Issues in Economic Systems and Institutions: Part V: Reputation

Parikshit Ghosh

Delhi School of Economics

April 3, 2023

An Entry Deterrence Game

- 2 periods, 2 players: Incumbent (I) and Rival (R).
- At $t = 1$, I chooses Fight (F) or Accommodate (A).
- Between dates 1 and 2, R can stay (S) or exit (E).
- If S, at $t = 2$, I again chooses F or A.
- I enjoys monopoly profits (π_m) after R exits, duopoly profits (π_d) after A and payoff from a price war (π_w) after F.

$$\pi_w < 0 < \pi_d < \pi_m$$

- R gets 0 after exit, π_d after A and π_w after F.

Incomplete Information

- Incumbent is either a rational type (prob $1-p$) or a crazy type (prob p).
- Rational type always plays best response, crazy type plays F in both periods.
- Two approaches: crazy type can be either "behavioral" or a rational player with a different payoff function.
- Rational I's strategies: $\{FF, FA, AA, AF\}$. Only FA or AA can be best response.
- R's beliefs: $\mu(\text{crazy}|A) = 0$ and $\mu(\text{crazy}|F) = \mu \in [0,1]$.
- R's strategies: $\{EE, ES, SE, SS\}$. First information set corresponds to F.

Separating Equilibrium

- Rational I's strategy: AA.
- Then $\mu = 1$ and R plays ES.
- I is playing best response if

$$2\pi_d \geq \pi_m + \pi_w \tag{1}$$

- Here fighting significantly enhances reputation but there is no reputation building (by rational incumbents).

Pooling Equilibrium

- Rational I's strategy: FA.
- I's strategy is best response only if R plays ES.
- Beliefs: $\mu = p$. Then I is playing best response after F if

$$p\pi_w + (1-p)\pi_d \leq 0 \tag{2}$$

- Here fighting preserves reputation, doesn't improve it.
- Rational incumbents pay the cost to preserve reputation.

Semi-Separating or Hybrid Equilibrium

- Suppose (1) and (2) are both violated.
- Alternative PBE: rational I plays F with prob α and R plays E with prob β after F.
- I's indifference implies:

$$
\begin{aligned}
2\pi_d &= \pi_w + \beta\pi_m + (1-\beta)\pi_d \\
\text{or,}\quad \beta &= \frac{\pi_d - \pi_w}{\pi_m - \pi_d}
\end{aligned}
$$

Semi-Separating or Hybrid Equilibrium

- R's indifference condition is

$$0 = \mu\pi_d + (1-\mu)\pi_w$$

$$\text{or,} \quad \mu = \frac{-\pi_w}{\pi_d - \pi_w}$$

- Using Bayes' Rule:

$$\mu = \frac{p}{p + (1-p)\alpha}$$

- Combining:

$$\alpha = \left(\frac{p}{1-p}\right)\left(\frac{1-\mu}{\mu}\right) = -\frac{\pi_d}{\pi_w}\left(\frac{p}{1-p}\right)$$

Remarks on Hybrid Equilibrium

- Here reputation is elastic both ways: improves after F and deteriorates after A.
- Rational incumbents may choose to invest in reputation.
- The higher the initial reputation (p), the more likely the incumbent is to defend it (α).
- Higher the initial reputation, the less it improves after F.
- "Soft bigotry of low expectations." Initial conditions matter.

- Long horizon: Consider the T stage version (one long run incumbent against a series of short run rivals). If T is large enough, even the rational incumbent fights with probability 1 in all but the last few periods.

Reputational Herding (Scharfstein and Stein 1990)

- Variation on the theme of informational cascades.
- Agents have no direct payoff from the decision—they are investing other people's money.
- Agents want to enhance their reputation for expertise.
- A critical assumption:
  - signals of good experts are correlated (great minds think alike)
  - signals of bad experts are uncorrelated (fools often differ)
- Reputation is enhanced by: (a) taking the right decision (b) agreeing with other experts.
- (b) may be so strong that all experts mimic the choices of their predecessors.

The Model

- Two fund managers: A and B.
- Each manager chooses to invest (I) or not (N).
- Return on investment is either high ($x_H > 0$) or low ($x_L < 0$), with equal likelihood.
- Manager A chooses first, then manager B.
- Each manager receives a private signal, which is either good (s_G) or bad (s_B).
- The second manager can observe the first manager's action but not his signal.

Information

- Each manager is either smart (prob θ) or dumb (prob $1-\theta$).
- Managers do not know the quality of their own signals.
- Smart manager's signal (informative but noisy):

	s_G	s_B
x_H	p	$1-p$
x_L	$1-p$	p

- Dumb manager's signal (pure noise):

	s_G	s_B
x_H	$\frac{1}{2}$	$\frac{1}{2}$
x_L	$\frac{1}{2}$	$\frac{1}{2}$

- Smart signals perfectly correlated, dumb signals independent.

Reputation

- $\Pr[s_G|\text{smart}] = \Pr[s_G|\text{dumb}]$, hence each signal in itself conveys no information about expertise.
- Two signals together convey some information about expertise (matched signals good news).
- Market observes each manager's action but not signal.
- Also learns the state of the world (x_H or x_L) eventually.
- Revises the probability that an expert is smart to some $\widehat{\theta}$.
- Experts are interested in maximizing expected value of reputation ($\widehat{\theta}$) because their future salaries are linked to it.

Benchmark 1A: Single Investor, One Signal

- Suppose there is a single investor who invests his own money, i.e., cares about returns, not reputation.
- Conditional probabilities after each signal:

$$\mu_G = \Pr[x_H | s_G] = \theta p + (1-\theta).\frac{1}{2}$$
$$\mu_B = \Pr[x_H | s_B] = \theta(1-p) + (1-\theta).\frac{1}{2}$$

- Assume:

$$\mu_B x_H + (1-\mu_B)\, x_L < 0 < \mu_G x_H + (1-\mu_G)\, x_L$$

- Optimal decision is dependent on the signal.

Benchmark 1B: Single Investor, Two Signals

- Suppose the investor knows two signals.
- Given previous assumption, the optimal decision rule (which maximizes returns) is:
  - if (s_G, s_G), choose I.
  - if (s_B, s_B), choose N.
  - if (s_G, s_B) or (s_B, s_G), depends.
- If the signals are opposite they wash out, i.e., $\Pr[x_H | s_G, s_B] = \frac{1}{2}$.
- Then I if $x_H + x_L > 0$, and N if $x_H + x_L < 0$.
- It is not optimal to simply follow the first signal.

Benchmark 2: Single Manager

- Suppose there is a single manager who cares about reputation, not returns.
- If the expert is expected to invest under s_G and not invest under s_B, will he behave accordingly (i.e. reveal his signal)?
- The manager's reputation goes up when he is correct and goes down when he is wrong:

$$\widehat{\theta}\left(s_G, x_H\right) = \widehat{\theta}\left(s_B, x_L\right) > \theta > \widehat{\theta}\left(s_G, x_L\right) = \widehat{\theta}\left(s_B, x_H\right)$$

- The manager thinks the state is more likely to be what his signal indicates:

$$\Pr[x_H | s_G] = \Pr[x_L | s_B] > \frac{1}{2} > \Pr[x_H | s_B] = \Pr[x_L | s_G]$$

- The expected reputation is greater if he reveals his signal than if he misreports it.

Equilibrium

Theorem

In equilibrium, manager A always invests if he gets s_G and does not invest if he gets s_B.

- We will show that manager B always mimics manager A, regardless of his own signal.
- Then, A's reputation is affected only by his own actions.
- A single manager will always reveal his signal (previous slide).
- There is also a "perverse" equilibrium where he reveals his signal by taking the wrong action. Rule it out (suppose he has a small stake in returns).

Equilibrium

Theorem

There is no equilibrium in which manager B always reveals his signal.

- Suppose there is a revealing equilibrium.
- Suppose (w.l.o.g) A's signal is revealed as s_B, but manager B's signal is s_G.
- We will show that manager B will misreport his signal as s_B.

Outline of Proof

- Two conflicting signals cancel each other out (nobody knows who is the dumb one!)

$$\Pr[x_H | s_B, s_G] = \frac{1}{2}$$

- Incentive constraint for truthfully reporting s_G:

$$\frac{1}{2}\widehat{\theta}\left(s_B, s_G, x_H\right) + \frac{1}{2}\widehat{\theta}\left(s_B, s_G, x_L\right)$$

$$\geq \quad \frac{1}{2}\widehat{\theta}\left(s_B, s_B, x_H\right) + \frac{1}{2}\widehat{\theta}\left(s_B, s_B, x_L\right)$$

- In words:

expected reputation (truth-telling) $\geq$ expected reputation (lying)

Outline of Proof

- However

$$\underbrace{\widehat{\theta}\left(s_B, s_G, x_H\right)}_{\text{agrees with state alone}} < \underbrace{\widehat{\theta}\left(s_B, s_B, x_L\right)}_{\text{agrees with both}}$$

$$\underbrace{\widehat{\theta}\left(s_B, s_G, x_L\right)}_{\text{disagrees with both}} < \underbrace{\widehat{\theta}\left(s_B, s_B, x_H\right)}_{\text{disagrees with state alone}}$$

- When signals conflict, either prediction has the same chance of being correct.
- Ceteris paribus, agreeing with other expert will increase the likelihood of being perceived smart.
- It is better to be wrong with others than to be wrong alone!

Equilibrium

Theorem

There is an equilibrium in which manager B always does the same thing as manager A, regardless of his own signal..

- "Reasonable" off-the-equilibrium-path beliefs: if B disagrees with A, then his signal is s_G if he chose I and s_B if he chose N.
- Incentive constraint:

$$\theta \geq \frac{1}{2}\widehat{\theta}\left(s_B, s_G, x_H\right) + \frac{1}{2}\widehat{\theta}\left(s_B, s_G, x_L\right)$$

$$\text{or,} \quad \theta - \widehat{\theta}\left(s_B, s_G, x_L\right) \leq \widehat{\theta}\left(s_B, s_G, x_H\right) - \theta$$

- Intuition: reputation loss from being wrong alone is greater than reputation gain from being right alone.

Political Correctness (Morris 2001)

- Listeners are often unsure about the bias of speakers (are they sexist/racist/casteist)?
- What is said reveals something about both the speaker's information as well as his motives.
- Message affects reputation and vice versa.
- Political correctness: even unbiased speakers lie! They lie in a direction opposite to the suspected bias.
- In the extreme, political correctness leads to babbling: everyone says the "safe" thing.
- PC may be socially inefficient due to informational loss. Important policy issues are not discussed frankly.
- Possibility of multiple equilibria: different speech cultures.

The Model

- Two periods: $t = 1, 2$. State-of-the-world at date t: $\omega_t \in \{0, 1\}$. Equi-probable and time-independent.
- Two players: decision maker (D) and advisor (A).
- D chooses action $a_t \in \mathbf{R}$ each period. D's payoff:

$$U_D = -x_1(a_1 - \omega_1)^2 - x_2(a_2 - \omega_2)^2$$

- A can be good (prob λ_1) or bad (prob $1 - \lambda_1$).
- Good advisors have the same preference as D, bad advisors always want higher action:

$$U_A^b = y_1 a_1 + y_2 a_2$$

- $\frac{x_2}{x_1}$ and $\frac{y_2}{y_1}$ represent relative importance of the future.

The Model

- The advisor gets an independent noisy signal of the state each period: $s_t \in \{0, 1\}$.
- Accuracy of the signal is $\gamma \in \left(\frac{1}{2}, 1\right)$:

	$s = 0$	$s = 1$
$\omega = 0$	γ	$1 - \gamma$
$\omega = 1$	$1 - \gamma$	γ

- Each period, A sends message $m_t \in \{0, 1\}$.
- After date 1, D learns ω_1 and updates his belief to λ_2.
- $\lambda_2 (\lambda_1, m_1, \omega_1)$ is the advisor's new reputation.

Full Information Benchmark

- If A was known to be good, messages would be truthful and credible.
- Optimal actions equal to expected value of the state:

$$
\begin{aligned}
a^*(0) &= \Pr[\omega = 1 | s = 0] = 1 - \gamma \\
a^*(1) &= \Pr[\omega = 1 | s = 1] = \gamma
\end{aligned}
$$

- Good advisor wants exactly these actions.
- Bad advisor always wants $a = 1$, regardless of state.
- Without reputational concerns, good advisor reports true value of signal. Bad advisor lies and reports 1 even if signal is 0.

Last Period

- Message strategies

	$s_2 = 0$	$s_2 = 1$
Good	0	1
Bad	1	1

- Optimal actions, $a_2\left(m_2; \lambda_2\right)$:

$$\begin{aligned} a_2(0; \lambda_2) &= 1 - \gamma \\ a_2(1; \lambda_2) &= \Pr\left[\omega_2 = 1 \middle| m_2 = 1\right] \\ &= \frac{1 - \lambda_2(1-\gamma)}{2 - \lambda_2} \end{aligned}$$

- $a_2(1; \lambda_2) \uparrow$ as $\lambda_2 \uparrow$ and $\rightarrow \gamma$ as $\lambda_2 \rightarrow 1$.

Value of Reputation

- Let $v_G(\lambda_2)$ and $v_B(\lambda_2)$ denote expected payoff to good and bad advisor in the last period.
- The bad advisor always reports $m_2 = 0$:

$$v_B(\lambda_2) = y_2 a_2(1; \lambda_2) \uparrow \text{ in } \lambda_2$$

- The good advisor always reports truthfully:

$$v_G(\lambda_2) = -x_2 . \frac{1}{2}\left[\mathbf{E}\left((a_2 - \omega_2)^2 \,|s_2 = 0\right) + \mathbf{E}\left((a_2 - \omega_2)^2 \,|s_2 = 1\right.\right.$$

- First term is minimized at $a_2 = 1 - \gamma$, second term at $a_2 = \gamma$.
- $a_2(1; \lambda_2) \uparrow$ towards γ as $\lambda_2 \uparrow$. Both types value reputation.

First Period

- Assume equilibrium with no political correctness, i.e., good A reports truthfully.
- Bad A must falsely report the truth sometimes. Suppose not:

	$s_1 = 0$	$s_1 = 1$
Good	0	1
Bad	0	1

- Then $\lambda_2(\lambda_1, m_1, \omega_1) = \lambda_1$ for any m_1 and ω_1. Message does not affect reputation.
- Also, message is believed completely: $a_1(0) = 1 - \gamma < \gamma = a_1(1)$. Bad advisor will lie and always report 1.

First Period

- Assume good advisor always reports truthfully; bad advisor sometimes lies.

	$s_1 = 0$	$s_1 = 1$
Good	0	1
Bad	0 (prob $1 - v$) 1 (prob v)	1

- Then $m_1 = 0 \Rightarrow s_1 = 0$, but $m_1 = 1$ does not $\Rightarrow s_1 = 1$.
- Equilibrium actions $a_1(m_1)$:

$$a_1(0) = 1 - \gamma$$

$$a_1(1) \in \left(\frac{1}{2}, \gamma\right)$$

Reputation

- Reputation is enhanced if message is
  - (i) factually correct
  - (ii) politically correct.
- Ceteris paribus, initial reputation positively affects final reputation.
- Comparison for $\lambda_2(\lambda_1, m_1, \omega_1)$:

$$\lambda_2(\lambda_1, 0, 1) = \lambda_2(\lambda_1, 0, 0) > \lambda_1 > \lambda_2(\lambda_1, 1, 1) > \lambda_2(\lambda_1, 1, 0)$$

- Exact expressions involve υ and can be calculated using Bayes' Rule.

Reputation Rankings: Intuition

- Politically correct message ($m_1 = 0$) enhances reputation regardless of factual correctness. Reason:
  - the good type sends $m_1 = 0$ more often than the bad type.
  - $m_1 = 0$, $\omega_1 = 1$ is always an honest mistake.
- Politically incorrect message ($m_1 = 1$) harms reputation even when factually correct!
  - the bad type sends $m_1 = 1$ more often than the good type.
  - $m_1 = 1$, $\omega_1 = 1$ is sometimes dishonest yet accidental accuracy.
- $m_1 = 1$ causes further damage to reputation when factually incorrect.
  - $m_1 = 1$, $\omega_1 = 0$ is sometimes a deliberate mistake.

First Period Equilibrium

- **Condition 1:** when $s_1 = 0$, the bad advisor must be indifferent between truth-telling and lying:

$$y_1 a_1(0) + \mathbf{E}\left[v_B(\lambda_2)\,|0,0\right] = y_1 a_1(1) + \mathbf{E}\left[v_B(\lambda_2)\,|1,0\right]$$

$$\underbrace{y_1\left[a_1(1) - a_1(0)\right]}_{\text{gain from lying}} = \underbrace{\mathbf{E}\left[v_B(\lambda_2)\,|0,0\right] - \mathbf{E}\left[v_B(\lambda_2)\,|1,0\right]}_{\text{expected reputational loss}}$$

- When $s_1 = 1$, the bad advisor strictly prefers to tell the truth (implication of above).
- When $s_1 = 0$, the good advisor always wants to tell the truth (better outcome as well as better reputation).
- **Condition 2:** the good advisor must prefer to tell the truth when $s_1 = 1$ (algebra ommitted).

Informative Equilibria: General Properties

- Good advisor sends $m_1 = 0$ whenever $s_1 = 0$. He announces $m_1 = 1$ with (weakly) positive probability if $s_1 = 1$.
  - When $s_1 = 0$, both current and reputational payoffs are higher for $m_1 = 0$.
- Bad advisor sends $m_1 = 1$ more often than the good advisor.
  - Otherwise there would be no reputational cost to sending $m_1 = 1$.
- There is a strict reputational incentive to be politically correct:

$$\lambda_2(\lambda_1, 0, 1) = \lambda_2(\lambda_1, 0, 0) > \lambda_1 > \lambda_2(\lambda_1, 1, 1) > \lambda_2(\lambda_1, 1, 0)$$

  - Reasons as before.

Extreme Political Correctness

- If $\frac{x_2}{x_1}$ is high enough, the only equilibrium is babbling.
- One way to depict the strategies:

	$s_1 = 0$	$s_1 = 1$
Good	0	0
Bad	0	0

- Both type of advisors say the "safe" thing so as not to damage their reputation and influence in the future.
- Reason: the bad advisor's indifference condition pins down reputational gain from $m_1 = 0$. Good advisor will want to capture this if x_2 is high enough.
- Under PC, **nothing** is learnt about the state or the speaker!

Welfare

- Benchmark: a model where D in the 2nd period does not know what happened in the 1st period (no reputational incentives).
- Without reputation, strategies are in both periods are as in last period of the game with reputation.
- Reputation creates 3 effects:
  1. *Discipline effect*: bad advisor tells the truth more often (announce $m_1 = 0$ when $s_1 = 0$). $(+)$
  2. *Sorting effect*: decision maker learns something about the speaker and his trustworthiness. $(+)$
  3. *Political correctness effect*: even the good advisor starts lying sometimes (send $m_1 = 0$ when $s_1 = 1$). (-)
- When we have an extreme PC (babbling) equilibrium, 3 dominates $1 + 2$.

What's In A Name?

- Bertrand and Mullainathan (*AER*, 2004): RCT for testing discrimination in job applications.
- Fictitious CVs randomly matched with white (Emily, Greg) and black (Lakisha, Jamal) sounding names.
- 5000 resumes submitted in response to 13000 employment ads in Boston and Chicago.
- Call back rates: whites = 9.65%, blacks = 6.45%. Whites have 50% higher call back.
- As resume quality improves, response rate goes up faster for whites than blacks.
- Blind auditions improve the chances of women violinists significantly (Goldin and Rouse, *AER* 2000).

Other Examples of “Discrimination”

- Insurance premiums: young drivers (under 25) and older health insurees (over 65) face higher rates.
- Airport screening: Middle Eastern males more likely to be terrorists than Swiss nuns.
- Residential choice and segregation: racial composition of a neighbourhood is a predictor of crime. Avoid the Bronx.
- Credit: 97% of Grameen bank loans are given to women.
- Racial profiling: Search rates from Knowles, Persico and Todd's (*JPE* 2001) data on 1,590 stop-and-search operations by Maryland police, 1995 - 99:
  - Blacks = 63%, whites = 29%.
  - Men = 93%, women = 7%.

Three Notions of Discrimination

- **Incidental discrimination:** groups are not treated differently, but there is differential impact because groups differ statistically in behaviour (prison population disproportionately male because lawbreakers are disproportionately male).
- **Statistical discrimination:** groups are treated differently, but only insofar as group affiliation is a statistical predictor of behaviour (young people drive rash, on average).
- **Taste based discrimination:** groups are treated differently because doing so for its own sake generates utility (racial or caste-based segregation). Prejudice, pure and simple.

How do we empirically separate these strands?

Discrimination in Monitoring

- Crime as a rational choice: criminals commit an illegal act after weighing benefits against expected punishment cost (Becker, 1968).
- Two races, $r = A, W$. Observationally distinct to the police.
- Observable non-racial characteristic c follows distribution $n_r(c)$, with totals

$$N_r = \int n_r(c)dc$$

- Distribution of x, i.e. legal income opportunities: $F_r(x; c)$.
- Benefit of carrying drugs $= B$, penalty if caught $= P$.
- Police have enough resources to search only S people.

The Game and Equilibrium

- Individual decision: carry drugs or enter a legal profession?
- Police simultaneously choose $\sigma_r(c)$ —what proportion of group (r, c) to search, subject to the budget constraint

$$\sum_{r=A}^{W} \int n_r(c)\sigma_r(c)dc = S$$

- Police objective function:

$$U = \sum_{r=A}^{W} \int n_r(c)\sigma_r(c)\left[\theta_r(c) + u_r\right] dc$$

where u_r is the intrinsic utility/disutility of searching race r, and $\theta_r(c)$ is the "hit rate" among group (r, c).

The Testable Implication

- Interior equilibrium: police must be indifferent across groups, i.e. $\theta_r(c) + u_r$ is a constant.
- Expected payoff from drug peddling:

$$q(\sigma_r(c)) = (1 - \sigma_r(c))B - \sigma_r(c)P$$

- Those whose legal incomes are less, choose to commit crime:

$$\theta_r(c) = F_r(q(\sigma_r(c)))$$

- Equality is preserved after integrating over non-racial characteristics. Let $\theta_r = \int \theta_r(c)n_r(c)dc$. Then

$$\theta_A + u_A = \theta_W + u_W$$

Testable Implication

- If police are unprejudiced ($u_A = u_W$), **hit rates** will be equalized across groups, even if **search rates** are very different.
- If police are prejudiced against blacks ($u_A > u_W$), equilibrium hit rates will be **lower** among blacks.
- The general idea: compensating differentials.
- Onerous jobs pay more, attractive cities/neighbourhoods have higher rents.
- Does not require the econometrician to know or control for all the variables that police condition their decisions on.

Highway Search Data: Hit Rates

- Test of equality of means (Pearson's chi-squared):

$$\sum_{r=1}^{R} \frac{(\widehat{p}_r - \widehat{p})^2}{\widehat{p}_r} \sim \chi^2(R-1)$$

- Observed hit rates across racial and gender groups

	Black	White	Hispanic	All Races
Both sexes	.34	.32	.11*	.30
Male	.34	.33	.11	.32
Female	.44*	.22*	-	.36

- Null hypothesis of equality is not rejected when only whites and blacks are used. Rejected when Hispanics are added.
- Evidence of taste based discrimination against Hispanics and white females, but not males or blacks.

Efficiency

Theorem

If $u_A = u_W$ and $F_A = F_W = F$ is concave, the equilibrium allocation maximizes rather than minimizes aggregrate crime rate.

- Let $N_A = N_W = N$ and $\sigma = \frac{S}{N}$. Budget constraint:

$$\frac{1}{2}[\sigma_A + \sigma_W] = \sigma$$

- Aggregrate crime rate:

$$F_A(q(\sigma_A)) + F_W(q(2\sigma - \sigma_A))$$

- First-order condition holds at $f_A = f_W$ ($q'(.)$ is constant)):

$$f_A(q(\sigma_A)) = f_W(q(2\sigma - \sigma_A))$$

- Since the objective function is concave, this is a maximum!

Equilibrium, Fairness and Efficiency

- In general, they are described by three different conditions:
  - Equilibrium: $F_A = F_W$, i.e. crime rates are equal.
  - Fairness: $\sigma_A = \sigma_W$, i.e. search rates are equal.
  - Efficiency: $f_A = f_W$, i.e. response elasticities are equal.
- Mandatory quotas (e.g. $\sigma_A = \sigma_W$) to improve fairness may improve efficiency or worsen it.
- Intuition: there is a conflict between two different objectives:
  - discouraging crime (deterrence).
  - catching as many criminals as possible (retribution).
- Dynamic inconsistency and problem of commitment.

An Example

- Extreme case: $F_W(q(2\sigma)) = 1$, i.e., Ws are completely unresponsive to higher scrutiny (zero response elasticity).
- Efficient allocation: $\sigma_W = 0$ and $\sigma_A = 2\sigma$.
- Equilibrium allocation is fairer but less efficient than this benchmark.

Statistical Discrimination

- Skills a function of both innate ability and effort.
- Noisy measurement of skill $\Rightarrow$ priors will affect posteriors.
- Positive feedback loop between employer perceptions and worker effort:
  - if employers hold optmistic beliefs, workers may have a strong incentive to become skilled.
  - if employers hold pessimistic beliefs, workers may have a weak incentive to become skilled.
- Multiple equilibria are possible for some parameters.
- Different groups (e.g. blacks and whites) may get locked into different equilibria.
- Affirmative may improve or worsen stereotypes (beliefs about skill distribution in target population).

The Coate-Loury model

- Two races: W (proportion λ) and B.
- Two kinds of tasks: unskilled (0) and skilled (1).
- Two worker types: qualified (q) and unqualified (u).
- Every worker (either race) could become qualified at a cost c which is private information. Within each group, $c \sim U[0,1]$.
- Payoffs to employers and workers:

	Skilled (1)	Unskilled (0)
Qualified (q)	x_q, w	$0, 0$
Unqualified (u)	$-x_u, w$	$0, 0$

- Noisy test of qualification:

	Pass	Unclear	Fail
Qualified (q)	$1 - p_q$	p_q	0
Unqualified (u)	0	p_u	$1 - p_u$

Employer and Worker Best Response

- Employers: Pass $\to$ 1, Fail $\to$ 0, Unclear $\to$?
- Let prior $= \pi$. Then posterior

$$\sigma = \Pr(q|\text{Unclear}) = \frac{\pi p_q}{\pi p_q + (1-\pi)p_u}$$

- Assigning skilled task after unclear test result is optimum iff

$$\sigma x_q + (1-\sigma)x_u \geq 0$$

$$\pi \geq \frac{p_u x_u}{p_u x_u + p_q x_q} = \widehat{\pi}$$

- Workers invest in qualification iff cost below a cutoff:

$$\begin{aligned} \phi(\pi) &= \bar{c} = (1-p_u)w \text{ if } \pi \geq \widehat{\pi} \\ &= (1-p_q)w \text{ if } \pi < \widehat{\pi} \end{aligned}$$

Multiple Equilibria

- **Equilibrium with liberal beliefs:** suppose $\pi \geq \widehat{\pi}$ in equilibrium $\Rightarrow$ unclear test result leads to skilled task. Condition:
$$\pi_l = \phi(\pi_l) = (1 - p_u)w \geq \widehat{\pi}$$
- **Equilibrium with conservative beliefs:** suppose $\pi < \widehat{\pi}$ in equilibrium $\Rightarrow$ unclear test result leads to unskilled task. Condition:
$$\pi_c = \phi(\pi_c) = (1 - p_q)w < \widehat{\pi}$$
- Multiple equilibria exist if
$$\pi_c < \widehat{\pi} \leq \pi_l$$
$$\text{or} \quad (1 - p_q)w < \frac{p_u x_u}{p_u x_u + p_q x_q} \leq (1 - p_u)w$$

Numerical Example

- Stability: suppose employer beliefs evolve the following way:

$$\pi^{t+1} = \phi(\pi^t)$$

  The both equilibria are locally stable.

- Payoffs to employers and workers:

	Skilled	Unskilled
Qualified	$1,1$	$0,0$
Unqualified	$-1,1$	$0,0$

- Noisy test of qualification:

	Pass	Unclear	Fail
Qualified	$\frac{1}{4}$	$\frac{3}{4}$	0
Unqualified	0	$\frac{1}{2}$	$\frac{1}{2}$

Employer's Hiring Strategy

- Employers: pass $\longrightarrow$ manager, fail $\longrightarrow$ clerk, unclear $\longrightarrow$?
- Let fraction of qualified workers $= \pi$. After unclear test result

$$\sigma = \Pr(s|\text{Unclear}) = \frac{\pi.\frac{3}{4}}{\pi.\frac{3}{4} + (1-\pi).\frac{1}{2}} = \frac{3\pi}{2+\pi}$$

- Assigning as manager after unclear test result is optimum iff

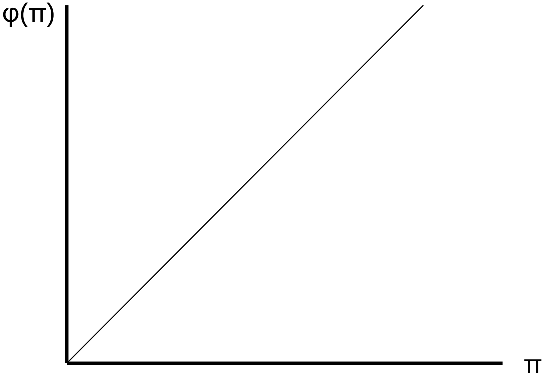
$$\sigma.1 + (1-\sigma)(-1) \geq 0 \Rightarrow \sigma \geq \frac{1}{2}$$

$$\text{or} \quad \pi \geq \frac{2}{5}$$

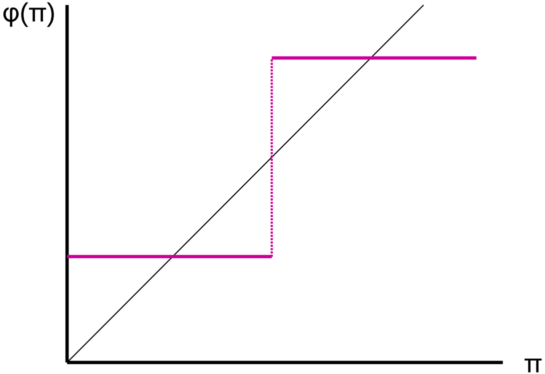
Worker's Investment Strategy

- Let $\phi(\pi)$ = fraction of workers who **actually** acquire skills when employers **think** this fraction is π.
- **Optmistic beliefs:** $\pi \geq \frac{2}{5}$ (unclear $\longrightarrow$ manager).
  - Hiring probability if qualified = 1.
  - Hiring probability if unqualified = $\frac{1}{2}$.
  - Expected gain from skill = cost threshold for investment = fraction of qualified workers = $\phi(\pi) = \frac{1}{2}$.
- **Pessimistic beliefs:** $\pi \geq \frac{2}{5}$ (unclear $\longrightarrow$ clerk).
  - Hiring probability if qualified = $\frac{1}{4}$.
  - Hiring probability if unqualified = 0.
  - Expected gain from skill = cost threshold for investment = fraction of qualified workers = $\phi(\pi) = \frac{1}{4}$.
- In equilibrium, employers' beliefs must be fulfilled: $\phi(\pi) = \pi$.

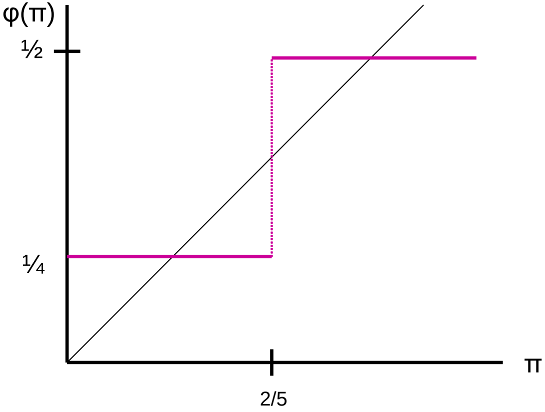
Multiple Equilibria



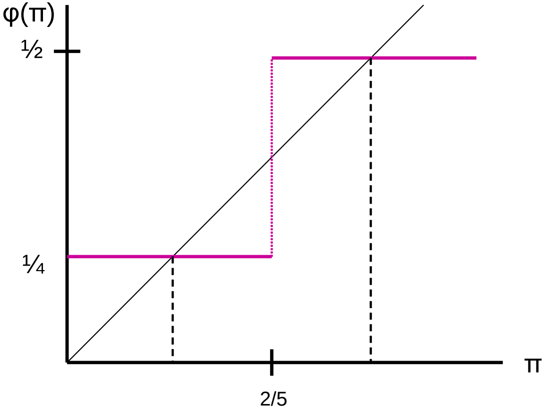
Multiple Equilibria



Multiple Equilibria



Multiple Equilibria



Affirmative Action, Incentives and Stereotype

- Does affirmative action destroy incentives, investment in skills, etc.?
- Does affirmative action worsen stereotypes of beneficiaries?
- Theoretically, it can go either way:
  - since AA makes entry easier, fewer members may invest.
  - under statistical discrimination, if entry is too hard, it may discourage investment.
- The effect can only be determined empirically.
- Our model illustrates potential positive effect on stereotypes in two senses:
  - the bad equilibrium improves locally (temporary effect).
  - the bad equilibrium is destroyed altogether (permanent effect).

Affirmative Action, Incentives and Stereotype

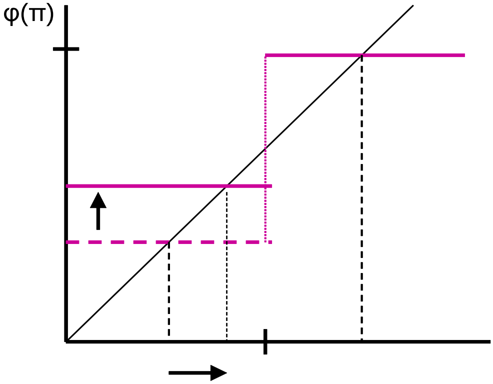
- Suppose A's are in bad equilibrium, W's are in good equilibrium.
- If quotas are introduced, employers may be forced to hire a fraction α of A's with unclear test results.
- For given α, a A worker invests if

$$\begin{aligned} \underbrace{\left(1 - p_q + \alpha p_q\right) w - c}_{} &\geq \underbrace{\alpha p_u w}_{} \\ \text{qualified payoff} &\geq \text{unqualified payoff} \end{aligned}$$

$$\text{or,} \quad c \leq \left[1 - p_q + \alpha\left(p_q - p_u\right)\right] w$$

- If $p_q > p_u$ (condition for multiple equilibria), the cost threshold for investment increases with α.
- Incentives and stereotype improve in the bad equilibrium.

Case 1: Local Improvement, Temporary Effect



Case 2: Global Improvement, Permanent Effect