Some Unintended Negative Consequences of Affirmative Action Policies: Theory and Empirics

Discussion Group on Affirmative Action Policies

Institute for Advanced Study in Toulouse
Glenn C. Loury, October 9, 2018

Two aspects of a racial preference policy may work against the interests of its beneficiaries:

Two aspects of a racial preference policy may work against the interests of its beneficiaries:

(1) Incentive and Reputation Issues:

(Theory) By relaxing selection criteria one alters an applicant's incentives to acquire skills and the inferences that observers are inclined to make about their abilities. (Coate/Loury 1993)

Two aspects of a racial preference policy may work against the interests of its beneficiaries:

(1) Incentive and Reputation Issues:

(Theory) By relaxing selection criteria one alters an applicant's incentives to acquire skills and the inferences that observers are inclined to make about their abilities. (Coate/Loury 1993)

(2) Sorting and Matching Issues:

(Empirics) When complementarities exist between individual and organizational traits (e.g., between a person's abilities and those of his/her classmates), AA may distort the allocation of persons across organizations and in that way disadvantage its beneficiaries'. (Arciadicono and Lovenheim, 2016)

Basic Set-up of Coate-Loury's Statistical Discrimination Model

Basic Set-up of Coate-Loury's Statistical Discrimination Model

- Employers and workers interact under incomplete information about a worker's qualifications. All workers belong to one of two visible groups, either B or W.

Basic Set-up of Coate-Loury's Statistical Discrimination Model

- Employers and workers interact under incomplete information about a worker's qualifications. All workers belong to one of two visible groups, either B or W.

- Let $\lambda \epsilon (0,1)$ denote the fraction of workers belonging to (advantaged) group W.

Basic Set-up of Coate-Loury's Statistical Discrimination Model

- Employers and workers interact under incomplete information about a worker's qualifications. All workers belong to one of two visible groups, either B or W.
- Let $\lambda \epsilon (0,1)$ denote the fraction of workers belonging to (advantaged) group W.
- Workers can choose to become qualified, at the cost = c ($I=1$) or not ($I=0$). Costs are distributed in the worker population according to a CDF, $G(c)$, in the same way for each group. Employers observe a noisy signal of a worker's qualification, θ.

Basic Set-up of Coate-Loury's Statistical Discrimination Model

- Employers and workers interact under incomplete information about a worker's qualifications. All workers belong to one of two visible groups, either B or W.

- Let $\lambda \epsilon (0,1)$ denote the fraction of workers belonging to (advantaged) group W.

- Workers can choose to become qualified, at the cost = c (I=1) or not (I=0). Costs are distributed in the worker population according to a CDF, G(c), in the same way for each group. Employers observe a noisy signal of a worker's qualification, θ.

- π_i denotes the fraction of group i=B,W workers employers believe to be qualified. (In an equilibrium, employers' beliefs about each group of workers are correct.)

Basic Set-up of Coate-Loury's Statistical Discrimination Model

- Employers and workers interact under incomplete information about a worker's qualifications. All workers belong to one of two visible groups, either B or W.

- Let $\lambda \epsilon (0,1)$ denote the fraction of workers belonging to (advantaged) group W.

- Workers can choose to become qualified, at the cost = c (I=1) or not (I=0). Costs are distributed in the worker population according to a CDF, G(c), in the same way for each group. Employers observe a noisy signal of a worker's qualification, θ.

- π_i denotes the fraction of group i=B,W workers employers believe to be qualified. (In an equilibrium, employers' beliefs about each group of workers are correct.)

- Employers assign a worker to some desirable position (A=1), or not (A=0). Doing so generates the net employer gain of x_q if that worker has become qualified (I=1), and the net loss of x_u if that worker is not qualified (I=0). Not assigning the worker to the desirable position (A=0) generates a net payoff of zero for the employer.

Basic Set-up of Coate-Loury's Statistical Discrimination Model

- Employers and workers interact under incomplete information about a worker's qualifications. All workers belong to one of two visible groups, either B or W.

- Let $\lambda \epsilon (0,1)$ denote the fraction of workers belonging to (advantaged) group W.

- Workers can choose to become qualified, at the cost = c (I=1) or not (I=0). Costs are distributed in the worker population according to a CDF, G(c), in the same way for each group. Employers observe a noisy signal of a worker's qualification, θ.

- π_i denotes the fraction of group i=B,W workers employers believe to be qualified. (In an equilibrium, employers' beliefs about each group of workers are correct.)

- Employers assign a worker to some desirable position (A=1), or not (A=0). Doing so generates the net employer gain of x_q if that worker has become qualified (I=1), and the net loss of x_u if that worker is not qualified (I=0). Not assigning the worker to the desirable position (A=0) generates a net payoff of zero for the employer.

- Workers receive the gross payoff w if assigned to the desirable position (A=1), and they receive the gross payoff of zero if not so assigned. (A=0)

Schematic representation of employer/worker interactions:

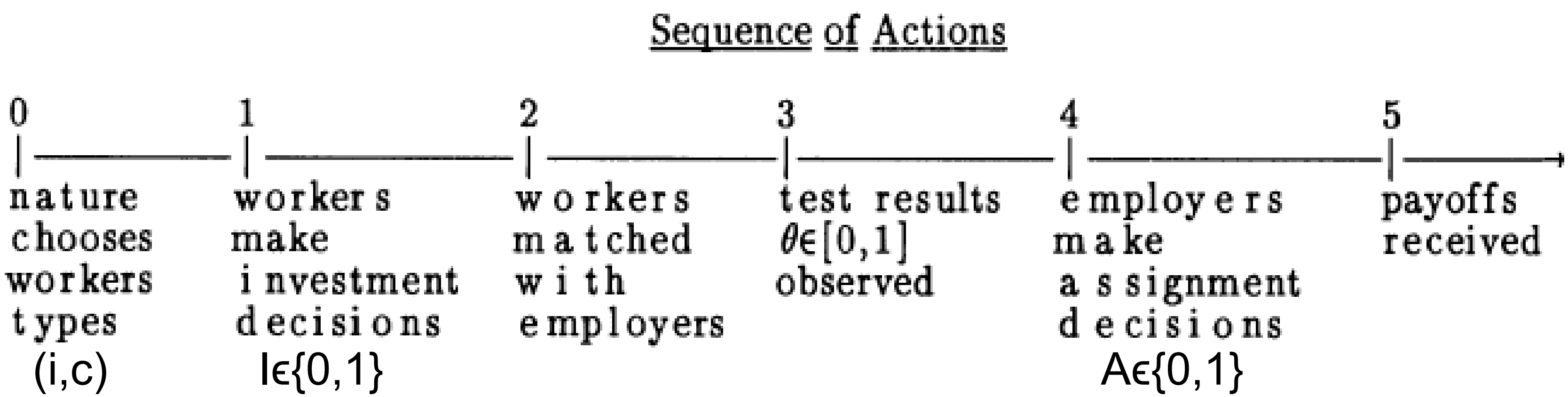


FIGURE 1. SEQUENCE OF ACTIONS

Schematic representation of employer/worker interactions:

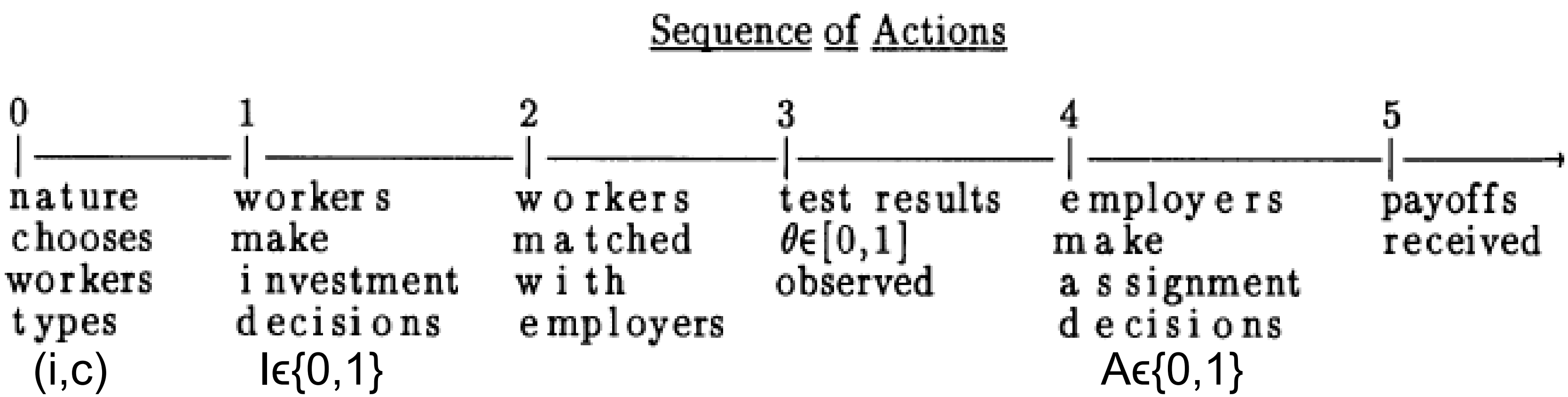


FIGURE 1. SEQUENCE OF ACTIONS

- Let us now consider employers' assignment behavior in the context of this model.

Employers' Assignment Behavior:

- Employers cannot directly observe a worker's qualification before they make an assignment. Instead they observe the noisy signal, $\theta \epsilon [0,1]$.

Employers' Assignment Behavior:

- Employers cannot directly observe a worker's qualification before they make an assignment. Instead they observe the noisy signal, $\theta \epsilon [0,1]$.

- $F_q(\theta)$ and $F_u(\theta)$ are signal CDF's conditional on worker qualification. The respective PDF's are $f_q(\theta)$ and $f_u(\theta)$. Assume that the likelihood ratio $\varphi(\theta)=[f_u(\theta)/f_q(\theta)]$ is a strictly decreasing function of θ. (MLRP)

Employers' Assignment Behavior:

- Employers cannot directly observe a worker's qualification before they make an assignment. Instead they observe the noisy signal, $\theta \epsilon [0,1]$.

- $F_q(\theta)$ and $F_u(\theta)$ are signal CDF's conditional on worker qualification. The respective PDF's are $f_q(\theta)$ and $f_u(\theta)$. Assume that the likelihood ratio $\varphi(\theta)=[f_u(\theta)/f_q(\theta)]$ is a strictly decreasing function of θ. (MLRP)

- Let $\xi(\pi,\theta)$ be posterior probability that worker with signal θ is qualified, given that the fraction of his group who are qualified is π. Then, using *Bayes Rule*:

Employers’ Assignment Behavior:

- Employers cannot directly observe a worker’s qualification before they make an assignment. Instead they observe the noisy signal, $\theta \in [0,1]$.

- $F_q(\theta)$ and $F_u(\theta)$ are signal CDF’s conditional on worker qualification. The respective PDF’s are $f_q(\theta)$ and $f_u(\theta)$. Assume that the likelihood ratio $\varphi(\theta)=[f_u(\theta)/f_q(\theta)]$ is a strictly decreasing function of θ. (MLRP)

- Let $\xi(\pi,\theta)$ be posterior probability that worker with signal θ is qualified, given that the fraction of his group who are qualified is π. Then, using *Bayes Rule*:

$$\xi(\pi,\theta) \equiv \pi f_q(\theta)/[\pi f_q(\theta)+(1-\pi)f_u(\theta)]$$

$$=1/\{1+[(1-\pi)/\pi]\varphi(\theta)\}.$$

Employers' Assignment Behavior:

- Employers cannot directly observe a worker's qualification before they make an assignment. Instead they observe the noisy signal, $\theta \epsilon [0,1]$.

- $F_q(\theta)$ and $F_u(\theta)$ are signal CDF's conditional on worker qualification. The respective PDF's are $f_q(\theta)$ and $f_u(\theta)$. Assume that the likelihood ratio $\varphi(\theta)=[f_u(\theta)/f_q(\theta)]$ is a strictly decreasing function of θ. (MLRP)

- Let $\xi(\pi,\theta)$ be posterior probability that worker with signal θ is qualified, given that the fraction of his group who are qualified is π. Then, using *Bayes Rule*:

$$\xi(\pi,\theta) \equiv \pi f_q(\theta) / [\pi f_q(\theta) + (1-\pi) f_u(\theta)]$$

$$= 1/\{1 + [(1-\pi)/\pi]\varphi(\theta)\}.$$

- Notice that, given our assumptions, $\xi(\pi,\theta)$ strictly increases with π and strictly increases with θ, for all $(\pi,\theta) \in (0,1)^2$.

Employers' Assignment Behavior (cont'd):

- The expected net gain to an employer from assigning some worker to the desirable task – if the worker's group is skilled at rate π and if the worker presents the noisy signal θ – is positive if and only if:

Employers' Assignment Behavior (cont'd):

- The expected net gain to an employer from assigning some worker to the desirable task – if the worker's group is skilled at rate π and if the worker presents the noisy signal θ – is positive if and only if:

$$\xi(\pi,\theta)x_q > [1-\xi(\pi,\theta)]x_u$$

Employers' Assignment Behavior (cont'd):

- The expected net gain to an employer from assigning some worker to the desirable task – if the worker's group is skilled at rate π and if the worker presents the noisy signal θ – is positive if and only if:

$$\xi(\pi,\theta)x_q > [1-\xi(\pi,\theta)]x_u$$

- So, the employer will assign this worker to the desirable task if and only if his expected net gain is non-negative. That is, if and only if $\{\theta \geq s\}$, where s is the *assignment threshold* which satisfies the following equation:

Employers' Assignment Behavior (cont'd):

- The expected net gain to an employer from assigning some worker to the desirable task – if the worker's group is skilled at rate π and if the worker presents the noisy signal θ – is positive if and only if:

$$\xi(\pi,\theta)x_q > [1-\xi(\pi,\theta)]x_u$$

- So, the employer will assign this worker to the desirable task if and only if his expected net gain is non-negative. That is, if and only if $\{\theta \geq s\}$, where s is the *assignment threshold* which satisfies the following equation:

$$(1)\quad \frac{\pi x_q}{(1-\pi)x_u} = \varphi(s)$$

Employers' Assignment Behavior (cont'd):

- The expected net gain to an employer from assigning some worker to the desirable task – if the worker's group is skilled at rate π and if the worker presents the noisy signal θ – is positive if and only if:

$$\xi(\pi,\theta)x_q > [1-\xi(\pi,\theta)]x_u$$

- So, the employer will assign this worker to the desirable task if and only if his expected net gain is non-negative. That is, if and only if $\{\theta \geq s\}$, where s is the *assignment threshold* which satisfies the following equation:

$$(1)\quad \frac{\pi x_q}{(1-\pi)x_u} = \varphi(s)$$

- Here we are using the fact that $\frac{\xi(\pi,\theta)}{1-\xi(\pi,\theta)} = \left(\frac{\pi}{1-\pi}\right) \bullet \left(\frac{f_q(\theta)}{f_u(\theta)}\right)$

Employers' Assignment Behavior (cont'd):

- The expected net gain to an employer from assigning some worker to the desirable task – if the worker's group is skilled at rate π and if the worker presents the noisy signal θ – is positive if and only if:

$$\xi(\pi,\theta)x_q > [1-\xi(\pi,\theta)]x_u$$

- So, the employer will assign this worker to the desirable task if and only if his expected net gain is non-negative. That is, if and only if $\{\theta \geq s\}$, where s is the *assignment threshold* which satisfies the following equation:

$$(1)\quad \frac{\pi x_q}{(1-\pi)x_u} = \varphi(s)$$

- Here we are using the fact that $\frac{\xi(\pi,\theta)}{1-\xi(\pi,\theta)} = \left(\frac{\pi}{1-\pi}\right) \bullet \left(\frac{f_q(\theta)}{f_u(\theta)}\right)$

- So, if employers think the fraction π of some group are qualified, they choose an assignment threshold s for that group satisfying equation (1).

Workers’ Skill Acquisition Behavior

Workers’ Skill Acquisition Behavior

- Let the gross gain to a worker from obtaining a desirable position be $w>0$. Assume workers correctly anticipate employers’ assignment threshold, s.

Workers’ Skill Acquisition Behavior

- Let the gross gain to a worker from obtaining a desirable position be $w>0$. Assume workers correctly anticipate employers’ assignment threshold, s.

- Then a worker with cost c, facing an employer believed to be using the assignment threshold s, will gain from becoming qualified if and only if:

Workers' Skill Acquisition Behavior

- Let the gross gain to a worker from obtaining a desirable position be $w>0$. Assume workers correctly anticipate employers' assignment threshold, s.

- Then a worker with cost c, facing an employer believed to be using the assignment threshold s, will gain from becoming qualified if and only if:

$$w[1-F_q(s)] - c \geq w[1-F_u(s)]$$

Workers' Skill Acquisition Behavior

- Let the gross gain to a worker from obtaining a desirable position be $w>0$. Assume workers correctly anticipate employers' assignment threshold, s.

- Then a worker with cost c, facing an employer believed to be using the assignment threshold s, will gain from becoming qualified if and only if:

$$w[1-F_q(s)] - c \geq w[1-F_u(s)]$$

that is, if and only if:

$$w[F_u(s) - F_q(s)] = w\Delta F(s) \geq c,$$

Workers' Skill Acquisition Behavior

- Let the gross gain to a worker from obtaining a desirable position be $w>0$. Assume workers correctly anticipate employers' assignment threshold, s.

- Then a worker with cost c, facing an employer believed to be using the assignment threshold s, will gain from becoming qualified if and only if:

$$w[1-F_q(s)] - c \geq w[1-F_u(s)]$$

that is, if and only if:

$$w[F_u(s) - F_q(s)] = w\Delta F(s) \geq c,$$

where $\Delta F(s)$ is the amount by which becoming qualified increases the worker's probability of being assigned to the desirable task when the assignment threshold is s. (Note: $\Delta F(0)=\Delta F(1)=0$, and $\Delta F(s)$ is single-peaked.)

Workers' Skill Acquisition Behavior

- Let the gross gain to a worker from obtaining a desirable position be $w>0$. Assume workers correctly anticipate employers' assignment threshold, s.

- Then a worker with cost c, facing an employer believed to be using the assignment threshold s, will gain from becoming qualified if and only if:

$$w[1-F_q(s)] - c \geq w[1-F_u(s)]$$

that is, if and only if:

$$w[F_u(s) - F_q(s)] = w\Delta F(s) \geq c,$$

where $\Delta F(s)$ is the amount by which becoming qualified increases the worker's probability of being assigned to the desirable task when the assignment threshold is s. (Note: $\Delta F(0)=\Delta F(1)=0$, and $\Delta F(s)$ is single-peaked.)

- So, if workers believe that employers will use assignment threshold s for their group, the fraction of them becoming qualified satisfies equation (2):

$$(2)\quad \pi = G(w\Delta F(s))$$

Equilibrium Worker-Employer Interactions

Equilibrium Worker-Employer Interactions

- Thus, the behaviors of employers and workers are defined by the following two equations connecting the fraction of a group of workers who become qualified, π, to the employer's assignment threshold, s:

Equilibrium Worker-Employer Interactions

- Thus, the behaviors of employers and workers are defined by the following two equations connecting the fraction of a group of workers who become qualified, π, to the employer's assignment threshold, s:

- (1) Employers (the EE curve): {(s,π) such that $\frac{\pi x_q}{(1-\pi) x_u} = \varphi(s)$}

Equilibrium Worker-Employer Interactions

- Thus, the behaviors of employers and workers are defined by the following two equations connecting the fraction of a group of workers who become qualified, π, to the employer's assignment threshold, s:

- (1) Employers (the EE curve): {(s,π) such that $\frac{\pi x_q}{(1-\pi) x_u} = \varphi(s)$}

- (2) Workers (the WW curve): {(s,π) such that $\pi = G(w\Delta F(s))$}

Equilibrium Worker-Employer Interactions

- Thus, the behaviors of employers and workers are defined by the following two equations connecting the fraction of a group of workers who become qualified, π, to the employer's assignment threshold, s:

- (1) Employers (the EE curve): $\{(s,\pi)$ such that $\frac{\pi x_q}{(1-\pi)x_u} = \varphi(s)\}$

- (2) Workers (the WW curve): $\{(s,\pi)$ such that $\pi = G(w\Delta F(s))\}$

- An *equilibrium* in this model may therefore be identified with any pair of numbers $(s^*,\pi^*)\epsilon[0,1]^2$ satisfying these two equations.

Equilibrium Worker-Employer Interactions

- Thus, the behaviors of employers and workers are defined by the following two equations connecting the fraction of a group of workers who become qualified, π, to the employer's assignment threshold, s:

- (1) Employers (the EE curve): $\{(s,\pi)$ such that $\frac{\pi x_q}{(1-\pi)x_u} = \varphi(s)\}$

- (2) Workers (the WW curve): $\{(s,\pi)$ such that $\pi = G(w\Delta F(s))\}$

- An *equilibrium* in this model may therefore be identified with any pair of numbers $(s^*,\pi^*)\epsilon[0,1]^2$ satisfying these two equations.

- *Any intersection of EE and WW curves is an equilibrium: The employers' anticipation of the group's qualification rate, π^*, and workers' anticipation of the employers' assignment threshold, s^*, are mutually consistent.*

Schematic Representation of Players' Payoffs

Employer's action, A

Worker's action, I

Payoffs to {W,E}	E assigns W to desired position: (A=1) when $\{\frac{\pi x_q}{(1-\pi)xu} \geq \varphi(\theta)\}$	E doesn't assign W to desired position: (A=0) when $\{\frac{\pi x_q}{(1-\pi)xu} < \varphi(\theta)\}$
Worker obtains costly skill: (I=1) when $\{w\Delta F(s) \geq c\}$	$\{w - c, x_q\}$	$\{-c,0\}$
Worker doesn't obtain costly skill: (I=0) when $\{w\Delta F(s) < c\}$	$\{w, -x_u\}$	$\{0,0\}$

Equilibria = { (π^*,s^*) such that: (1) $\frac{x_q \pi *}{x_u(1-\pi *)} = \varphi(s*)$ and (2) $\pi^* = G(w\Delta F(s^*))$}

Depicting Equilibrium With Group Stereotypes

Depicting Equilibrium With Group Stereotypes

Since the groups B and W are visibly distinct, the existence of multiple equilibria (s_w,π_w) and (s_b,π_b), both satisfying the foregoing equations, but with $\pi_w > \pi_b$, creates the possibility of ***rational negative stereotypes*** disadvantaging the Bs.

Depicting Equilibrium With Group Stereotypes

Since the groups B and W are visibly distinct, the existence of multiple equilibria (s_w, π_w) and (s_b, π_b), both satisfying the foregoing equations, but with $\pi_w > \pi_b$, creates the possibility of ***rational negative stereotypes*** disadvantaging the Bs.

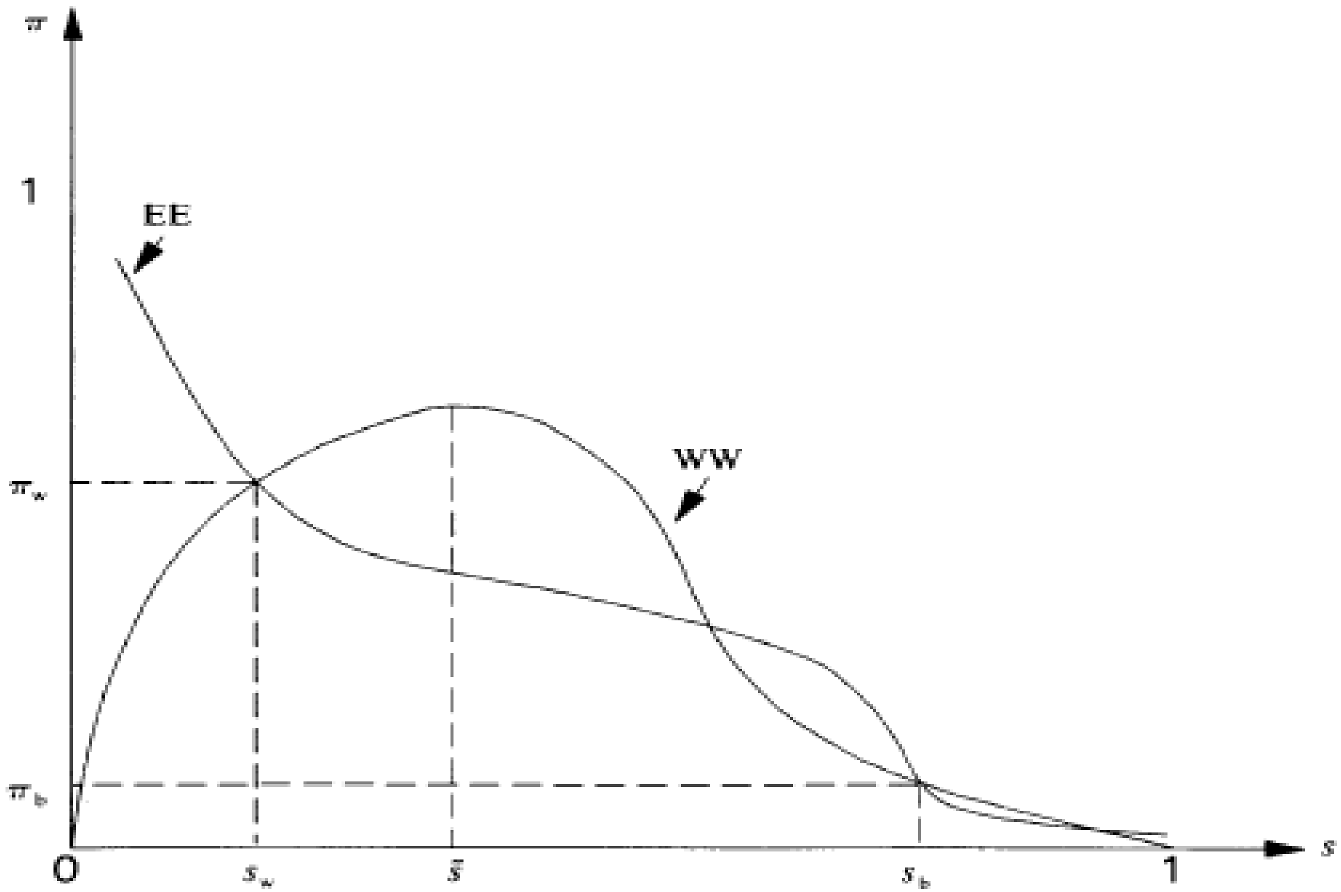


FIGURE 2. AN EQUILIBRIUM WITH NEGATIVE STEREOTYPES AGAINST B'S

Affirmative Action as Remedy for Negative Stereotypes

Affirmative Action as Remedy for Negative Stereotypes

- Notice that a group with fraction π skilled facing the threshold s will end up with fraction $\rho(s,\pi)$ being assigned to desired positions, where:

Affirmative Action as Remedy for Negative Stereotypes

- Notice that a group with fraction π skilled facing the threshold s will end up with fraction ρ(s,π) being assigned to desired positions, where:

$$\rho(s,\pi) \equiv \pi[1 - F_q(s)] + (1-\pi)[1 - F_u(s)]$$

Affirmative Action as Remedy for Negative Stereotypes

- Notice that a group with fraction π skilled facing the threshold s will end up with fraction ρ(s,π) being assigned to desired positions, where:

$$\rho(s,\pi) \equiv \pi[1 - F_q(s)] + (1 - \pi)[1 - F_u(s)]$$

- Also, the per capita expected payoff to an employer using assignment threshold s and facing group with fraction π skilled is P(s,π), where:

Affirmative Action as Remedy for Negative Stereotypes

- Notice that a group with fraction π skilled facing the threshold s will end up with fraction ρ(s,π) being assigned to desired positions, where:

$$\rho(s,\pi) \equiv \pi[1 - F_{\mathrm{q}}(s)] + (1-\pi)[1 - F_{\mathrm{u}}(s)]$$

- Also, the per capita expected payoff to an employer using assignment threshold s and facing group with fraction π skilled is P(s,π), where:

$$P(s,\pi) \equiv \pi[1 - F_{\mathrm{q}}(s)]x_{\mathrm{q}} - (1-\pi)[1 - F_{\mathrm{u}}(s)]x_{\mathrm{u}}$$

Affirmative Action as Remedy for Negative Stereotypes

- Notice that a group with fraction π skilled facing the threshold s will end up with fraction ρ(s,π) being assigned to desired positions, where:

$$\rho(s,\pi) \equiv \pi[1 - F_q(s)] + (1-\pi)[1 - F_u(s)]$$

- Also, the per capita expected payoff to an employer using assignment threshold s and facing group with fraction π skilled is P(s,π), where:

$$P(s,\pi) \equiv \pi[1 - F_q(s)]x_q - (1-\pi)[1 - F_u(s)]x_u$$

- Consider, then, an Affirmative Action Policy where a regulator who can observe an employer's assignment decisions, but not workers' skills or the noisy signals of their skills, requires that the employer must assign the workers from the two groups to desirable positions at equal rates. (Note: we are ruling out the possibility that the regulator, by observing the workers' noisy signals, could simply force the employer to use the same assignment threshold for the two groups.)

Affirmative Action as Remedy (continued):

- In order to comply with this regulation, an employer who believes that the fractions (π_b,π_w) of the two groups have acquired the skill, where $\pi_w \geq \pi_b$, must choose assignment thresholds $\{s_w \geq s_b\}$ in such a manner that:

Affirmative Action as Remedy (continued):

- In order to comply with this regulation, an employer who believes that the fractions (π_b,π_w) of the two groups have acquired the skill, where $\pi_w \geq \pi_b$, must choose assignment thresholds $\{s_w \geq s_b\}$ in such a manner that:

AA Constraint: $\rho(s_b, \pi_b) = \rho(s_w, \pi_w)$

Affirmative Action as Remedy (continued):

- In order to comply with this regulation, an employer who believes that the fractions (π_b,π_w) of the two groups have acquired the skill, where $\pi_w \geq \pi_b$, must choose assignment thresholds $\{s_w \geq s_b\}$ in such a manner that:

AA Constraint: $$\rho(s_b, \pi_b) = \rho(s_w, \pi_w)$$

- On the other hand, if the employer believes skill levels in the two groups are (π_b,π_w), then using the assignment thresholds (s_b,s_w), and considering the groups' population proportions, implies a per capita expected payoff of:

Affirmative Action as Remedy (continued):

- In order to comply with this regulation, an employer who believes that the fractions (π_b,π_w) of the two groups have acquired the skill, where $\pi_w \geq \pi_b$,, must choose assignment thresholds $\{s_w \geq s_b\}$ in such a manner that:

AA Constraint: $\rho(s_b, \pi_b) = \rho(s_w, \pi_w)$

- On the other hand, if the employer believes skill levels in the two groups are (π_b,π_w), then using the assignment thresholds (s_b,s_w), and considering the groups' population proportions, implies a per capita expected payoff of:

Employer's Expected Payoff = $[(1-\lambda)P(s_b, \pi_b) + \lambda P(s_w, \pi_w)]$

Affirmative Action as Remedy (continued):

- In order to comply with this regulation, an employer who believes that the fractions (π_b,π_w) of the two groups have acquired the skill, where $\pi_w \geq \pi_b$, must choose assignment thresholds {$s_w \geq s_b$} in such a manner that:

AA Constraint: $\rho(s_b, \pi_b) = \rho(s_w, \pi_w)$

- On the other hand, if the employer believes skill levels in the two groups are (π_b,π_w), then using the assignment thresholds (s_b,s_w), and considering the groups' population proportions, implies a per capita expected payoff of:

Employer's Expected Payoff = $[(1-\lambda)P(s_b, \pi_b) + \lambda P(s_w, \pi_w)]$

- So, we model an employer's response to this affirmative action regulation as him choosing assignment thresholds (s_b,s_w), given beliefs (π_b,π_w), so as to maximize his per capital expected payoff subject to the AA constraint.

Affirmative Action as Remedy (continued):

- Let γ be Lagrange multiplier on AA constraint in employers' maximization problem. If $\pi_b = \pi_w$ the constraint does not bind, so $\gamma = 0$. But if $\pi_b < \pi_w$ then the constraint must bind (since the EE curve is strictly decreasing) so $\gamma > 0$.

Affirmative Action as Remedy (continued):

- Let γ be Lagrange multiplier on AA constraint in employers' maximization problem. If $\pi_b=\pi_w$ the constraint does not bind, so $\gamma=0$. But if $\pi_b<\pi_w$ then the constraint must bind (since the EE curve is strictly decreasing) so $\gamma>0$.

- Intuitively, γ is the reduction in the employer's payoff associated with him being required to assign a marginally greater fraction of group B workers to the desired task, at the constrained optimal assignment thresholds (s_b,s_w).

Affirmative Action as Remedy (continued):

- Let γ be Lagrange multiplier on AA constraint in employers' maximization problem. If $\pi_b=\pi_w$ the constraint does not bind, so $\gamma=0$. But if $\pi_b<\pi_w$ then the constraint must bind (since the EE curve is strictly decreasing) so $\gamma>0$.

- Intuitively, γ is the reduction in the employer's payoff associated with him being required to assign a marginally greater fraction of group B workers to the desired task, at the constrained optimal assignment thresholds (s_b,s_w).

- Then, given λ, π_b and π_w, it may be shown that these first-order conditions fully characterize an employer's optimal response to the AA policy:

Affirmative Action as Remedy (continued):

- Let γ be Lagrange multiplier on AA constraint in employers' maximization problem. If $\pi_b=\pi_w$ the constraint does not bind, so $\gamma=0$. But if $\pi_b<\pi_w$ then the constraint must bind (since the EE curve is strictly decreasing) so $\gamma>0$.

- Intuitively, γ is the reduction in the employer's payoff associated with him being required to assign a marginally greater fraction of group B workers to the desired task, at the constrained optimal assignment thresholds (s_b,s_w).

- Then, given λ, π_b and π_w, it may be shown that these first-order conditions fully characterize an employer's optimal response to the AA policy:

- (3) The $EE_b(\gamma)$ curve: $\{(s_b,\pi_b)$ such that $\dfrac{\pi_b\,(xq+\frac{\gamma}{1-\lambda})}{(1-\pi_b)(xu-\frac{\gamma}{1-\lambda})}=\varphi(s_b)\}$

Affirmative Action as Remedy (continued):

- Let γ be Lagrange multiplier on AA constraint in employers' maximization problem. If $\pi_b=\pi_w$ the constraint does not bind, so γ=0. But if $\pi_b<\pi_w$ then the constraint must bind (since the EE curve is strictly decreasing) so γ>0.

- Intuitively, γ is the reduction in the employer's payoff associated with him being required to assign a marginally greater fraction of group B workers to the desired task, at the constrained optimal assignment thresholds (s_b,s_w).

- Then, given λ, π_b and π_w, it may be shown that these first-order conditions fully characterize an employer's optimal response to the AA policy:

- (3) The $EE_b(\gamma)$ curve: $\{(s_b,\pi_b)$ such that $\dfrac{\pi_b\,(xq+\frac{\gamma}{1-\lambda})}{(1-\pi_b)(xu-\frac{\gamma}{1-\lambda})}=\varphi(s_b)\}$

- (4) The $EE_w(\gamma)$ curve: $\{(s_w,\pi_w)$ such that $\dfrac{\pi_w\,(xq-\frac{\gamma}{\lambda})}{(1-\pi_w)(xu+\frac{\gamma}{\lambda})}=\varphi(s_w)\}$

Affirmative Action as Remedy (continued):

- Let γ be Lagrange multiplier on AA constraint in employers' maximization problem. If $\pi_b=\pi_w$ the constraint does not bind, so $\gamma=0$. But if $\pi_b<\pi_w$ then the constraint must bind (since the EE curve is strictly decreasing) so $\gamma>0$.

- Intuitively, γ is the reduction in the employer's payoff associated with him being required to assign a marginally greater fraction of group B workers to the desired task, at the constrained optimal assignment thresholds (s_b,s_w).

- Then, given λ, π_b and π_w, it may be shown that these first-order conditions fully characterize an employer's optimal response to the AA policy:

- (3) The $EE_b(\gamma)$ curve: $\{(s_b,\pi_b)$ such that $\dfrac{\pi_b\,(xq+\frac{\gamma}{1-\lambda})}{(1-\pi_b)(xu-\frac{\gamma}{1-\lambda})} = \varphi(s_b)\}$

- (4) The $EE_w(\gamma)$ curve: $\{(s_w,\pi_w)$ such that $\dfrac{\pi_w\,(xq-\frac{\gamma}{\lambda})}{(1-\pi_w)(xu+\frac{\gamma}{\lambda})} = \varphi(s_w)\}$

- (5) The AA constraint: $\{(s_b,\pi_b)$ and (s_w,π_w) such that $\rho(s_b,\pi_b)=\rho(s_w,\pi_w)$ $\}$

Affirmative Action as Remedy (continued):

- Let γ be Lagrange multiplier on AA constraint in employers' maximization problem. If $\pi_b=\pi_w$ the constraint does not bind, so $\gamma=0$. But if $\pi_b<\pi_w$ then the constraint must bind (since the EE curve is strictly decreasing) so $\gamma>0$.

- Intuitively, γ is the reduction in the employer's payoff associated with him being required to assign a marginally greater fraction of group B workers to the desired task, at the constrained optimal assignment thresholds (s_b,s_w).

- Then, given λ, π_b and π_w, it may be shown that these first-order conditions fully characterize an employer's optimal response to the AA policy:

- (3) The $EE_b(\gamma)$ curve: $\{(s_b,\pi_b)$ such that $\dfrac{\pi_b\,(xq+\frac{\gamma}{1-\lambda})}{(1-\pi_b)(xu-\frac{\gamma}{1-\lambda})}=\varphi(s_b)\}$

- (4) The $EE_w(\gamma)$ curve: $\{(s_w,\pi_w)$ such that $\dfrac{\pi_w\,(xq-\frac{\gamma}{\lambda})}{(1-\pi_w)(xu+\frac{\gamma}{\lambda})}=\varphi(s_w)\}$

- (5) The AA constraint: $\{(s_b,\pi_b)$ and (s_w,π_w) such that $\rho(s_b,\pi_b)=\rho(s_w,\pi_w)\}$

- (6) Finally, optimal worker behavior requires that: $\pi_i = G(w\Delta F(s_i))$, for i=B,W – that is, both (s_b,π_b) and (s_w,π_w) must also lie on the WW curve.

- *Define an* **Equilibrium under AA** *as a triple* $\{(s_b,\pi_b);(s_w,\pi_w);\gamma\}$ *satisfying (3)-(6), such that each of the* (s_i,π_i) *lie on both the* $EE_i(\gamma)$ *and WW curves, i=B,W, and such that the AA constraint holds as an equality.* (Notice that if $\pi_b=\pi_w$ in Equilibrium under AA, then $\gamma=0$, and there are no negative stereotypes! Unfortunately, other equilibria may exist.)

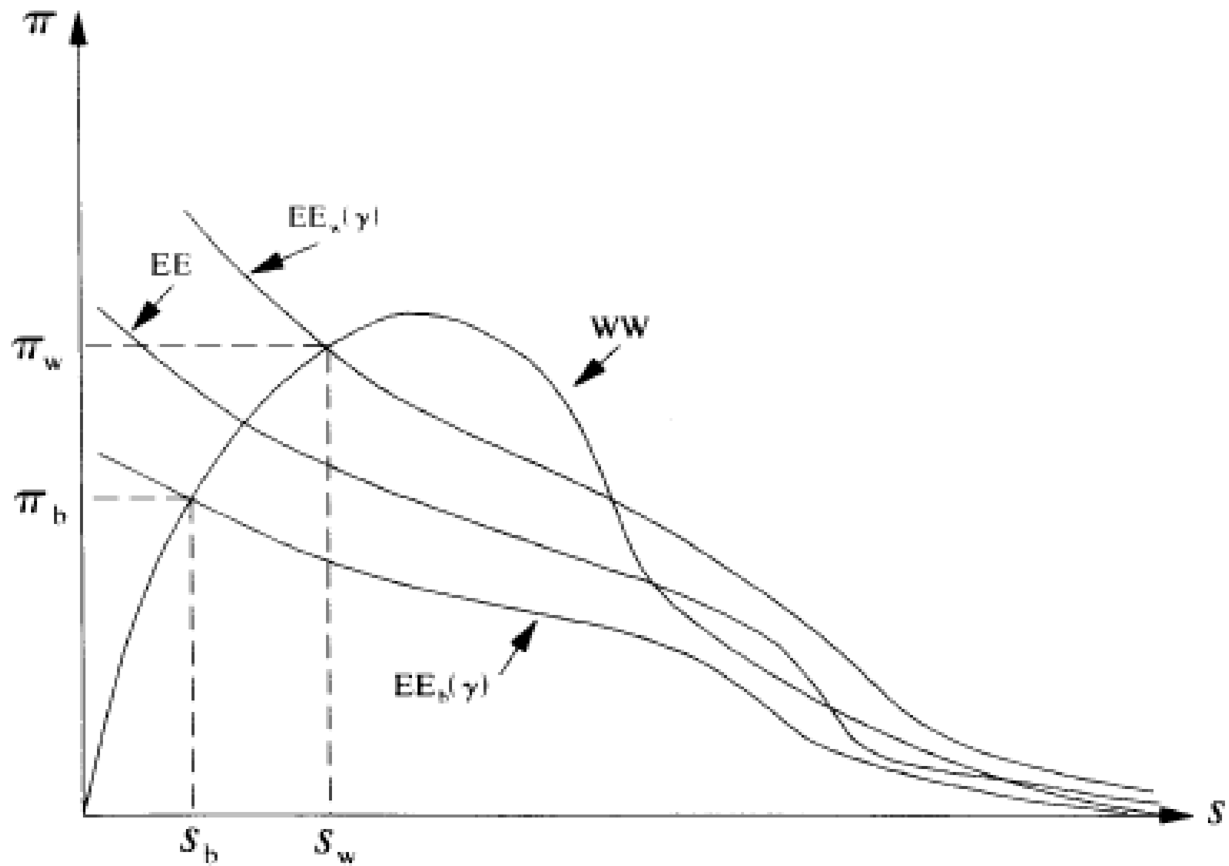


FIGURE 3. EMPLOYERS' OPTIMAL STANDARDS FOR B'S AND W'S GIVEN BELIEFS AND POSITIVE VALUE OF MULTIPLIER

So, When Will Affirmative Action Eliminate Negative Stereotypes?

So, When Will Affirmative Action Eliminate Negative Stereotypes?

- Define $\rho^*(s)$ as the fraction of a group facing hiring threshold s, and investing in skills accordingly, who end-up being assigned by employers to the desired task.

So, When Will Affirmative Action Eliminate Negative Stereotypes?

- Define $\rho^*(s)$ as the fraction of a group facing hiring threshold s, and investing in skills accordingly, who end-up being assigned by employers to the desired task.

- Therefore: $\rho^*(s) = \rho(s,G(w\Delta F(s)) = G(w\Delta F(s)[1-F_q(s)] + [1-G(w\Delta F(s)][1-F_u(s)]$

So, When Will Affirmative Action Eliminate Negative Stereotypes?

- Define $\rho^*(s)$ as the fraction of a group facing hiring threshold s, and investing in skills accordingly, who end-up being assigned by employers to the desired task.

- Therefore: $\rho^*(s) = \rho(s,G(w\Delta F(s)) = G(w\Delta F(s)[1-F_q(s)] + [1-G(w\Delta F(s)][1-F_u(s)]$

- Notice that $\rho^*(0) = 1$ and $\rho^*(1) = 0$. (I.e., when $s=0$, every worker is assigned to the desired task whether any are skilled or not. When $s=1$, no workers are so assigned.)

So, When Will Affirmative Action Eliminate Negative Stereotypes?

- Define $\rho^*(s)$ as the fraction of a group facing hiring threshold s, and investing in skills accordingly, who end-up being assigned by employers to the desired task.

- Therefore: $\rho^*(s) = \rho(s,G(w\Delta F(s)) = G(w\Delta F(s)[1-F_q(s)] + [1-G(w\Delta F(s)][1-F_u(s)]$

- Notice that $\rho^*(0) = 1$ and $\rho^*(1) = 0$. (I.e., when $s=0$, every worker is assigned to the desired task whether any are skilled or not. When $s=1$, no workers are so assigned.)

- In fact, as s rises, the changes in the function $\rho^*(s)$ embody two distinct effects:

So, When Will Affirmative Action Eliminate Negative Stereotypes?

- Define $\rho^*(s)$ as the fraction of a group facing hiring threshold s, and investing in skills accordingly, who end-up being assigned by employers to the desired task.

- Therefore: $\rho^*(s) = \rho(s,G(w\Delta F(s)) = G(w\Delta F(s)[1-F_q(s)] + [1-G(w\Delta F(s)][1-F_u(s)]$

- Notice that $\rho^*(0) = 1$ and $\rho^*(1) = 0$. (I.e., when $s=0$, every worker is assigned to the desired task whether any are skilled or not. When $s=1$, no workers are so assigned.)

- In fact, as s rises, the changes in the function $\rho^*(s)$ embody two distinct effects:

  - (a) *stricter rationing* of access to desired positions due to the use of a higher assignment threshold; and

So, When Will Affirmative Action Eliminate Negative Stereotypes?

- Define $\rho^*(s)$ as the fraction of a group facing hiring threshold s, and investing in skills accordingly, who end-up being assigned by employers to the desired task.

- Therefore: $\rho^*(s) = \rho(s,G(w\Delta F(s)) = G(w\Delta F(s)[1-F_q(s)] + [1-G(w\Delta F(s)][1-F_u(s)]$

- Notice that $\rho^*(0) = 1$ and $\rho^*(1) = 0$. (I.e., when $s=0$, every worker is assigned to the desired task whether any are skilled or not. When $s=1$, no workers are so assigned.)

- In fact, as s rises, the changes in the function $\rho^*(s)$ embody two distinct effects:

  - (a) *stricter rationing* of access to desired positions due to the use of a higher assignment threshold; and

  - (b) *altered incentives* for workers to become skilled, due to changes in $\Delta F(s)$.

So, When Will Affirmative Action Eliminate Negative Stereotypes? (continued)

- Note that $\rho^*(0) = 1$ and $\rho^*(1) = 0$. (If $s=0$ then every worker is assigned to the desired task whether any are skilled or not. If $s=1$ then no workers are so assigned.)

So, When Will Affirmative Action Eliminate Negative Stereotypes? (continued)

- Note that $\rho^*(0) = 1$ and $\rho^*(1) = 0$. (If $s=0$ then every worker is assigned to the desired task whether any are skilled or not. If $s=1$ then no workers are so assigned.)

- Notice also that the *stricter rationing effect* (a) is always negative. But the *altered incentive effect* (b) is positive at lower values of s, and it only becomes negative at higher values. (A stricter assignment threshold encourages skill acquisition when the threshold is low, but discourages acquiring skills as the threshold becomes higher.)

So, When Will Affirmative Action Eliminate Negative Stereotypes? (continued)

- Note that $\rho^*(0) = 1$ and $\rho^*(1) = 0$. (If $s=0$ then every worker is assigned to the desired task whether any are skilled or not. If $s=1$ then no workers are so assigned.)

- Notice also that the *stricter rationing effect* (a) is always negative. But the *altered incentive effect* (b) is positive at lower values of s, and it only becomes negative at higher values. (A stricter assignment threshold encourages skill acquisition when the threshold is low, but discourages acquiring skills as the threshold becomes higher.)

- Finally, observe that if one can restate the AA constraint as follows:

So, When Will Affirmative Action Eliminate Negative Stereotypes? (continued)

- Note that $\rho^*(0) = 1$ and $\rho^*(1) = 0$. (If $s=0$ then every worker is assigned to the desired task whether any are skilled or not. If $s=1$ then no workers are so assigned.)
- Notice also that the *stricter rationing effect* (a) is always negative. But the *altered incentive effect* (b) is positive at lower values of s, and it only becomes negative at higher values. (A stricter assignment threshold encourages skill acquisition when the threshold is low, but discourages acquiring skills as the threshold becomes higher.)
- Finally, observe that if one can restate the AA constraint as follows:

In any Equilibrium under AA it must be that $\rho^*(s_b) = \rho^*(s_w)$

So, When Will Affirmative Action Eliminate Negative Stereotypes? (continued)

- Note that $\rho^*(0) = 1$ and $\rho^*(1) = 0$. (If $s=0$ then every worker is assigned to the desired task whether any are skilled or not. If $s=1$ then no workers are so assigned.)

- Notice also that the *stricter rationing effect* (a) is always negative. But the *altered incentive effect* (b) is positive at lower values of s, and it only becomes negative at higher values. (A stricter assignment threshold encourages skill acquisition when the threshold is low, but discourages acquiring skills as the threshold becomes higher.)

- Finally, observe that if one can restate the AA constraint as follows:

In any Equilibrium under AA it must be that $\rho^*(s_b) = \rho^*(s_w)$

- Hence, we may conclude that if the function $\rho^*(s)$ is strictly decreasing throughout its range (0,1) then, in any Equilibrium under AA it must be the case that $s_b = s_w$ and, therefore, $\pi_b = \pi_w$. ***That is, strict monotonicity of $\rho^*(s)$ is a sufficient condition to insure that AA policy will always eliminate a negative stereotype against group B! It turns out that this is also a necessary condition for AA to have this effect!!***

So, When Will Affirmative Action Eliminate Negative Stereotypes? (continued)

- We can now state the main result of this paper:

So, When Will Affirmative Action Eliminate Negative Stereotypes? (continued)

- We can now state the main result of this paper:

Theorem: If the *stricter rationing effect* always outweighs the *incentive effect*, so ρ*(s) is decreasing throughout its range, then affirmative action policy as defined here always eliminates any negative stereotype against group B, in the sense that the only equilibria arising under the policy involve $\pi_a = \pi_b$ and $s_a = s_b$. However, if there is some range of assignment thresholds over which ρ*(s) is increasing (i.e., over which the *incentive effect* outweighs the *rationing effect*), then there is a non-negligible set of the parameter values (w,x_q,x_u,λ), for which equilibria exist under affirmative action with $\pi_a>\pi_b$ and $s_a>s_b$. Moreover, the larger is λ (that is, the smaller is the disadvantaged group), the larger will be the set of parameters (w, x_q, x_u) for which such *patronizing equilibria* exist.

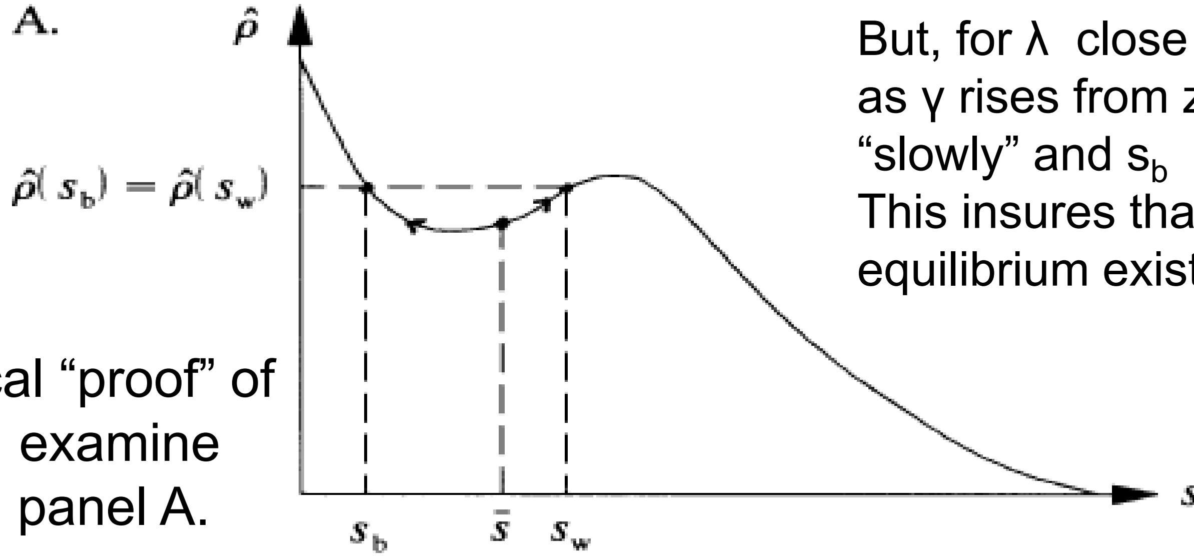
So, When Will Affirmative Action Eliminate Negative Stereotypes? (continued)

- We can now state the main result of this paper:

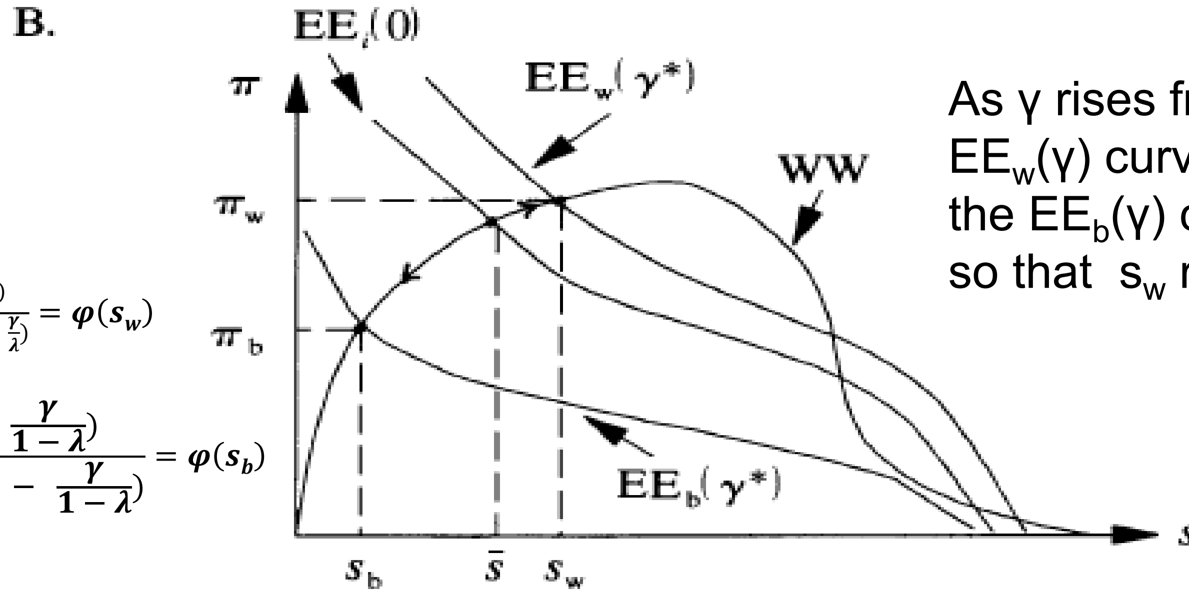
Theorem: If the *stricter rationing effect* always outweighs the *incentive effect*, so $\rho^*(s)$ is decreasing throughout its range, then affirmative action policy as defined here always eliminates any negative stereotype against group B, in the sense that the only equilibria arising under the policy involve $\pi_a = \pi_b$ and $s_a = s_b$. However, if there is some range of assignment thresholds over which $\rho^*(s)$ is increasing (i.e., over which the *incentive effect* outweighs the *rationing effect*), then there is a non-negligible set of the parameter values (w, x_q, x_u, λ), for which equilibria exist under affirmative action with $\pi_a > \pi_b$ and $s_a > s_b$. Moreover, the larger is λ (that is, the smaller is the disadvantaged group), the larger will be the set of parameters (w, x_q, x_u) for which such *patronizing equilibria* exist.

- We call an outcome under affirmative action where $\pi_a > \pi_b$ and $s_a > s_b$ a *patronizing equilibrium* because it has the following property: Employers think (correctly) that Bs are on average a less skilled group than As. So, being required by the regulation to assign As and Bs to desired positions at the same rate, employers believe they need to use a less rigorous assignment threshold in order to comply with the AA regulation. Yet, it is precisely because employers treat Bs less rigorously than As when making their assignment decisions (that is, they *patronize* Bs, and in so doing create a lower incentive for Bs than for As to acquire skills) that the negative stereotypes employers hold about group B workers exist! When this happens, the AA policy has backfired!

But, for λ close enough to one, as γ rises from zero, s_w rises "slowly" and s_b falls "quickly." This insures that a patronizing equilibrium exists for λ near one!



For a graphical "proof" of the Theorem, examine panel B, then panel A.



As γ rises from zero, the $EE_w(\gamma)$ curve shifts up and the $EE_b(\gamma)$ curve shifts down, so that s_w rises and s_b falls.

EE$_w$(γ): $\frac{\pi_w\,(xq - \frac{\gamma}{\lambda})}{(1-\pi_w)(xu + \frac{\gamma}{\lambda})} = \varphi(s_w)$

EE_b(γ): $\frac{\pi_b\,(xq + \frac{\gamma}{1-\lambda})}{(1-\pi_b)(xu - \frac{\gamma}{1-\lambda})} = \varphi(s_b)$

FIGURE 4. AN EQUILIBRIUM UNDER AFFIRMATIVE ACTION WITH NEGATIVE STEREOTYPE ABOUT B'S

E.g.: Consider this investment game with stereotypes. (Derived from Coate-Loury, "Will Affirmative Action Policies Eliminate Negative Stereotypes?", AER, 1993)

(1) Players

- An employer – who decides whether to hire workers for skilled positions
- Many workers – decides whether to invest and thereby become qualified for skilled work

(2) Employer do not observe workers' investments, but see "test" which is correlated with worker investment

(3) Investment costly to worker; getting hired into a skilled position always benefits workers, but only benefits employer if worker has invested.

Thus,

Payoff matrix

		Employer	
		A=0	A=1
Worker	I=0	0,0	1,-2
	I=1	-c,0	1-c,1

c is distributed as uniform [0,1]

Payoff=(worker, employer)

Assume “test” has three outcomes: “pass”, “fail”, and “unclear.” Assume investors cannot fail, and non-investors cannot pass.

Specifically, we take it that:

(i) Probability {test=“fail” given no investment} = 2/3
(ii) Probability {test=“unclear” given no investment} = 1/3

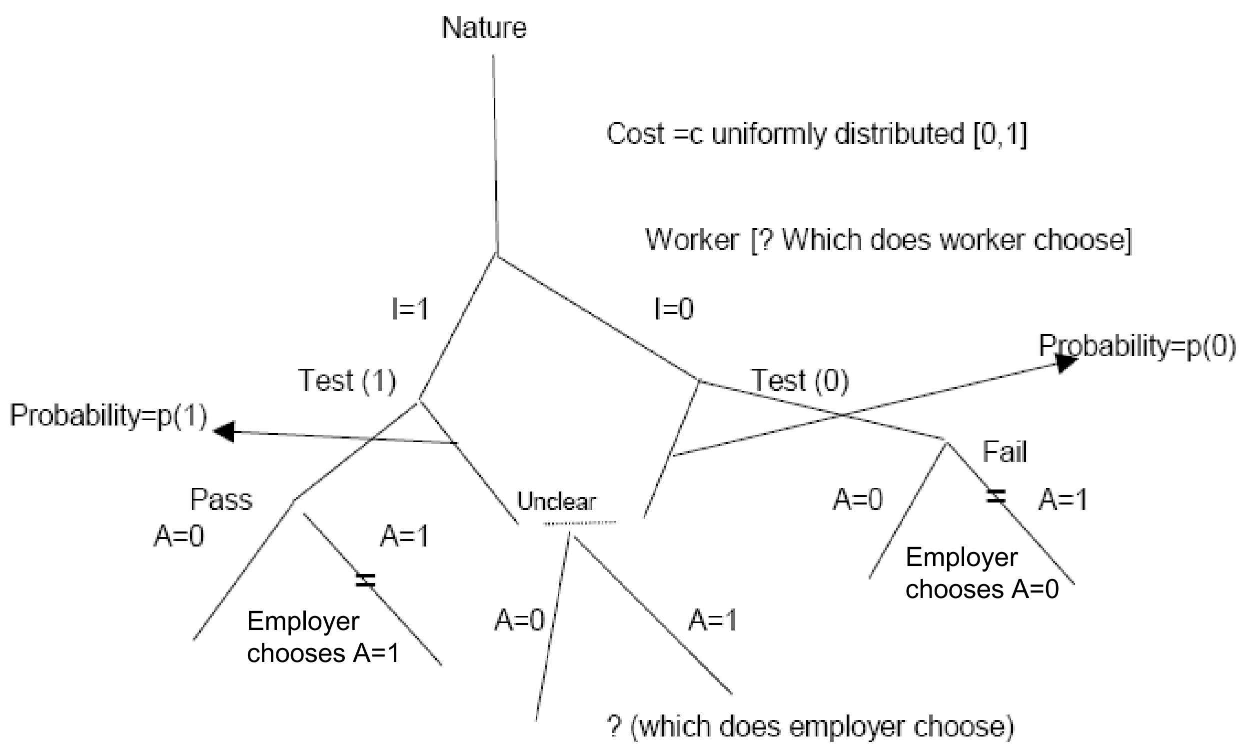
also

(iii) Probability {test=“pass”, given investment} = 1/3
(iv) Probability {test=“unclear”, given investment} = 2/3

[Note: we assume test is better at identifying those who do NOT invest than it is at detecting those who invest.]

Game Tree of Self-Confirming Racial Stereotypes Model (Coate-Loury 1993)

Game Tree



Here employer would want to choose A=1, giving worker "benefit of the doubt" only if the he thinks the odds that the worker has chosen I=1 are at least two-to-one

Now, let “q” denote a worker’s perception of the probability of being hired and let “s” be employer’s perceived probability that a given worker has invested.

(5) Next, we ask: when will the worker invest in becoming qualified?

- Let $q_1 \in (0,1)$ be the probability of getting hired in worker’s mind, if he invests.
- q_0= probability hired if not invest. Then:
- q_1*1-c= expected net benefit if I=1; and
- q_0*1-0 = expected net benefit if I=0
- $\Rightarrow$ a worker with cost c chooses I=1 if and only if $(q_1-q_0)\geq c$

(6) And, we also ask, when will the employer hire?

- Let $s\in(0,1)$ be employer’s probability worker has invested. Then:
- 0 = benefit to employer if he chooses A=0; and,
- s*1+(1-s)*(-2) = expected benefit if A=1. Hence A=1 if and only if $s\geq 2/3$

(7) But, what do workers and employers believe?

- Let test have 3 outcomes, t: Pass, fail, unclear. If a worker passes, then employer knows I=1; if he fails, then employer knows I=0
- Let Pr{“unclear”/I=0}=1/3 and Pr{“unclear”/I=1}=2/3
- Let the employer begin with the prior belief that a fraction $\Pi \in (0,1)$ of the overall worker population has invested. (In equilibrium this belief must be borne out by actual behavior in worker population.)

(8) From all of this, employing Bayes's Rule for conditional probabilities, it follows that the employer's posterior beliefs when seeing an “unclear” test are:

$$s = \frac{\Pi*(2/3)}{\Pi*(2/3)+(1-\Pi)*(1/3)} = \frac{2\pi}{1+\pi} = \frac{\{\#\ with\ I = 1\ and\ t = \text{unclear}\}}{\{total\ \#\ with\ t = \text{unclear}\}}$$

if the t=”unclear”; while s=0, if t=“fail”, and s=1, if t=“pass”

Finding Equilibrium in this Investment-Hiring Game:

1) Recall that the employer hires a worker for the skilled job if s = {probability the worker has invested} > 2/3.

2) But, if the employer believes that the fraction π of some group of workers have invested, and if a worker belonging to that group presents an "unclear" test result, then the employer's posterior probability of this particular worker being one who invested is $\frac{2\pi}{1+\pi}$, which exceeds 2/3 only if $\pi > ½$

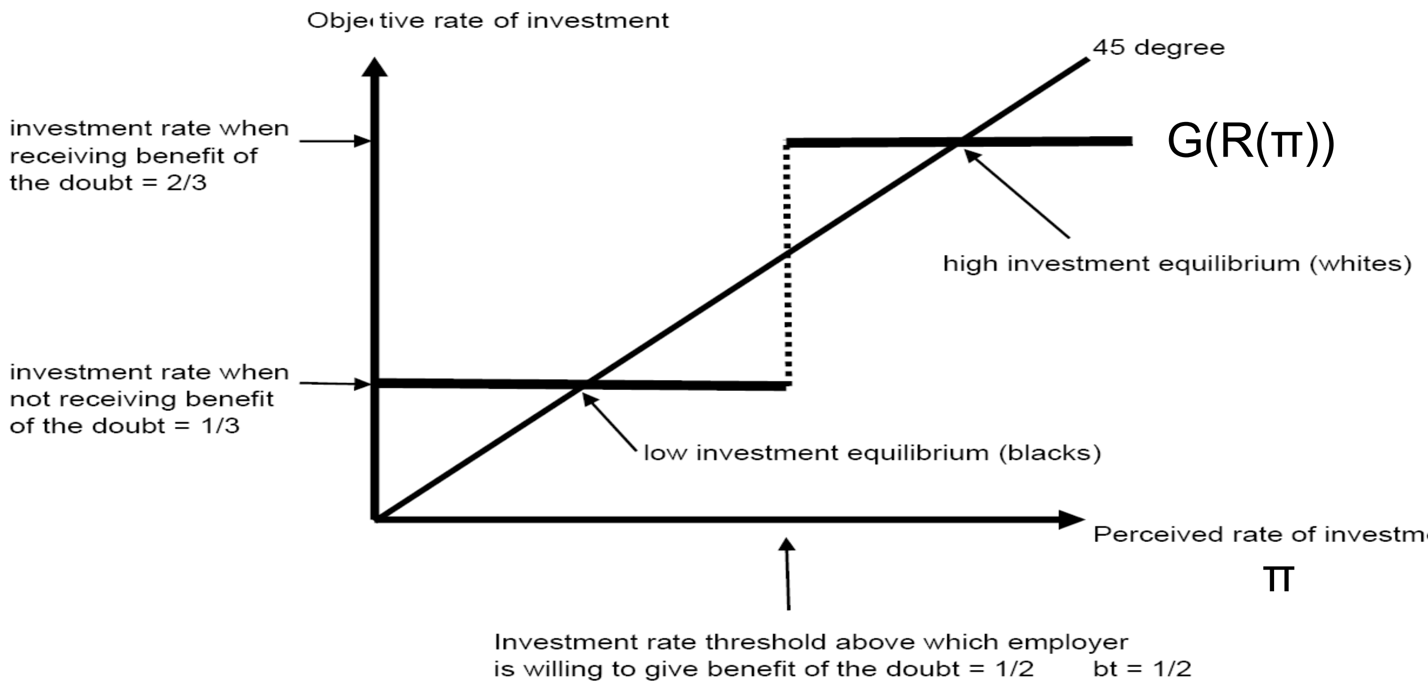
3) Hence, a worker gets the benefit of the employer's doubt only if the employer believes more than half of that worker's group have invested.

4) A worker who anticipates receiving the benefit of the doubt is guaranteed a skilled job if he invests, and has probability 1/3 of getting if he does not invest. Hence, that worker anticipates a return from investing equals [1 – 1/3] = 2/3

5) A worker who anticipates not receiving the benefit of the doubt has no chance of a skilled job without investing, and has probability 1/3 of getting the job if he invests. Hence, that worker's return from investing equals [1/3 – 0] =1/3

6) We conclude that in this example $R(\pi)=2/3$, for $\pi>1/2$, and $R(\pi)=1/3$, for $\pi<1/2$

7) Since $G(c)=c$ (a uniform cost distribution), we conclude that $\pi^* = G(R(\pi^*))$ for π a self-confirming stereotypic belief, π^*, only if either $\pi^* = 1/3$ or $\pi^* = 2/3$



Assumptions:

(1) Worker invests only if expected benefit outweighs cost

(2) Employer hires only if expected benefit is non-negative

(3) Only workers who invest (do not invest) can pass (fail) the test

Racial inequality due to a stereotype may lead to a demand for an affirmative intervention equalizing assignment rates:

Moreover, suppose that there are two visibly distinct population subgroups, B's and W's, and that the B's are in the low investment equilibrium ($\pi_B = \pi_L = 1/3$) while the W's are in the high investment equilibrium ($\pi_W = \pi_H = 2/3$). Let ρ_i denote the fraction of group i assigned to the skilled task in equilibrium. Then:

$$\rho_B = (1/3) * (1/3) = 1/9 \quad \text{while } \rho_W = (2/3) + (1/3) * (1/3) = 7/9$$

This circumstance may lead to some political demand that more B's be assigned to the skilled task. This is what we are calling "affirmative action."

How "Patronization" under Affirmative Action Works:

Notice, however, that if a policymaker aims to generate the same assignment rate of 7/9 to the skilled task for B's, but thinks only 1/3 of B's are investing (while 2/3 of W's are), then he will believe that it is necessary to **assign one-half of the B's who fail the test** to the skilled task because, counting all investors & non-investors:

$$(7/9) = (2/3) + (1/3)(1/3) = \{\text{fraction of W's to task one}\}$$

$$= (1/3) + (2/3)[(1/3) + (1/2)(2/3)] = \{\text{fraction of B's to task one}\}$$

Yet, if B's anticipate a 50% chance of getting assigned to the skilled task even when failing the test then not investing gives a (2/3) = (1/3) + (1/2)(2/3) chance of success, while investing guarantees success. So, only 1/3 of them would want to invest! Hence, *in this case, AA would remedy the **underrepresentation** problem for B's, but would NOT fix the **underinvestment** problem!! I.e., the B's would be **patronized**!*