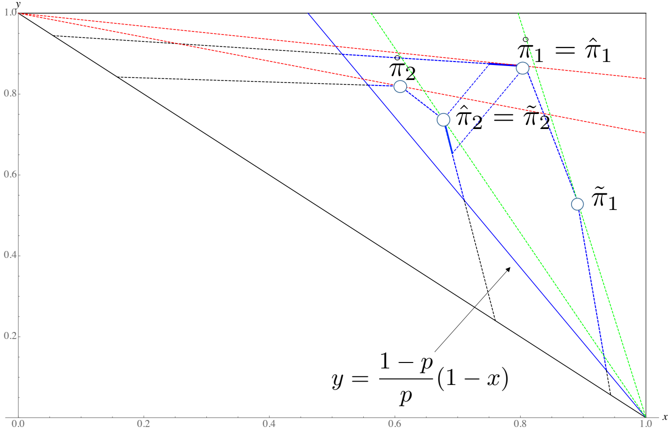
If $\hat{\pi}_i$ Blackwell-dominates every $\pi_i \in S_i$
(i.e., $\hat{y}_i \geq y_i$ and $\hat{x}_i \geq x_i$ for any $\pi_i = (x_i, y_i) \in S_i$,)
an optimal decision rule should have $\pi_i = \hat{\pi}_i$ for $i = 1, 2$,
i.e., $\{(\hat{\pi}_1, A_g^1, A_b^1), (\hat{\pi}_2, A_g^2, A_b^2)\}$.



Optimal Decision Rule 1

Proposition

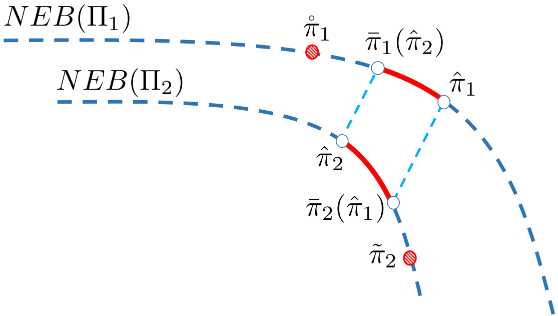
If (i) $\hat{\pi}_1 = \mathring{\pi}_1$ *and* (ii) $\hat{\pi}_2 = \tilde{\pi}_2$, an optimal decision rule should have $\pi_i = \hat{\pi}_i$ for $i = 1, 2$, i.e., $\{(\hat{\pi}_1, A_g^1, A_b^1), (\hat{\pi}_2, A_g^2, A_b^2)\}$.



Optimal Decision Rule 2

Proposition

If $\hat{\pi}_1$ is "far from" $\mathring{\pi}_1$ and $\hat{\pi}_2$ is "far from" $\tilde{\pi}_2$, an optimal decision rule must achieve the ex post optimal decisions: $(A^i_g, A^i_b) = (1, 0)$ for $i = 1, 2$.



Optimal Decision Rule 1

- $\{(\hat{\pi}_1, A_g^1, A_b^1), (\hat{\pi}_2, A_g^2, A_b^2)\}$

Proposition

(a) If θ_1 is "more likely" than θ_2,
$(A_g^1 = 1, A_b^1 = 0)$ and $(A_g^2 < 1, A_b^2 = 0)$ are optimal.

(b) If θ_1 is "less likely" than θ_2,
$(A_g^1 = 1, A_b^1 > 0)$ and $(A_g^2 = 1, A_b^2 = 0)$ are optimal.

- The principal sacrifices *ex post optimality* to incentivise the agent to conduct $\hat{\pi}_i$.
- Ex post optimality is sacrificed for the type which is "less likely" than the other.

Appendix

Optimal Decision Rule 2

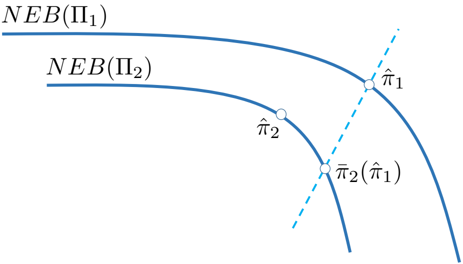
- $\{(A_g^1, A_b^1), (A_g^2, A_b^2)\}$ comes for free
- What is left is to find an appropriate (π_1, π_2)
- If $NEB(\Pi_i)$ is a well-defined function, it can be written as $N_i(x_i)$.

$$
\begin{aligned}
& \max_{x_1, x_2} \left\{ t \cdot (pN_1(x_1) - (1-p)x_1) + (1-t) \cdot (pN_2(x_2) - (1-p)x_2) \right\} \\
s.t. \quad & pN_1(x_1) + (1-p)x_1 = pN_2(x_2) + (1-p)x_2 \\
& \underline{x}_1 \leq x_1 \leq \bar{x}_1 \\
& \underline{x}_2 \leq x_2 \leq \bar{x}_2
\end{aligned}
$$

Concluding Remarks

- Study a simple principal-agent problem
  - Mechanism Designer vs. Information Designer
- Optimal decision rules
  - First-best outcome might be achievable
  - Two kinds of optimal decision rules
  1. Favorite experiments with distortions in ex post decision-making
     - Optimal when $\hat{\pi}_i$ meets some quality requirements
  2. Ex post optimal decisions with distortions in assigning experiments
     - Optimal when $\hat{\pi}_i$ fails to meet some quality requirements

Simplifying the Problem: Experiments

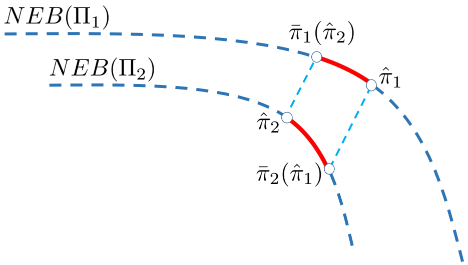


Definition of $\bar{\pi}_j(\pi_i)$

Given π_i on $NEB(\Pi_i)$, $\bar{\pi}_j(\pi_i)$ is an experiment on $NEB(\Pi_j)$ such that $\{(\pi_i, 1, 0), (\bar{\pi}_j(\pi_i), 1, 0)\}$ is incentive compatible (or $py_i + (1-p)(1-x_i) = p\bar{y}_j + (1-p)(1-\bar{x}_j)$.)

Return

Simplifying the Problem: Experiments



Proposition

An optimal M must have π_1 “between” $\bar{\pi}_1(\hat{\pi}_2)$ and $\hat{\pi}_1$ *and* π_2 “between” $\hat{\pi}_2$ and $\bar{\pi}_2(\hat{\pi}_1)$.

Return

Simplifying the Problem: Action-probability pairs

- Given $M = \{(\pi_1, A_g^1, A_b^1), (\pi_2, A_g^2, A_b^2)\}$, IC constraint for type θ_1

$$A_g^1(py_1 + (1-p)(1-x_1)) + A_b^1(1 - (py_1 + (1-p)(1-x_1)))$$
$$\geq A_g^2(py_2 + (1-p)(1-x_2)) + A_b^2(1 - (py_2 + (1-p)(1-x_2)))$$
$$\Rightarrow \quad M = \{(\pi_1, 1, A_b^1), (\pi_2, A_g^2, 0)\}$$

$$A_b^1 + (1 - A_b^1)(py_1 + (1-p)(1-x_1) \geq A_g^2(py_2 + (1-p)(1-x_2))$$

Lemma

An optimal M has

(a) a binding IC constraint and
(b) either $A_b^1 \geq 0$ or $A_g^2 \leq 1$.

$\Rightarrow M^1 = \{(\pi_1, 1, A_b^1 \geq 0), (\pi_2, 1, 0)\}$ or $M^2 = \{(\pi_1, 1, 0), (\pi_2, A_g^2 \leq 1, 0)\}$

Return

Simplifying the Problem: Action-probability pairs

Proposition

Given (π_1, π_2) such that $py_1 + (1-p)(1-x_1) \leq py_2 + (1-p)(1-x_2)$,

(a) if $\tau(\pi_1, \pi_2) > \frac{1-t}{t}$, $(A_g^1 = 1, A_b^1 = 0)$ and $(A_g^2(\pi_1, \pi_2) \leq 1, A_b^2 = 0)$ are optimal,

(b) if $\tau(\pi_1, \pi_2) < \frac{1-t}{t}$, $(A_g^1 = 1, A_b^1(\pi_1, \pi_2) \geq 0)$ and $(A_g^2 = 1, A_b^2 = 0)$ are optimal,

where

$$\tau(\pi_1, \pi_2) = \left(\frac{py_2+(1-p)(1-x_2)}{py_2-(1-p)(1-x_2)}\right) \cdot \left(\frac{1-2p+py_1-(1-p)(1-x_1)}{1-(py_1+(1-p)(1-x_1))}\right),$$

$$A_g^2(\pi_1, \pi_2) = \frac{py_1+(1-p)(1-x_1)}{py_2+(1-p)(1-x_2)}, \text{ and}$$

$$A_b^1(\pi_1, \pi_2) = \frac{(py_2+(1-p)(1-x_2))-(py_1+(1-p)(1-x_1))}{1-(py_1+(1-p)(1-x_1))}.$$

Return