Complementary Bias:
A Model of Two-Sided Statistical Discrimination

Ashley C. Craig and Roland G. Fryer, Jr*

Harvard University

January 2018

Abstract

We introduce a model of two-sided statistical discrimination in which worker and firm beliefs are complementary. Firms try to infer whether workers have made investments required for them to be productive, and simultaneously, workers try to deduce whether firms have made investments necessary for them to thrive. When multiple equilibria exist, group differences are sustained by both sides of the interaction – workers and firms. Strategic complementarity complicates both empirical analysis designed to detect discrimination and policy meant to alleviate it. Affirmative action is much less effective than in traditional statistical discrimination models. More generally, we demonstrate the futility of policies that are designed to correct gender and racial disparities but do not address both sides of the coordination problem. We propose a two-sided version of "investment insurance" – a policy in which the government (after observing a noisy version of the employer's signal) offers to hire any worker who it believes to be qualified and whom the employers does not offer a job – and show that it (weakly) dominates any alternative. The paper concludes by proposing a way to identify statistical discrimination by employers when beliefs are complements.

*We are grateful to Joseph Altonji, David Card, Kerwin Charles, Tanaya Devi, Christian Dustmann, Nicole Fortin, Edward Glaeser, Matthew Gentzkow, Louis Kaplow, Lawrence Katz, Kevin Lang, Glenn C. Loury, Costas Meghir, Jesse Shapiro, Betsey Stevenson, and seminar participants at NBER Labor Studies and Harvard University for comments and suggestions. Financial Support from the Harvard Kennedy School Women and Public Policy Program (Craig), an Inequality and Wealth Concentration Ph.D. Scholarship from the Multidisciplinary Program in Inequality & Social Policy at Harvard University (Craig) and Harvard University (Fryer) is gratefully acknowledged.

1 Introduction

Strategic complements, a term coined by Bulow, Geanakoplos, and Klemperer (1985), refer to decisions of at least two players that mutually reinforce one another. As Bulow et al. (1985) write, conventional substitutes and complements can be distinguished by whether a more "aggressive" strategy by firm A (e.g., a lower price in price competition or greater quantity in quantity competition) lowers or raises firm B's *total* profits. *Strategic* substitutes or complements are analogously defined by whether a more "aggressive" strategy by A lowers or raises B's *marginal* profits.

Given the history of widespread gender, racial, and ethnic discrimination around the globe, it is natural to think of worker and firm actions as strategic complements. Examples abound. As in traditional one-sided models, statistical discrimination against women could be generated by employers' asymmetric beliefs about the competence of men and women. But alternatively, given a history of bias in hiring and inflexible workplaces, women may believe that they will be treated unfairly, will encounter a hostile culture and will ultimately fail to be promoted. Women may therefore invest in such a way that causes employers to adjust their beliefs downward, even if they were initially homogeneous – confirming womens' suspicions. This mechanism could contribute to disparities in particular sectors, driving occupational segregation, as well as throughout the labor market more generally. As we show, such ingrained pessimism on both sides makes a given disparity in treatment much more difficult to address.

A similar dynamic may undergird racial disparities. Racial inequality in the 20^{th} century was generated by explicit racism and discrimination in almost every aspect of life. Parts of the South were plastered with signs that read "Negroes need not apply" (U.S. Congress 1963). In our model, if employers openly discriminate against blacks then, as a best response, black workers decide it does not make economic sense to invest. Now imagine that with the signing of the Civil Rights Act of 1964, employers stopped discriminating and hired with homogeneous beliefs, but workers were not convinced firms were amenable to minority workers. In traditional models of discrimination, minority outcomes improve immediately. In our model, however, even if employers had homogeneous beliefs immediately following the civil rights legislation, the equilibrium remains unchanged. Minorities, after years of subjugation, continue to believe that firms are hostile to minority workers and consequently do not invest. If they do not invest, employers adjust their beliefs downward. This illuminates the basic economics of our approach.

In this paper, we expand on the intuition from the gender and race examples above by building a model in which worker and firm actions are strategic complements vis-à-vis a two-sided statistical discrimination model.[1] Nature distributes costs of investment to workers and firms. We think of worker investment as classical Becker (1964) human capital. Firm investment is a fixed cost of creating a work environment conducive

[1]It is important to note at the outset that our game is not supermodular. There is not a strategic complementarity between *every* worker and *every* firm, but there is strategic complementary between *sides* (workers and firms).

to workers (e.g. flexible work hours for women or affinity groups for minority workers).[2] Workers (resp. firms) observe their costs and decide whether to invest. Conditional on this investment, nature distributes a signal to firms (regarding worker investment) and another to workers (regarding firm investment). Then workers, given their beliefs and observed signal, choose whether to apply to firms; and, conditional upon receiving an application, firms decide whether to hire.

Our model nests the classic one-sided treatment of statistical discrimination (e.g., Coate and Loury, 1993). Equilibria analogous to those in a one-sided model always exist in ours if the rate of firm investment is approximately fixed, which shuts down the strategic complementarity that drives our results. However, statistical discrimination can also be generated and sustained by *worker* pessimism. The resulting complementarity in beliefs between workers and firms makes the analysis of disparities between groups more complex but also considerably richer.

We begin our analysis of policy by considering affirmative action in the sense of a requirement that firms make job offers to members of both groups with equal probability. In models of statistical discrimination without strategic complementarity (classic one-sided models), such a requirement leads to homogeneous employer beliefs when lower hiring standards do not undermine worker investment, but negative stereotypes about minorities may persist if low standards are too de-motivating (Coate and Loury 1993).[3] Affirmative action can have the same issues in our model. But worse, affirmative action can undermine firm investment incentives and trigger zero investment by minority workers. Such severe inequality can be sustained indefinitely in our model despite affirmative action. Moreover, it may simply be impossible for affirmative action to eliminate discrimination because firms have less incentive to invest in a numerical minority if investment costs are fixed. Our model also allows us to analyze a more ambitious form of affirmative action – employment quotas – which requires employers to hire members of each group in proportion to their representation in the population. Employment quotas can cause firms to be overly aggressive in their attempts to hire minority workers, which can severely undermine minority worker investment, as well as harming the majority.

Generally, we demonstrate a kind of "impossibility result". Not only does every policy analyzed by Coate and Loury (1993) have the potential to be harmful, but *any* policy that fails to address the expectations of both sides (employers and workers) simultaneously will be ineffective. This result stems naturally from the two-sided nature of our model. Workers fail to invest both because of harsh hiring standards and because they are pessimistic about how they would fare in the workplace if they were hired. At the same time, firms are hesitant to make investments to support minority workers both because they think the workers they would attract would be unqualified and because minorities are not applying.

Consistent with this result, we provide suggestive evidence that two-sided policies are more effective

[2] In the main model, we assume this cost is fixed. In section 8.3.2, we provide some intuition for how results change if costs are proportional to the number of hires.

[3] Altonji and Blank (1999) show that these "patronizing equilibria" are eliminated if investment is a continuous variable.

at increasing the wages of disadvantaged groups. We focus on recent examples of job training programs (e.g., *Year Up* and *Per Scholas*) that have been unusually successful at increasing wages for disadvantaged youth. These programs combine worker investment with additional signals to firms, while at the same time demonstrating to workers that firms are investing. We argue that these programs are consistent with the policies suggested by our model. Firms can be more confident that workers have the relevant cognitive and non-cognitive skills to be successful, and workers know for sure that their investments will pay off because they are matched with an employer with demand for their type.

Randomized evaluations of *Year Up* and *Per Scholas* demonstrate that treatment youth earn, on average, 30% more than control youth (Roder and Elliot 2014, Hendra, Greenberg, Hamilton, et al. 2016). In an analysis of 207 other (one-sided) job training programs in Card (2015), the treatment effect on monthly earnings is 9.3%. We view this evidence as preliminary and incomplete but, coupled with the model, these results may help design future programs.

Based on our model, we propose a new policy – which we label "investment insurance" – as a simple solution to statistical discrimination. Imagine that the government can observe a noisy version of the signals employers and workers receive, and offers them contracts. If the government believes an individual invested, it will subsidize them. The same assurance is provided to employers regarding their investment. This provides assurance to both workers and firms that their investments will pay off. In our model, this weakly dominates any other policy – including affirmative action, employment quotas or wage subsidies.[4] Unlike these other policies, investment insurance can never be harmful to minority workers. Quite to the contrary, there is always a policy of this type that leads to full equality, and in weakly less time than any alternative. The underlying economics is similar to the concept of 'insulating tariffs' as discussed by Weyl (2010): the government effectively insulates workers and firms from uncertainty about investment by the other side.[5]

The paper concludes by deriving a model-based empirical test for statistical discrimination by employers. The test builds conceptually on the work of Altonji and Pierret (2001), Lang and Lehmann (2012), and Fryer, Pager and Spenkuch (2013), but it is designed to be robust to the confounds of worker belief formation and complementarity between firm and worker investment. Our analysis focuses directly on the mechanism through which rational stereotyping affects incentives: pessimistic employers shrink their estimates of worker productivity toward the group mean, causing a flattening of the relationship between productivity and wages. Based on this insight, we propose examining workers who switch firms. Under the assumption that firms gain some private information about a worker's ability with tenure, we demonstrate that wage profiles should flatten more for minority than majority workers when they move.

The paper proceeds as follows. The next section provides a brief review of the literature. Section 3 introduces our model and derives the basic implications of two-sided statistical discrimination and how it

[4] An analog also dominates affirmative action in traditional one-sided statistical discrimination models.
[5] We are grateful to Jesse Shapiro for pointing out this connection.

differs from traditional one-sided models. Section 4 discusses policies such as affirmative action, employment quotas, and wage subsidies. Section 5 considers two-sided policies. Sections 6 describes empirical implications of the two-sided approach. Section 7 concludes. Section 8, an Online Appendix, contains technical proofs, derivations omitted from the main analysis, and extensions of the basic model.

2 A Brief Review of the Literature

Our paper lies at the intersection of two important literatures: models of discrimination and models with strategic complementarities.[6] We briefly discuss each in turn.

A. Models of Discrimination

The two main theories of discrimination are a theory based on tastes pioneered by Becker (1957) and a statistical theory posited by Phelps (1972) and Arrow (1973).[7] Statistical models rely on imperfect observability of a worker's productivity to account for employers' use of a worker's group identity in their decision-making.

Phelps (1972) assumes available measures of productivity to be noisier for minority workers. One prediction of this model – developed by Aigner and Cain (1977) – is that there will be a wage gap at the top of the income distribution favoring whites and another gap at the bottom of the income distribution favoring blacks. Arrow (1973) demonstrates that statistical discrimination can occur even when there is no such unexplained group heterogeneity. The key insight in Arrow (1973) is that when employee productivity is endogenous, employer prejudice can be self-fulfilling.

An important contribution to this literature is Coate and Loury (1993) who formalize the insights in Arrow (1973) using a job assignment model in the spirit of Milgrom and Oster (1987). Coate and Loury (1993) provide sufficient conditions for multiple equilibria to exist and then demonstrate that an affirmative action policy may fail in the presence of statistical discrimination by perpetuating stereotypes.

There have been several important extensions of the Coate and Loury (1993) model. Moro and Norman (2004) embed Coate and Loury (1993) in a general equilibrium framework (with endogenous wages) and demonstrate that discrimination can occur even when the corresponding model with a single group has a unique equilibrium. Fang (2001) allows individuals to choose their group identity (i.e. social culture) and shows that allowing firms to give preferential treatment based on some seemingly irrelevant (chosen) group identity allows society to overcome an informational free-riding problem. Fryer (2007) develops a multi-stage

[6] Our work is also related to the literature on two-sided markets: e.g., Caillaud and Jullien (2003), Rochet and Tirole (2003), Anderson and Coate (2005), Armstrong (2006), and Weyl (2010). The labor market in our game could be viewed as a 'platform', with each side (workers vs. firms) benefiting from participation by the other. Strategic interactions are complicated in our model by imperfect information but this complementarity between sides is fundamental to the model and our policy recommendations.

[7] See Fang and Moro (2011) for a nice review of models of discrimination.

model of statistical discrimination and explores what happens to individuals who nonetheless overcome the initial discrimination. If an employer discriminates against a group in the first stage, she may actually favor members of that group when she makes promotion decisions within the firm.

Our paper builds on this literature – being close in spirit to Coate and Loury's work. The simple idea is that the original Arrow (1973) insight applies to both sides of the market: employers' prejudicial beliefs can be self-fulfilling but so too can workers' prejudicial beliefs about employers.[8] This insight has far-reaching implications for policy and empirical analysis. As we show, policies that have been proposed based on one-sided models are ineffective and potentially harmful. However, our analysis suggests under-explored and potentially promising two-sided alternatives. Our paper also builds qualitatively on important work by Lang, Manove and Dickens (2005) who provide a model in which disparate outcomes can be sustained despite discriminatory employer preferences being arbitrarily weak.[9]

Finally, our proposed method to identify statistical discrimination by employers is related to a small but burgeoning literature on empirical tests for discrimination. Altonji and Pierret (2001) provide a classic test for discrimination in wage-setting based on the dynamics of employer learning and implied trajectories of black and white workers. Lang and Lehmann (2012) suggest an alternative test, which is based on the same data but is more robust to changing black-white relative productivity. Fryer, Pager and Spenkuch (2010) use a labor market search model to derive a conservative test for discrimination based on changes in the average wages of black and white workers who switch to new firms. Our contribution to this literature is to propose a test that would ensure that any finding of discrimination is robust to confounding variation in firm investment and worker belief formation.[10]

B. Strategic Complementarities

As mentioned in the Introduction, the terms *strategic complements* and *strategic substitutes* were first used by Bulow et al. (1985) who used the two concepts to shed light on results in oligopoly theory.[11] When two players' actions are strategic complements, each player's set of best responses weakly increases with the actions of the other. In our model, there is strategic complementarity between firm and worker investment: a higher level of firm (worker) investment raises the incentive for any worker (firm) to invest. This is the fundamental logic that underlies our ultimate policy prescription: two-sided investment insurance.

[8] There are also parallels between our model and others in which two sides invest before being matched. For example, Noldeke and Samuelson (2015) develop a model with simultaneous investment followed by *ex post* matching. Although information is complete in their context, the models share a complementarity in decisions that underlies inefficient equilibria.

[9] Another related paper is Filippin (2009). In that model, unequal opportunities between groups are self-fulfilling, but driven by incorrect minority worker beliefs. Minority workers, perhaps due to a history of poor treatment, believe incorrectly that employers are biased. This causes them to supply low levels of effort, which precludes them from being promoted. Since minority workers never provide high levels of effort, they never observe whether they would have been promoted.

[10] In addition to the papers cited here, alternative tests have been suggested in other contexts. For example, Knowles, Persico and Todd (2001) use search and success rates to test for racial discrimination in searches for contraband by police officers.

[11] In another early paper, Cooper and John (1988) use the concept of strategic complementarity to analyze macroeconomic coordination failures.

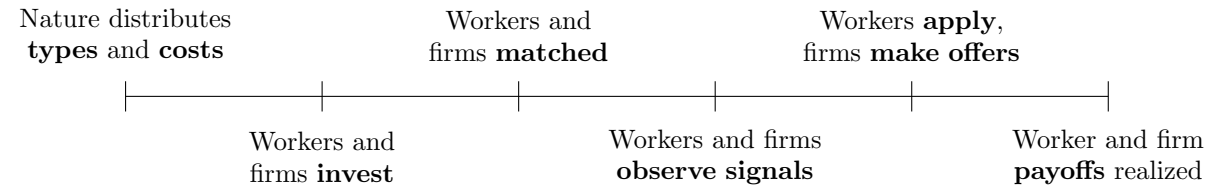


Figure 1: Sequence of Actions

Vives (1990) builds on these concepts and analyzes *supermodular* games, where all players' payoffs satisfy monotonicity properties that are closely related to strategic complementarity.[12] He demonstrates that such games have appealing properties. For example, the equilibrium set is always non-empty and equilibria can be pareto-ranked. Additionally, he describes robust stability properties of these games.

Despite strategic complementarity between worker and firm investment, the game we describe is not supermodular, and the results discussed by Vives (1990) do not apply. One can see this from the following observation. Fixing a level of firm investment δ, there is a set of possible worker investment levels $\pi^*(\delta)$ and corresponding employer hiring standards s^* that are mutually consistent. In general, raising δ raises some elements of $\pi^*(\delta)$ but *lowers* others: i.e., the correspondence $\pi^*(\delta)$ is not monotonic.

3 The Basic Model

A. Building Blocks

Imagine a large number of employers and a larger population of workers. Each employer is randomly matched with many workers from this population. Workers belong to one of two identifiable groups, $j \in \{A, B\}$. Denote by λ_A the fraction of As in the population and $\lambda_B = 1 - \lambda_A$ the fraction of Bs. One can imagine groups being race, gender, or any other protected class.

Nature moves first and assigns a type to each worker and a type to each employer. The worker's type, denoted by $c \in (0, \overline{c})$, $\overline{c} < \infty$, depicts her cost of investment in human capital. Let the fraction of workers with costs no greater than c be represented by $G^W(c)$ – a smooth and continuous cumulative distribution function – with $g^W(c)$ the associated density. Similarly, employers have the opportunity to invest at a cost $k_j \in \left(0, \overline{k}\right)$, $\overline{k} < \infty$, to make their workplaces desirable and productive places to work for workers of type j. The fraction of employers with investment cost no greater than k_j is $G^E(k_j)$, with $g^E(k_j)$ the associated density. Superscripts "W" and "E" refer to workers and employers, respectively.

Consistent with Lang (1986), we assume that firm investment costs are fixed. This is done for analytical

[12] Supermodular games were originally developed by Topkis (1979) but were first applied to economics by Vives (1990) and further analyzed by Milgrom and Roberts (1990).

simplicity and symmetry – and has no impact on our main results. Section 8.3.2 describes how the model changes if we assume that investment costs are proportional to the number of workers hired.

After observing their costs, workers (resp. employers) make a dichotomous investment decision, choosing to become "qualified" or "unqualified," with no in-between. For workers, "qualified" implies that they are productive. For firms, "qualified" implies they are establishments that are desirable for a given type of worker. Nature then distributes a signal to employers regarding each worker's investment decision and, simultaneously, a signal to workers regarding the employer's investment decision. Specifically, let $\theta \in [0,1]$ denote a noisy, but informative, signal to employers about whether or not a particular worker chose to invest. There is an associated smooth and continuous cumulative distribution function $F_i^W(\theta)$, and density function, $f_i^W(\theta)$, where $i \in \{q,u\}$. We assume that $\phi(\theta) \equiv \frac{f_u^W(\theta)}{f_q^W(\theta)}$ is non-increasing in θ (i.e., $f_i^W(\theta)$ satisfies the monotone likelihood ratio property).

The signal structure for employer investment is similar: nature distributes a noisy but informative signal $\psi \in [0,1]$ to workers about whether or not the employer chose to invest. There is an associated smooth and continuous cumulative distribution function, $F_i^E(\psi)$, and density function, $f_i^E(\psi)$, where $i \in \{q,u\}$. We assume that $\tau(\psi) \equiv \frac{f_u^E(\psi)}{f_q^E(\psi)}$ is non-increasing in ψ.

Next, workers observe the signal they receive from the employer and decide whether to "apply."[13] If they receive information that suggests that a given workplace will be a poor fit for their type, they may refrain from applying. Firms observe θ for all who apply and make a deterministic hiring decision: hire or reject. Production occurs and payoffs are received.

B. Payoffs

If the worker is hired and works for an employer who has made a group j investment, she receives a fixed payoff of $\omega_q - c$ if she chose to invest and ω_q if not. If the worker is hired and works for an employer who has *not* made a group j specific investment, she receives $-\omega_u - c$ if she invested and $-\omega_u$ if she did not. If she does not work for any employer, she receives $-c$ if she invested or zero otherwise.[14] We assume that both ω_q and ω_u are positive and exogenously determined. This, again, is purely for analytical convenience and ease of exposition in our baseline model and does not change our main results. We endogenize wages in two places: (1) to demonstrate robustness of our approach, we discuss policy when workers are paid either by ex-post bargaining or by their expected marginal product in section 8.3.1; and (2) we allow for continuous wage-setting when discussing our proposed empirical test for statistical discrimination in section 6.

[13] We view this model as a static approach to what is likely a dynamic process. In a dynamic version, a worker's choice to refrain from applying would reflect the option value of waiting for a better offer.

[14] A mathematically equivalent assumption (see section 8.2.1) is that the worker receives an unemployment payment $\bar{U}$ – or other outside option – if she does not apply. In this case, the key assumption is that workers prefer unemployment to being matched to an employer who has not invested. A third alternative is to imagine that application to a firm is costly.

The employer receives $\chi_q - k_j$, $\chi_q > 0$, if it hires a qualified worker and makes group j-specific investments, and χ_q if it hires a qualified worker and chooses not to invest in group j amenities. Similarly, the employer's payoffs are $-\chi_u - k_j$ where $-\chi_u < 0$ if it hires an unqualified worker and makes group j-specific investment, and $-\chi_u$ if it hires an unqualified worker and chooses not to invest in group j-specific amenities. If no worker is hired, the employer receives $-k_j$ if it invested and zero otherwise.

C. Strategies

The worker's strategy consists of a pair of functions – an investment decision and an application decision, which we write as $I^W : \{A, B\} \times [0, \bar{c}] \to [0, 1]$ and $A^W : \{A, B\} \times [0, 1] \times [0, 1] \times [0, \bar{c}] \to [0, 1]$. The employer's strategy also consists of a pair of functions – an investment decision and an assignment decision – $I^E : \{A, B\} \times \left[0, \bar{k}\right] \to [0, 1]$, $A^E : \{A, B\} \times [0, 1] \times [0, 1] \times \left[0, \bar{k}\right] \to [0, 1]$.

D. Expected Payoffs

Employer Offer Threshold

Let $\pi_j \in [0, 1]$ denote the employer's prior belief that a randomly drawn worker of group j is qualified. The expected payoff for the employer is a function of its beliefs, investment decisions, the signal it receives, and net payoffs. An employer does not intrinsically care about which type of worker it hires – save investment costs, which are sunk at the time that it makes an offer – but it may have different priors about the likelihoods that workers of different types are qualified.

Given the prior π_j and observed signal θ, the employer formulates the posterior probability, using Bayes' rule, that a worker of group j is qualified: $\kappa(\pi_j, \theta) = \frac{\pi_j f_q^W(\theta)}{\pi_j f_q^W(\theta) + (1-\pi_j) f_u^W(\theta)}$. The expected payoff to hiring a worker of group j can be written as: $\kappa(\pi_j, \theta)\chi_q - (1 - \kappa(\pi_j, \theta))\chi_u$. Recall that the payoff for not hiring a worker is 0. The condition that this expected payoff be positive defines a standard, which is a critical threshold in the signal θ such that the employer will choose to hire only if a worker's signal exceeds this threshold: $s_j^*(\pi_j) \equiv \min\left\{\theta \in [0, 1] | \frac{\chi_q}{\chi_u} > \left(\frac{1-\pi_j}{\pi_j}\right)\phi(\theta)\right\}$.

Worker Application Threshold

Let $\delta_j \in [0, 1]$ denote the prior belief that a worker of type $j \in \{A, B\}$ has that an employer made the investment relevant to her group. The worker's expected payoff is a function of her beliefs, investment decision, the signal she receives, and net payoffs. Given δ_j and observed signal ψ, workers calculate the posterior probability that a particular employer has invested, again using Bayes' rule: $\xi(\delta_j, \psi) = \frac{\delta_j f_q^E(\psi)}{\delta_j f_q^E(\psi) + (1-\delta_j) f_u^E(\psi)}$. The worker's expected payoff of applying can be written as: $\xi(\delta_j, \psi)\omega_q - (1 - \xi(\delta_j, \psi))\omega_u$. Thus, similar to before, the worker will only apply if the employer's signal exceeds the following threshold: $t_j^*(\delta_j) \equiv \min\left\{\psi \in [0, 1] | \frac{\omega_q}{\omega_u} > \left(\frac{1-\delta_j}{\delta_j}\right)\tau(\psi)\right\}$.

Investment Decisions

We begin with the worker. With probability $\delta_j \left(1 - F_q^E\left(t^*\left(\delta_j\right)\right)\right)$ a group j worker will be matched to an employer who made the group j investment and will apply to that employer because the signal she receives exceeds her application threshold. However, with probability $\left(1-\delta_j\right)\left(1 - F_u^E\left(t^*\left(\delta_j\right)\right)\right)$ she will apply to an employer who did *not* invest. In total, the worker's expected payoff from successfully obtaining a job is $\overline{\omega}\left(\delta_j\right) = \delta_j \left(1 - F_q^E\left(t^*\left(\delta_j\right)\right)\right)\omega_q - \left(1-\delta_j\right)\left(1 - F_u^E\left(t^*\left(\delta_j\right)\right)\right)\omega_u$.

Investing in human capital increases the likelihood that a worker is accepted by an employer. If a worker of type j invests, she gets expected gross payoff: $\left(1 - F_q^W\left(s_j\right)\right)\overline{\omega}\left(\delta_j\right)$. Conversely, if she does not invest, she gets $\left(1 - F_u^W\left(s_j\right)\right)\overline{\omega}\left(\delta_j\right)$. Thus, the net return on investment for workers can be written as:

$$\beta_W\left(s_j, \delta_j\right) \equiv \left[F_u^W\left(s_j\right) - F_q^W\left(s_j\right)\right]\overline{\omega}\left(\delta_j\right). \tag{1}$$

Now, consider the employer's investment decision. Similar to the worker, the employer's expected net payoff from hiring a worker is $\overline{\chi}\left(\pi_j\right) = \pi_j \left(1 - F_q^W\left(s_j^*\left(\pi_j\right)\right)\right)\chi_q - \left(1-\pi_j\right)\left(1 - F_u^W\left(s_j^*\left(\pi_j\right)\right)\right)\chi_u$. If the employer makes the group j investment, it gets gross payoff $\lambda_j[1 - F_q^E\left(t_j\right)]\overline{\chi}\left(\pi_j\right)$. If the employer does not invest, it gets $\lambda_j[1 - F_u^E\left(t_j\right)]\overline{\chi}\left(\pi_j\right)$. Recall that λ_j is the fraction of workers who are members of group j. Thus, the net return to investment for firms is:

$$\beta_E\left(t_j, \pi_j|\lambda_j\right) \equiv \lambda_j \left[F_u^E\left(t_j\right) - F_q^E\left(t_j\right)\right]\overline{\chi}\left(\pi_j\right). \tag{2}$$

E. Bayesian Nash Equilibrium

A pair of beliefs – one for employers and one for workers – will be self-confirming if, by choosing standards optimal for those beliefs, the actions of each group of agents induce the other to become qualified at exactly the rate posited by the initial beliefs. This intuition leads to the following definition of equilibrium:

Definition. *An equilibrium of the game is a pair of beliefs* $\{\pi, \delta\}$ *satisfying:*

$$\pi_j = G^W\left(\beta_W\left(s_j, \delta_j\right)\right) \tag{3}$$

$$\delta_j = G^E\left(\beta_E\left(t_j, \pi_j|\lambda_j\right)\right) \tag{4}$$

This definition of equilibrium implies that both employer and worker beliefs are confirmed in equilibrium vis-à-vis a self-confirming feedback loop. Fix δ, and suppose that an employer believes that a fraction π of workers are qualified. Expecting this, each worker calculates her net benefit of investment and invests if and only if her costs are less than the net benefit. In equilibrium, the fraction of workers who invest must be equal to the employer's beliefs π. Workers' beliefs about firms must also be self-confirming.

For any *fixed* δ, our model nests the seminal Coate and Loury (1993) model where the wage is $\overline{\omega}(\delta)$. Similarly, fixing π at some level induces a version of the same model in which roles are reversed: workers' pessimistic beliefs drive disparate outcomes. In summary, discrimination can be generated by either side in the two-sided model and is generically sustained by both (e.g., in Silicon Valley, employers may discount the skills of women, while at the same time potential female software engineers discount employers' proclamations about workplace flexibility, sexual harassment, and discrimination).

Beliefs in this model exhibit extensive complementarity. First, a belief that one side is more likely to invest *increases* the expected return to investment of the other side, since it strictly increases the likelihood (π or δ) of getting a positive payoff from a given match. The signal threshold, $s^*(\pi)$ or $t^*(\delta)$, used by the opposing side is also lowered. For example, a rise in the fraction of firms investing causes $t^*(\delta)$ to fall as workers become more optimistic and more willing to apply for jobs. Conversely, a rise in π causes $s^*(\pi)$ to fall as firms become more optimistic about workers and more willing to hire.

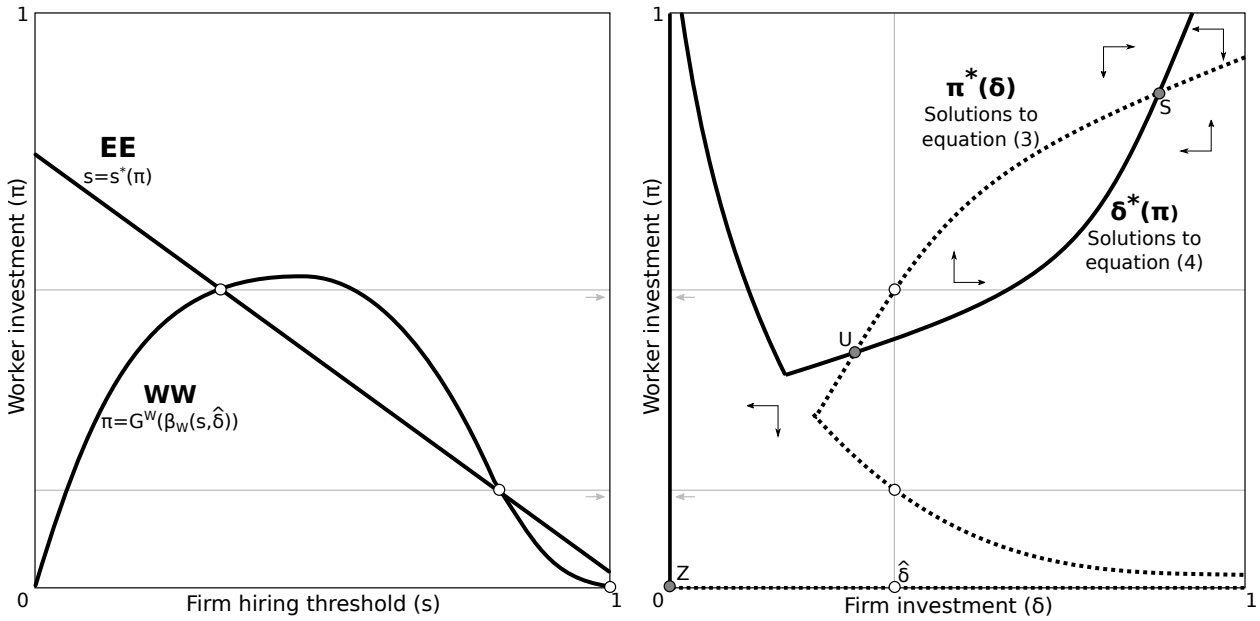
In our baseline model with fixed investment costs, there is an inherent disadvantage to being a numerical minority. If $\lambda_A = \lambda_B$ then the set of equilibria are fully symmetric. However, if $\lambda_A > \lambda_B$ then, for any given beliefs, firms have a strictly lower incentive to invest in amenities for Bs. At the extreme, as $\lambda_B \to 0$, no firms will ever be motivated to accommodate Bs. This intuition is straightforward to see from equation (2). For any fixed π and t, λ scales the return, while investment costs are independent of population size.[15]

A discriminatory equilibrium is one in which employers do not have homogeneous beliefs (e.g., $\pi_A > \pi_B$). This can occur whenever the system defined by (3) and (4) has multiple solutions, for then both workers and employers understand that workers of group A are more qualified than workers of group B and employers are less likely to make the workplace suitable for Bs than they are for As. One can imagine the familiar refrain: employers would be delighted to hire Bs but they are just not qualified. And Bs retort that they would invest if only they could trust that their efforts would be rewarded by employers. Discrimination, in this sense, is a classic coordination problem.

To understand the mechanics of equilibrium, it is instructive to consider fixing the level of firm investment at some $\hat{\delta}$. This fixes the worker application threshold at $t^*(\hat{\delta})$ and induces a model that is isomorphic to Coate and Loury (1993) with worker wage $\overline{\omega}(\hat{\delta})$. This is shown in Figure 2a. Holding fixed the level of firm investment, equilibrium is characterized by two graphs in $\{\pi, s\}$ space: an EE curve, which embodies the employer's hiring threshold; and a WW curve, which describes optimal worker investment as a function of that hiring threshold. The EE curve is downward sloping, since more optimistic firms set more generous (lower) thresholds. The WW curve is hump-shaped, reflecting the fact that there is little incentive for workers to invest if employers set very high or very low standards.

A high enough level of δ ensures the existence of at least two non-zero solutions to equation (3). At each

[15] In section 8.3.2, we present an extension in which costs are incurred only for workers of group j who apply and are hired.



(a) Solutions to equation (3) when $\delta = \hat{\delta}$

(b) Two-sided equilibria at Z, U and S

Figure 2: Equilibria in the two-sided model

of these solutions, employers' hiring standards are optimally set and every worker is making her investment decision optimally. The solutions for one arbitrary chosen $\hat{\delta}$ are shown as white dots in Figures 2a and 2b. As δ rises, the WW curve shifts upward, and the value of π that solves equation (3) rises if the EE curve crosses from above and falls if it crosses from below. Varying δ in this way traces out the solutions to equation (3) in $\{\pi, \delta\}$ space as shown by $\pi^*(\delta)$ in Figure 2b.

An entirely analogous thought exercise can be conducted for any fixed π, which induces a similar one-sided model in which employers invest and workers decide whether to apply. Varying π allows us to trace out the solutions to equation (4) as shown by $\delta^*(\pi)$ in Figure 2b.

A "zero investment equilibrium" always exists in the two-sided model, since $\pi = \delta = 0$ and $s = t = 1$ always satisfy the equilibrium equations. Yet, other equilibria may also exist. This should not be surprising given the strategic complementarity between workers and firms that we have described. For example, there are exactly three equilibria for the parameterization shown in Figure 2b: Z, U and S. The model generally has multiple solutions if workers and employers are responsive enough to each others' investments.

Proposition 1. *Let $\pi^*(\delta)$ and $\delta^*(\pi)$ be the sets of solutions to equations (3) and (4) respectively. Assume that $\phi(\theta)$ and $\tau(\psi)$ are continuous, strictly decreasing and strictly positive on $[0,1]$, and that $G^W(c)$ and $G^E(k)$ are continuous with full support on $[0,\overline{c}]$ and $[0,\overline{k}]$ with $G^W(0) = G^E(0) = 0$. Further assume that for some $\underline{\delta}$, there exists an s for which $G^W(\beta_W(s,\delta)) > \phi(s)/[\chi_q/\chi_u + \phi(s)]$. Similarly assume that for*

some $\underline{\pi}$*, there exists a* t *for which* $G^E\left(\beta_E\left(t,\pi|\lambda\right)\right) > \tau\left(t\right)/\left[\omega_q/\omega_u + \tau\left(t\right)\right]$*. Then non-zero elements of* $\pi^*\left(\delta\right)$ *and* $\delta^*\left(\pi\right)$ *exist for any* $\delta \geq \underline{\delta}$ *and* $\pi \geq \underline{\pi}$ *respectively. If there is a set of beliefs* $\{\pi, \delta\}$ *such that* $\delta \in \delta^*\left(\pi\right)$ *and* $\pi < \max\left\{\pi^*\left(\delta\right)\right\}$ *then there exist multiple solutions to the two-sided model.*

All technical proofs are presented in Section 8.1. To better understand the logic behind Proposition 1, see Figure 2b. First, we know that the assumptions about $\phi\left(\theta\right)$ and $\tau\left(\psi\right)$ guarantee that the solutions to each equation are bounded below one. Second, we assume that there exists some pair of beliefs $\{\pi, \delta\}$ such that $\delta \in \delta^*\left(\pi\right)$ and $\pi < \max\left\{\pi^*\left(\delta\right)\right\}$, a condition that must hold for large enough χ_q and ω_q. Combined with our regularity assumptions, this is enough to ensure that $\pi^*\left(\delta\right)$ and $\delta^*\left(\pi\right)$ intersect at multiple points with $\delta > 0$ and $\pi > 0$, which define equilibria with non-zero investment. This is the case in Figure 2b.

F. Dynamics

To analyze dynamics, we define a simple learning process that describes how employers and workers adjust their beliefs and actions in response to a shock.

$$\pi_{t+1} = G^W\left(\left[F_u^W\left(s^*\left(\pi_t\right)\right) - F_q^W\left(s^*\left(\pi_t\right)\right)\right] \cdot \left[\delta_t\left(1 - F_q^E\left(t^*\left(\delta_t\right)\right)\right)\omega_q - \left(1-\delta_t\right)\left(1 - F_u^E\left(t^*\left(\delta_t\right)\right)\right)\omega_u\right]\right)$$
$$\delta_{t+1} = G^E\left(\lambda\left[F_u^E\left(t^*\left(\delta_t\right)\right) - F_q^E\left(t^*\left(\delta_t\right)\right)\right] \cdot \left[\pi_t\left(1 - F_q^W\left(s^*\left(\pi_t\right)\right)\right)\chi_q - \left(1-\pi_t\right)\left(1 - F_u^W\left(s^*\left(\pi_t\right)\right)\right)\chi_u\right]\right)$$

This rule is backward-looking, with each generation of workers and firms choosing their actions based on the decisions of the preceding generation.[16]

We can use this learning process to analyze the robustness of equilibria to small errors of perception. Following Coate and Loury (1993), we consider an arbitrary but small perturbation to both firm and worker investments. Under the adjustment process above and the assumptions of our existence proposition, the zero investment equilibrium is always locally stable. To see this, observe that below some strictly positive δ, no worker applies because workers are too pessimistic about firm investment. Similarly, low enough π guarantees that no firms make offers. As long as δ and π remain below these critical values, there is no incentive for either party to invest.

To analyze stability more generally, we can linearize this two-dimensional system around the equilibrium. For ease of exposition, define the following derivatives.

$$WW_1' = G^{W\prime} \cdot \left[f_u^W\left(s^*\left(\pi\right)\right) - f_q^W\left(s^*\left(\pi\right)\right)\right] \qquad WW_2' = G^{E\prime} \cdot \left[f_u^E\left(t^*\left(\delta\right)\right) - f_q^E\left(t^*\left(\delta\right)\right)\right]$$
$$EE_1' = 1/s^{*\prime}\left(\pi\right) \qquad EE_2' = 1/t^{*\prime}\left(\delta\right)$$
$$RR_1' = \overline{\omega}'\left(\delta\right) \cdot \left[F_u^W\left(s^*\left(\pi\right)\right) - F_q^W\left(s^*\left(\pi\right)\right)\right] \cdot G^{W\prime} \qquad RR_2' = \overline{\chi}'\left(\pi\right) \cdot \lambda \cdot \left[F_u^E\left(t^*\left(\delta\right)\right) - F_q^E\left(t^*\left(\delta\right)\right)\right] \cdot G^{E\prime}$$

[16] Our results are qualitatively robust to generalizations along the lines of Kim and Loury (2012), which features overlapping generations who are forward-looking. Introducing this level of complexity in the dynamic adjustment process is beyond the scope of this paper.

Intuitively, WW_1' is the slope of the WW curve in Figure 2a and captures the impact of a less favorable (higher) firm signal threshold on worker incentives. Similarly, EE_1' is the slope of the EE curve and captures the effect of *lower* worker investment on the signal threshold that firms optimally set. The direct impact of higher firm investment on the worker's payoff from being hired is RR_1' and could be shown by an upward shift in the WW curve. The firm equivalents – WW_2', EE_2' and RR_2' – are analogous.

These definitions allow us to write the Jacobian of the linearized system compactly.

$$\begin{bmatrix} WW_1' \frac{1}{EE_1'} & RR_1' \\ RR_2' & WW_2' \frac{1}{EE_2'} \end{bmatrix}$$

The system is stable if both eigenvalues of this matrix have absolute values strictly less than one, and the following condition is necessary and sufficient for this (Neusser 2016).

$$\left| WW_1' \frac{1}{EE_1'} + WW_2' \frac{1}{EE_2'} \right| < 1 + \left(WW_1' \frac{1}{EE_1'} \cdot WW_2' \frac{1}{EE_2'} \right) - (RR_1' \cdot RR_2') < 2.$$

Since both eigenvalues strictly less than one guarantees that the equilibrium is hyperbolic, this is also sufficient for the non-linear system to be locally asymptotically stable.

To understand the stability condition, consider two special cases. First, suppose that worker and firm signal thresholds are locally unresponsive to changes in investment: i.e., the two thresholds s^* and t^* are approximately fixed. In the limit, this implies that $1/EE_1' \to 0$ and $1/EE_2' \to 0$, which causes the condition for stability to collapse to: $-1 < RR_1' \cdot RR_2' < 1$. This condition is intuitive. If worker and firm payoffs $\overline{\omega}$ and $\overline{\chi}$ change too sharply with each others' investments, then a small perturbation causes a reinforcing dynamic through beliefs, which moves the system further away from the original equilibrium.

It is also instructive to consider another extreme in which worker investment responds very strongly to the firm signal threshold s^*, which in turn is highly responsive to worker investment. This implies that $|WW'/EE'| > 1$, violating the first inequality of the stability condition. Instability arises in this case because a small perturbation to worker investment is compounded through changes in firms' hiring thresholds.

3.1 An Example with Uniform Cost and Signal Distributions

To further fix ideas, we now introduce a simple example to provide intuition for the model. Let costs for workers and firms be distributed uniformly on $[0, 1]$ so that $G^W(c) = c$ and $G^E(k) = k$.

Worker signals are also uniformly distributed, but with the support depending on the investment decision. A qualified worker's signal is distributed uniformly on $[\theta_q, 1]$, while an unqualified worker's signal is distributed uniformly on $[0, \theta_u]$ with $\theta_q < \theta_u$. Thus, a worker is surely qualified if $\theta > \theta_u$, surely unqualified if $\theta < \theta_q$, and there is a constant likelihood ratio $\hat{\phi} = \frac{1-\theta_q}{\theta_u}$ for $\theta \in [\theta_q, \theta_u]$.

We make analogous assumptions for firms. The signal sent by a firm that invested is uniformly distributed on $[\psi_q, 1]$, while that sent by a firm that did not invest is distributed uniformly on $[0, \psi_u]$ with $\psi_q < \psi_u$. Thus, a firm certainly invested if $\psi > \psi_u$, certainly did not invest if $\psi < \psi_q$, and there is a constant likelihood ratio $\hat{\tau} = \frac{1-\psi_q}{\psi_u}$ for $\psi \in [\psi_q, \psi_u]$.

The employer will always reject workers with clear fail signals and always accept those with clear pass signals. However, employers will make an offer to unclear workers if and only if they are "optimistic" enough in the sense that their prior π_j is greater than a fixed threshold $\hat{\pi}_j$. In symbols:

$$\pi_j \geq \left(\frac{\hat{\phi}}{\chi_q/\chi_u + \hat{\phi}} \right) = \hat{\pi}_j$$

Thus, the employer will set the hiring threshold at either $s^* = \theta_q$ or $s^* = \theta_u$. Similarly, workers will always apply to firms with clear pass signals, never to firms with clear fail signals, and will apply to firms with unclear signals if and only if they are optimistic enough about firms: $\delta_j \geq \hat{\delta}_j$. The worker will therefore set the threshold at either $t^* = \psi_q$ or $t^* = \psi_u$.

To make this example especially simple, we adopt parameter values that make firms and workers symmetric. Specifically, let $\theta_q = \psi_q = \frac{1}{3}$, $\theta_u = \psi_u = \frac{2}{3}$, $\omega_q = 3$, $\omega_u = 1$, $\chi_q = 6$, $\chi_u = 2$ and $\lambda = 0.5$. With these functional form assumptions and parameter values, the returns to investment for workers (β_W) and employers (β_E) are piecewise-linear functions.[17]

$$\beta_W = \begin{cases} \frac{7}{4}\delta - \frac{1}{4} & \text{if } \delta \geq \hat{\delta} \\ \frac{3}{4}\delta & \text{if } \delta < \hat{\delta} \end{cases} \qquad \beta_E = \begin{cases} \frac{7}{4}\pi - \frac{1}{4} & \text{if } \pi \geq \hat{\pi} \\ \frac{3}{4}\pi & \text{if } \pi < \hat{\pi} \end{cases}$$

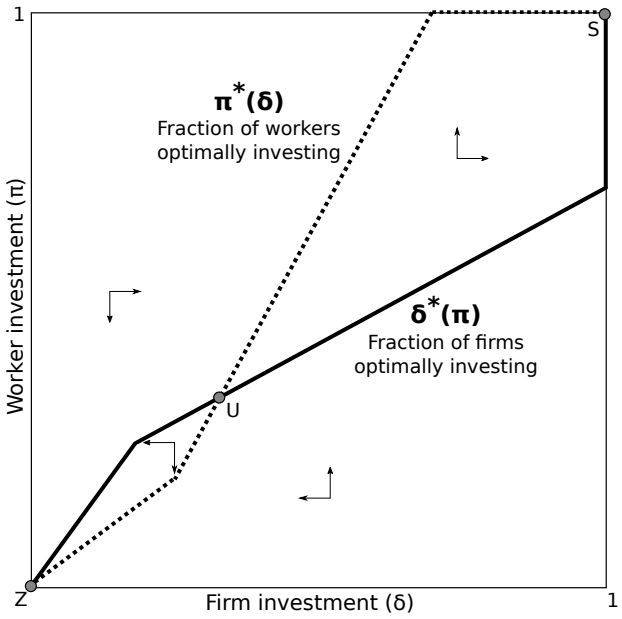
The fraction of workers who invest is $\pi = \min\{\max\{G_W, 0\}, 1\}$ and the fraction of firms who invest is $\delta = \min\{\max\{G_E, 0\}, 1\}$.

The equilibria in this example are shown in Figure 3, which is the equivalent of Figure 2b for this example. For any fixed belief π about the fraction of workers who are investing, the solid line, $\delta^*(\pi)$, shows the fraction of firms who optimally invest. Similarly, the dotted line, $\pi^*(\delta)$, shows the fraction of workers who invest for a given belief δ about the fraction of firms who are investing. There are three equilibria, shown as S, U and Z. Only at these points are the actions and beliefs of firms and workers mutually consistent.

Also evident in Figure 3 are the effects of changing signal thresholds. As firms become more optimistic (higher π), there are two effects. First, favorable beliefs about workers directly raise the return to firm investment, since the expected payoff from a match is higher. Secondly, firms eventually become so optimistic that they accept workers with 'unclear' test scores. At this point, firm returns become more sensitive to

[17] The unparameterized returns are derived in Appendix 8.2.2.

Figure 3: Equilibria in the Clear / Unclear Example



worker investment, since a match is more likely when workers are given the benefit of the doubt. This is reflected in a change in the slope of $\delta^*(\pi)$. The explanation of the shape of $\pi^*(\delta)$ is analogous.

This example also permits a transparent discussion of dynamics, since it corresponds to the special case in which signal thresholds are locally unresponsive to beliefs (see Section 2). The phase arrows in Figure 3 show the direction of adjustment in any given region. With the assumptions of this example, an equilibrium is stable if and only if the solid line is flatter than the dotted line: i.e., only the two extreme equilibria, S and Z, are stable. To understand the instability of the equilibrium at U, imagine a small upward perturbation to both π and δ such that the economy is at a point above the dotted line and below the solid line. From here, the fraction of firms and workers investing both increase further, moving us *further away* from U.

Both Proposition 1 and this simple example demonstrate that multiple equilibria can occur in our model. In the next two sections, we explore policies that can help minimize the set of "bad" equilibria (i.e. those with low investment and pessimistic beliefs on both sides).

4 Extending the Basic Model: One-Sided Policies

We now consider how the government or another third party intermediary might intervene with some policy to break equilibria that disadvantage some groups relative to others. We are particularly interested in a policy that: (i) eliminates equilibria without homogeneous beliefs; (ii) never harms its recipients; and (iii) achieves equality as quickly as possible. Given its historical and current prominence across the world – and

the controversy that typically ensues – we begin with affirmative action.

Affirmative Action – Executive Order # 11246, signed by Lyndon B. Johnson – has been promulgated around the world from Malaysia to South Africa to Lebanon. Affirmative action policies generally entail the preferential treatment of persons who possess certain social traits based on a presumption that, on average, individuals of those traits are less effective in the competition for scarce resources because of some social or historical handicap.

The simplest affirmative action policy insists that employers make color-blind assignment – requiring that As and Bs with identical "signals" be treated equally. Unfortunately, this policy can only be enforced if – in every instance – a regulator can observe and verify all information upon which employers rely when making hiring decisions. We assume this type of extreme informational requirement – essentially requiring government regulators to sit in all interviews – is impractical. Instead, we explore two potential definitions of affirmative action – equality in offers and equality in employment. We begin with the former.

A. Equality in Offers

In statistical discrimination models without strategic complementarity between worker and firm investment, affirmative action can be quite successful. In fact, affirmative action rules out the existence of any equilibrium with zero investment by workers of one group but positive investment by members of the other. Furthermore, firms can satisfy the affirmative action requirement by setting $s_A = s_B$, which achieves full equality in investment rates and correspondingly homogeneous beliefs: i.e., $\pi_A = \pi_B$. Nonetheless, as Coate and Loury (1993) make clear, affirmative action may not lead to homogeneous beliefs if more generous hiring standards undermine investment and become demotivating. As we will show, affirmative action is even more problematic in a model with strategic complementarity.

The behavior of workers is not directly affected by affirmative action. They continue to make their decisions as before. Affirmative action changes an employer's problem, however, because standards and investments can no longer be chosen independently for the two groups.

Consider a group of workers about which an employer believes a fraction π are qualified and for which it uses assignment standard s. For each group, let $\rho(s_j, \pi_j) \equiv \pi_j \left[1 - F_q^W(s_j)\right] + (1 - \pi_j)\left[1 - F_u^W(s_j)\right]$ be the probability that the employer assigns to making an offer to a randomly drawn worker, and let $P(s_j, \pi_j, i_j)$ denote the expected payoff from hiring such a worker, where $i_j \in \{q, u\}$ captures the firm investment decision. In symbols: $P(s_j, \pi_j, i_j) = \pi_j \left[1 - F_q^W(s_j)\right]\left[1 - F_{i_j}^E(t_j)\right]\chi_q - (1 - \pi_j)\left[1 - F_u^W(s_j)\right]\left[1 - F_{i_j}^E(t_j)\right]\chi_u$.

In the modified game, each employer must ensure that, whatever standards it uses, anticipated hiring rates for each group are equal: i.e., $\rho(s_A, \pi_A) = \rho(s_B, \pi_B)$. Given beliefs (π_A, π_B) and worker application standards (t_A, t_B), it will choose hiring standards (s_A, s_B) and make an investment decision $i_j \in \{q, u\}$ for

each group $j \in \{A, B\}$ to solve the following optimization problem:

$$\max_{s_A, s_B, i_A, i_B} \left[\lambda_B P\left(s_B, \pi_B, i_B\right) + \lambda_A P\left(s_A, \pi_A, i_A\right)\right] \quad \text{s.t.} \quad \rho\left(s_B, \pi_B\right) = \rho\left(s_A, \pi_A\right). \tag{5}$$

Namely, an employer's best response under affirmative action regulations is to choose standards and make investment decisions that maximize its expected payoff, subject to the affirmative action constraint. This suggests the following definition of equilibrium under affirmative action.

Definition. *An equilibrium under affirmative action is a set of beliefs* (π_A, π_B), (δ_A, δ_B), *worker standards* (t_A, t_B) *and employer standards* (s_A, s_B) *satisfying the following conditions:*

(a) Firm signal thresholds (s_A, s_B) *solve problem (5), given* (π_A, π_B, t_A, t_B).[18]

(b) $t_j = t_j^*\left(\delta_j\right)$, $j \in \{A, B\}$

(c) $\pi_j = G^W\left(\beta_W\left(s_j, \delta_j\right)\right)$, $j \in \{A, B\}$

(d) $\delta_j = G^E\left(\lambda_j\left[F_u^E\left(t_j\right) - F_q^E\left(t_j\right)\right]\left[\pi_j\left(1 - F_q^W\left(s_j\right)\right)\chi_q - \left(1 - \pi_j\right)\left(1 - F_u^W\left(s_j\right)\right)\chi_u\right]\right)$, $j \in \{A, B\}$

The only requirement that affirmative action adds is the constraint that $\rho\left(s_B, \pi_B\right) = \rho\left(s_A, \pi_A\right)$. Without this requirement, we obtain the unconstrained version of (6), which is solved by $s_A = s^*\left(\pi_A\right)$, $s_B = s^*\left(\pi_B\right)$ and firm investment rate $\delta_j = G^E\left(\beta_E\left(t_j, \pi_j | \lambda_j\right)\right)$. It is also clear that if an equilibrium with homogeneous beliefs exists *without* affirmative action, then an equilibrium with the same beliefs exists *with* the constraint. This follows directly from the fact that the affirmative action constraint is non-binding in any equilibrium in which employers have homogeneous beliefs ($\pi_A = \pi_B$).

However, unlike Coate and Loury (1993), it is generally impossible to guarantee that homogeneous beliefs will prevail because an equilibrium with zero B investment but positive A investment *always* satisfies the affirmative action constraint. We formalize this result in Proposition 2, but the intuition is simple: if Bs do not apply, firms are not punished for not hiring.

Proposition 2. *Assume that, without affirmative action, there exists an equilibrium with positive investment. Then there exists an equilibrium under affirmative action without homogeneous beliefs.*

A second problem is that there is generally no equilibrium with positive investment and homogeneous employer beliefs unless unless $\lambda_A = \lambda_B$. The reason for this is that firm investment returns are lower for smaller groups. More formally, suppose that $\pi_A = \pi_B = \pi$. The affirmative action constraint is then satisfied with equal firm hiring standards $s_A = s_B = s^*\left(\pi\right)$. Worker beliefs cannot be homogeneous in this case since

[18] We can limit our analysis to solutions in which all firms set the same signal thresholds since any one value of $\rho\left(s_j, \pi_j\right)$ can only be achieved with two different thresholds s^x and s^y if $F_q^W\left(s^x\right) = F_q^W\left(s^y\right)$, in which case the mass of type j individuals who receive offers is identical under the two thresholds.

firm investment incentives are strictly lower for the minority for any given firm beliefs π. However, strictly lower firm investment rates for Bs ($\delta_B < \delta_A$) combined with equal hiring thresholds must lead to lower investment returns for B workers ($\pi_B < \pi_A$). This is a contradiction, leading to the following result.

Proposition 3. *Assume that $\phi(\theta)$ and $\tau(\psi)$ are continuous, strictly decreasing and strictly positive on $[0,1]$. Further assume that $\lambda_A \neq \lambda_B$ and that $G^E(k)$ and $G^W(c)$ are strictly increasing. Then no equilibrium with positive investment and homogeneous employer beliefs exists (with or without affirmative action).*

A final drawback of this type of affirmative action in our model is that it can make outcomes for B workers strictly worse than under the status quo. Following the intuitive learning process we described in Section 3, suppose that employer and worker beliefs are fixed in the short run. At these fixed beliefs, the imposition of an affirmative action constraint can cause firm investment returns for Bs to become *negative*, ensuring that no firm invests and thus no B workers apply. This triggers reversion to zero investment by firms, and ultimately also workers ($\delta_B = \pi_B = 0$). We formalize this result in Proposition 4.

Proposition 4. *Assume that $\phi(\theta)$ and $\tau(\psi)$ are continuous, strictly decreasing and strictly positive on $[0,1]$. Further suppose that the A and B markets start with $\pi_A > \pi_B > 0$ and $\delta_A > \delta_B > 0$. For fixed beliefs $\{\pi_A, \pi_B, \delta_A, \delta_B\}$ and low enough δ_B and π_B, imposing affirmative action causes zero firms to invest in B amenities and zero B workers to invest.*

The intuition here is simple. Without affirmative action, firms were already hiring the few minority workers who they expected to be qualified. Affirmative action forces them to hire a potentially large number of additional minority workers if they apply, and these additional workers are expected to generate a loss for the firm on average. As a result, the few employers who were making investments to garner additional applications from minority workers now have less incentive to do so. If this effect is strong enough, they may even have an incentive to actively deter such applications. In summary, the affirmative action requirement may hurt minority workers' interests by undermining employers' efforts to attract them.

The mechanisms behind the failure of affirmative action here are substantively different from one-sided models and apply far more generally. For example, in Coate and Loury (1993), affirmative action eliminates equilibria with zero investment by one group but not the other, and equilibria with homogeneous beliefs always exist. The problem that arises is that homogeneous beliefs may not obtain if the parameters allow more generous employer hiring standards to be sufficiently de-motivating. Instead, there may be a solution to the one-sided equivalent of problem (5) that features lower standards but also lower investment by the minority group. Since our model nests Coate and Loury (1993), this is also a concern here. However, affirmative action in the two-sided model is additionally complicated by worker belief formation and the

inherent disadvantage that minorities face because employers have less incentive to adapt their workplaces to accommodate smaller groups.[19] Affirmative action may even backfire in our model by causing employers to scale back their efforts to attract minorities.

B. Equality in Employment

We now consider employment quotas, which require that members of groups A and B are hired in proportion to their population sizes.[20] This articulation of affirmative action may be closer to the original spirit of early affirmative action (Revised Philadelphia Plan 1969). Under this type of constraint, employers cannot use the excuse that they would like to hire minorities but are not receiving applications. However, employment quotas can trigger a severe version of *patronization* in which aggressive hiring by employers undermines minority workers' incentive to invest.

Let $\rho_H(s_j, \pi_j, i_j) = \pi_j \left[1 - F_q^W(s_j)\right] \left[1 - F_{i_j}^E(t_j)\right] + (1 - \pi_j) \left[1 - F_u^W(s_j)\right] \left[1 - F_{i_j}^E(t_j)\right]$ be the probability the employer assigns to *hiring* a randomly drawn group j worker, where $i_j \in \{q, u\}$. The employment quota requires that $\rho_H(s_A, \pi_A, i_A) = \rho_H(s_B, \pi_B, i_B)$. Thus, given beliefs (π_A, π_B) and worker application standards (t_A, t_B), an employer will again choose hiring standards (s_A, s_B) and make investment decisions (i_A, i_B) to solve the following problem:

$$\max_{s_A, s_B, i_a, i_B} \left[\lambda_B P(s_B, \pi_B, i_B) + \lambda_A P(s_A, \pi_A, i_A)\right] \quad \text{s.t.} \quad \rho_H(s_B, \pi_B, i_B) = \rho_H(s_A, \pi_A, i_A). \tag{6}$$

Note that this is identical to problem (5) except that $\rho(s_j, \pi_j)$ has been replaced with $\rho_H(s_j, \pi_j, i_j)$.

In the case of employment quotas, employers may set different standards depending on which investments they made, since the investments affect the ability of an employer to attract workers. We therefore use $s_j^{i_A, i_B}$ to denote the hiring threshold set for group j by a firm that made investment decisions i_A and i_B. Firms' investment decisions will now also be related across groups, with the critical cost threshold for a firm to invest in one group (k_j^*) depending on its cost of investment for the other groups (k_{-j}). As a result, the return expected by a worker, which we denote by $\overline{\beta}_W$, will be a more complicated function of firm costs and all four hiring thresholds for her group. This suggests the following definition of equilibrium under an employment quota.

Definition. *An equilibrium under an employment quota is a set of beliefs* (π_A, π_B), (δ_A, δ_B), *worker standards* (t_A, t_B) *and employer standards* $\left(s_j^{q,q}, s_j^{q,u}, s_j^{u,q}, s_j^{u,u}\right)$, $j \in \{A, B\}$ *satisfying the following conditions:*

(a) Each firm's investment decisions (i_A, i_B) *and thresholds* (s_A, s_B) *solve (6), given* (π_A, π_B, t_A, t_B)

[19] Our result that members of smaller minorities are worse off aligns with the predictions of search models (e.g., Black, 1995), although Becker's analysis of taste-based discrimination with perfect sorting of workers across firms predicts the opposite.

[20] We are grateful to Lawrence Katz for suggesting this exercise.

(b) $t_j = t_j^*(\delta_j)$, $j \in \{A, B\}$

(c) $\pi_j = G^W\left(\overline{\beta}_W\right)$, $j \in \{A, B\}$

(d) $\delta_j = \int_0^1 G^E\left(k_j^*(k_{-j})\right) dk_{-j}$

An advantage of an employment quota over a regulation that simply requires equality in offers is that an employment quota obviously eliminates the possibility of an equilibrium with zero investment by B workers but positive investment by A workers. It may even eliminate all discriminatory equilibria. For example, any equilibrium under an employment quota must entail homogeneous beliefs if two conditions hold: (i) the worker signal of firm investment is very informative; and (ii) the employer signal of worker investment is very uninformative. We formalize this in Proposition 5.

Proposition 5. *Assume that G^E has full support on $[0, \overline{c}]$ with $\overline{c} > \omega_q$, let $\phi(\theta)$ be strictly decreasing, and define $\tilde{s}$ as the firm signal threshold such that $\phi(\tilde{s}) = 1$. If firm investment is close enough to perfectly observable, any equilibrium under an employment quota must entail homogeneous beliefs if:*

$$\eta\left(\overline{\beta}(s)\right) < \frac{\phi(s_j)}{\phi(s_j) - 1}$$

for all $s \in [0, \tilde{s})$ where $\eta(c) = \frac{d[c \cdot G(c)]}{dc}$ and $\overline{\beta}(s) = \left[F_u^W(s) - F_q^W(s)\right]\omega_q$.

Fixing employer investment decisions, the inequality in Propositon 5 guarantees that no two levels of worker investment are consistent with the same probability of being hired. It is always satisfied if $\phi(0) = f_u^W(0)/f_q^W(0)$ is small enough, which implies that the employer signal of worker productivity is relatively uninformative. Next, near-perfect observability of employer investment ensures that firms will not be able to satisfy the employment quota unless they make *both* or *neither* of the investments. Thus, $\delta_A \approx \delta_B$. Combined, these two assumptions ensure that an employment quota eliminates any possibility of discrimination in equilibrium.

The result above is subject to an important caveat: even if homogeneous beliefs are achieved, this need not improve the outcomes of any individual. The policy may harm the majority rather than helping the minority, and can worsen outcomes for both groups. This is easiest to see from an extreme example in which $\pi_B = \delta_B = 0$ initially, implying that no minority workers apply. Holding beliefs fixed, the only way for a firm to satisfy an employment quota is to hire zero workers of type A. This is in stark contrast to Coate and Loury (1993), where firms can satisfy the quota by simply lowering the minority hiring standard.

The same logic applies more generally. Intuitively, employers are constrained in their ability to attract minority workers who are themselves pessimistic about firms. They are therefore forced by the employment

quota to aggressively lower their standards, which can severely undermine worker investment incentives. We formalize these intuitions in Proposition 6.

Proposition 6. *Assume that $\phi(\theta)$ and $\tau(\psi)$ are continuous, strictly decreasing and strictly positive on $[0,1]$. Further suppose that the A and B markets start with $\pi_A > \pi_B > 0$ and $\delta_A > \delta_B > 0$. For low enough δ_B and π_B, imposing an employment quota lowers employment of A workers. Furthermore, there exists an open set of parameters such that the policy leads to zero investment by B workers.*

C. Wage and Employment Subsidies

Another policy proposal put forward in the literature is to subsidize worker wages or employment. Indeed, wage subsidies are highlighted as particularly effective by Coate and Loury (1993). However, not only do these policies fail to eliminate zero investment as an equilibrium in our model, but both can actually be harmful. To see why, suppose a wage subsidy s is introduced, raising a worker's positive and negative payoffs to $s+\omega_q$ and $s-\omega_u$ respectively. A worker will now apply if and only if:

$$\xi(\delta_j,\psi)(s+\omega_q)+(1-\xi(\delta_j,\psi))(s-\omega_u)>0.$$

This subsidy lowers the worker's application threshold, which can undermine firms' incentive to invest. While this intuition is general, it can again be seen most clearly from the extreme: if $s \geq \omega_u$, the worker will *always* apply and zero firms will invest. Depending on the subsidy chosen, this policy therefore has the potential to harm its intended beneficiaries. The basic intuition is that the effect of a wage subsidy on worker application behavior can reduce the impact of firm investment on the number of workers that it attracts, lowering firms' return on investment. An analogous problem occurs if an employment subsidy is provided to firms, which raises both χ_q and χ_u.[21]

D. An "Impossibility" Result

Given the failure of the specific policies we have considered thus far, we now search more systematically for a simple and reliable way to rule out equilibria with zero investment, and to quickly eliminate discrimination without potential for unintended harm. Since there are two decisions for workers and two for employers, we ask – abstractly at first – which of these margins one should target with policy. The answer is surprisingly definitive: policy should simultaneously target the investment decisions of both workers and firms.

[21] An employment subsidy suffers from this problem in one-sided models but wage subsidies do not (see Coate & Loury, 1993).

Proposition 7. *Suppose that we seek to move to an equilibrium* $\{s^*, t^*, \pi^*, \delta^*\}$ *from another point with* $s_0 > s^*$, $t_0 > t^*$, $\pi_0 < \pi^*$ *by independently setting some combination* $\mathcal{C}$ *of* s, t, π *and* δ. *There exist interventions that achieve this aim for any* $\{\pi_0, \delta_0\}$ *if and only if* $\{\delta, \pi\} \in \mathcal{C}$, $\{t, \pi\} \in \mathcal{C}$ *or* $\{s, \delta\} \in \mathcal{C}$. *Targeting* $\{\delta, \pi\}$ *is faster than any alternative.*

An immediate implication of Proposition 7 is that policies that only affect one decision margin will fail to achieve the goals we have set forth. This includes not just affirmative action but also many policies that we have not considered explicitly. The result also provides some guidance for where to look for policy solution, which is a problem we take up in Section 5.

Suggestive Evidence of the Efficacy of Two-Sided Interventions

Empirical evidence suggests that two-sided policies are indeed more effective than their one-sided equivalents. We focus on job training programs, of which Job Corps is a canonical example. It is a residential program funded by the Department of Labor but operated mostly by private contractors, and typically lasts around eight months. Participants receive vocational and academic training, counseling, social skills training, health education and job search assistance. There is some limited input from business to incorporate specific proficiencies but little involvement of employers. A large-scale randomized evaluation suggested that Job Corps increased earnings by around four percent, one year after the program, but with little long term impact and no effect on hourly wages (Schochet, Burghardt and McConnell 2008).

Conversely, *WorkAdvance* programs are narrowly targeted and employers are deeply involved in designing the training. Participants are strongly encouraged to participate in work-based learning with an employer who offers a job, good benefits and the possibility of career advancement. Randomized evaluation suggests that *WorkAdvance* programs increase earnings by 14 percent on average and the effect is mostly driven by higher wages (Hendra, Greenberg, Hamilton, et al. 2016).

The success of these programs stands in stark contrast to the average job training program, a result that aligns with the predictions of our model. A trainee who participates *WorkAdvance* has an incentive to invest and gain the skills being offered: not only does this investment lead to being hired, but the worker knows she will be rewarded with a position with an employer who is offering a real opportunity. At the same time, employers can trust that they will receive workers who have both the cognitive and non-cognitive skills that they need to be productive. While we cannot prove that these are the reasons why these programs succeed, there does seem to be a pattern in which programs are most successful (Hossain and Bloom, 2015).

5 Extending the Basic Model: Two-Sided Policies

Building on the impossibility result we presented in Proposition 7, the next section formally considers interventions that target both sides of the coordination problem faced by workers and firms in our model.

A. Two-Sided Investment Insurance

We begin with a new policy: two-sided investment insurance. Specifically, we suppose that the government has access to informative (but possibly imperfect) signals of worker and firm investment, and that it offers incentive payments conditional on these signals. This solution is very effective in our model. We also believe some approximation is likely to be implementable in reality, given policymakers have access to increasingly rich administrative data that could be used to measure both worker qualification and firm investment.

Proposition 8. *Suppose that the government observes noisy but informative signals, θ^g and ψ^g, of worker and firm investment respectively. For any initial beliefs, there exist incentive payments ω^g and χ^g conditional on these signals that immediately ensure that $\pi_A = \pi_B$, $\delta_A = \delta_B$, $s_A = s_B$ and $t_A = t_B$. If and only if $\lambda_A \neq \lambda_B$, a non-zero permanent investment subsidy is required to maintain $\pi_A = \pi_B$.*

An especially attractive version of the worker subsidy is feasible if the government observes the same signal as firms: i.e., $\theta^g = \theta$. In this case, it can set $s^g = s_A$ and condition the worker payment on rejection by a firm, thereby insuring workers against the possibility that employers are discriminatory. The advantage of this policy is that no worker payments are made by the government, rendering the worker intervention costless. Intuitively, for a worker to receive a government payment, she would have to be rejected by an employer and then be "hired" by the government. But this is impossible if the employer and government signals and signal thresholds are identical. Essentially, a non-discriminatory government's use of its "market power" in standard setting achieves equality.

If θ^g is a noisy approximation of θ, the intuition is similar. In this case, we can characterize investment incentives using the conditional distribution of the government signal θ^g. The probability of rejection by the government is $\tilde{F}_i^W\left(s_B^g | \theta < s_B\right)$. If the government sets $s_B^g = s_A$, the fraction of B workers who invest is:

$$\pi_{B,t} = G^W\left(\beta_W\left(s^*\left(\pi_{B,t-1}\right), \delta_{B,t-1}\right) + \left[\tilde{F}_u^W\left(s_A | \theta < s^*\left(\pi_{B,t-1}\right)\right) - \tilde{F}_q^W\left(s_A | \theta < s^*\left(\pi_{B,t-1}\right)\right)\right]\omega^g\right).$$

Just like the policy described in Proposition 8, there is an incentive payment that ensures that $\pi_{B,t} = \pi_A$ for any π_A. Since this is achieved immediately, the actual cost of the worker payments are as follows.

$$\delta\left[1 - \tilde{F}_q\left(s_A | \theta < s_A\right)\right]\omega^g + (1-\delta)\left[1 - \tilde{F}_u\left(s_A | \theta < s_A\right)\right]\omega^g$$

This expected cost clearly shrinks to zero if the government signal is approximately identical to that of the firm. A small amount of additional noise adds to the cost.

In summary, two-sided investment insurance achieves all of our stated goals. It eliminates equilibria

with discriminatory beliefs weakly faster than any other policy, with no potential for negative side effects.[22] Some elements of investment insurance may already be provided via state subsidization of merit-based scholarships, systematic efforts to eliminate discrimination and harassment, and – for women – policies that promote workplace flexibility. But, no policy to date has simply guaranteed a market wage for those who it believes have invested.

Year Up – an organization that offers disadvantaged youth a combination of skills training and a six month internship with corporate partners such as JPMorgan, State Street or Google – is an example of an ambitious two-sided investment insurance program.[23] Unlike a traditional job training program, *Year Up* training is targeted to an industry or even a specific employer. Employers are actively involved in its operation; for instance, some design case studies and conduct mock interviews or customer interactions. Upon completing the program satisfactorily, participants are then rewarded with a well-paid internship with genuine opportunities for career advancement and ongoing support from *Year Up*. A key difference between *Year Up* and other training programs is that they also train employers on how to best deal with minority youth (e.g. require firm investment). They also guarantee an internship for every student: if a student successfully completes the program but does not obtain an internship at the end of their six month training period, *Year Up* itself will hire them.

This model has been remarkably successful, with a randomized evaluation indicating that treated individuals had 30 percent higher earnings over two years, and that this was mostly driven by higher wages (Roder and Elliot 2014).

B. Affirmative Action

Proposition 7 suggests that an alternative policy is to simultaneously target both the investment and hiring decisions of firms. One way to implement this is to combine affirmative action on the 'extensive' and 'intensive' margins. Essentially, firms would be encouraged to invest in amenities for the minority group (i.e., one-sided investment insurance) *and* change their hiring practices (i.e., affirmative action). If the gap between A and B workers is small, one-sided investment insurance alone can be effective, but it would have to be accompanied by affirmative action if B workers begin with zero investment. Intuitively, affirmative action ensures that at least some minorities are hired, and investment incentives ensure that workplaces are attractive for minority workers so that they will apply.

Note 1. A one-sided investment subsidy can eliminate discrimination after a one period delay if: (i) $s_B < 1$; (ii) $s_B \approx s_A$; and (iii) δ_A is small. If $s_B = 1$ but the remaining conditions hold, a one-sided subsidy can eliminate discrimination after one period if combined with affirmative action (equality in offers).

[22] A practical implementation of investment insurance would need to determine details such as which signals of investment should be subsidized, but this is beyond the scope of this paper.

[23] For more information, see `www.yearup.org`

A larger gap between groups limits the effectiveness of a one-sided investment subsidy. Even if combined with affirmative action, it may be impossible to achieve equality in finite time. There are two reasons for this. First, to be successful, the subsidy needs to raise firm investment in the B amenity to *above* the rate of investment in the A amenity. There is little scope to do this if nearly all firms are already making the A investment. Second, a harsh hiring standard for B workers limits the impact on worker returns of higher firm investment rates. These concerns imply that "two-sided affirmative action" is unlikely to be a policy that could eliminate severe inequalities. Two-sided investment insurance would be preferred.

Proposition 9. *Assume that $\phi(\theta)$ and $\tau(\psi)$ are continuous, strictly decreasing and strictly positive on $[0,1]$. For any $\pi_B \in [0,1)$ and $\pi_A \in (0,1)$ with $\pi_B < \pi_A$, there exist cost distributions G^W and G^E, a signal distribution $F_i^W(\theta)$ and parameters such that: (i) π_B and π_A are part of an equilibrium; and (ii) no one-sided investment subsidy can raise π_B to π_A in any finite number of periods T, even if combined with affirmative action.*

6 Interpreting Group Differences in the Presence of Two-Sided Statistical Discrimination

Two-sided statistical discrimination complicates empirical analysis, since differences between groups are generically a combination of both employer and worker decision-making. For example, consider a setting with employer learning as in Altonji and Pierret (2001). Under conditions they outline, the conditional expectation for log-wages can be written as a time-varying function of the form:

$$E(w_t|s_i, z_i, t) = b_{s,t}s_i + b_{z,t}z_i + H(t)$$

where: s_i is observable to both the employer and the econometrician; z_i is observable to econometrician but not initially observed (or at least not used) by the employer; t denotes experience; and $H(t)$ is an experience profile of productivity that is assumed not to depend on s_i and z_i.

Altonji and Pierret (2001) show that, if their assumptions hold, $b_{s,t}$ falls with experience, while $b_{z,t}$ rises. Their empirical results show that the racial wage gap rises with experience, which suggests that employers do not fully incorporate racial differences in productivity into their initial wage offer. The conclusion of their analysis is that statistical discrimination cannot explain racial differences in wage profiles.

As Altonji and Pierret (2001) acknowledge, the assumption that the experience profile of productivity is independent of race is restrictive. This motivates them to test for racial differences in training opportunities.

Any such difference could arise from employer discrimination but could also be driven by worker expectations. For example, suppose that workers make costly investments in human capital. The return on investment depends on whether higher productivity will be rewarded by employers. Early in a worker's career, the forces in our model would predict that black workers – if pessimistic – would invest less. This would cause the racial wage gap to widen with experience in the early years, an effect that would be sharpened by the fact that black workers often have disproportionately short tenure. However, information about a specific employer should eventually overwhelm any racial difference in priors. If investment returns are diminishing, we would expect to see convergence between the wages of black and white workers at "good" firms after many years of tenure. This aligns with the results found by Fryer, Pager and Spenkuch (2013), who demonstrate that racial wage gaps widen with experience but narrow after many years of tenure within the same firm.

An implication of a model in which ongoing investments depend on beliefs or otherwise depend on race is that empirical analysis designed to detect statistical discrimination may be misleading. Even if race is an s variable – i.e., employers statistically discriminate – underinvestment by black workers due to their own pessimism would lead to the false conclusion that employers do not discriminate. The same facts could be explained by a model in which the training opportunities offered to workers depend on employer beliefs (see Altonji and Pierret, 2001). The two reasons for underinvestment are thus inseparable with this approach.

Lang and Lehmann (2012) discuss a test that is robust to differing experience profiles of black and white workers. Let B_i indicate whether a worker is black. As before, z_i is correlated with productivity and initially unobserved by the employer. In a simplified model, Lang and Lehmann propose comparing two regressions.

$$E^*\left(w_t|B_i, z_i, t\right) = \alpha_1 + \alpha_2 B_i + \alpha_3 t + \alpha_4 B_i t + \alpha_5 z_i$$
$$E^*\left(w_t|B_i, z_i, t\right) = \beta_1 + \beta_2 B_i + \beta_3 t + \beta_4 B_i t + \beta_5 z_i + \beta_6 z_i t$$

Since employers gradually learn about z_i, low z_i workers would initially be overpaid but their wages would converge to their productivity over time. If black workers have lower z_i on average and employers statistically discriminate, we would expect $\gamma_4 < 0$ and $\gamma_2 > 0$ in the following auxiliary regression.

$$E^*\left(z_t t|B_i, z_i, t\right) = \gamma_1 + \gamma_2 B_i + \gamma_3 t + \gamma_4 B_i t + \gamma_5 z_i$$

Assuming that the weight on z_i increases over time as predicted by employer learning (i.e., $\beta_6 > 0$), this implies that $\alpha_2 < \beta_2$ and $\alpha_4 > \beta_4$, which is precisely what Altonji and Pierret find using the Armed Forces Qualification Test (AFQT) as z_i. As Lang and Lehmann (2012) argue, these results therefore suggest a model in which black-white productivity differences widen over time and employers statistically discriminate.

A. Detecting Employer Discrimination

Even in the presence of complementarity, one can make progress identifying employer discrimination. The approach we suggest is to focus directly on the mechanism through which statistical discrimination affects incentives: pessimistic employer beliefs lower the return to investment for blacks relative to whites. Specifically, we propose a test of whether there is a racial difference in the degree to which imperfect information lowers the return to improving one's own productivity.

Consider the following highly stylized thought experiment. Statistical discrimination should imply that a group j worker who is 10 percent more productive would be paid $\beta_j \leq 10$ percent more because employers shrink their estimates of productivity toward the mean of the group. If statistical discrimination causes β_B to be lower than β_W then investment is undermined for blacks relative to whites. The statistical discrimination literature suggests two reasons why we might expect this to be true.

1. Productivity may be harder to assess for minority workers.

2. Lower investment returns should be expected to compress the productivity distribution for blacks. This implies that priors – if approximately correct – are tighter and that lower returns are self-fulfilling.

Although the latter effect is what we focus on in our model, we do not attempt to provide a way of empirically distinguishing the two competing reasons for statistical discrimination.

The test that we derive involves measuring the relationship between a worker's past wage and her current wage, and comparing the coefficient for black and white workers of similar tenure at their previous firms. The intuition is that if past wages better reflect productivity, and productivity is imperfectly observed at the time of hiring by a new firm, then black workers' past wages should be less predictive of current wages.[24] Models of taste-based discrimination do not share this prediction.

To facilitate empirical analysis, we adapt our model to allow for continuous investment and adopt the assumption that workers are paid their expected marginal product. To begin, assume that output is produced at constant returns to scale using a mass of quality-adjusted labor Q_j and group-specific 'capital' K_j provided by the firm. Each worker provides a unit of physical labor but individuals are heterogenous in their ability a_i. Effective labor is $Q_j = L_j \cdot \overline{a}_j$ where L_j is the aggregate amount of physical labor supplied by group j workers, and $\overline{a}_j$ is the average productivity of those workers. For convenience we adopt a Cobb-Douglas specification for production: $Y_j = K_j^{1-\gamma} Q_j^{\gamma}$. Thus, letting $k_j = K_j / \overline{a}_j L_j$ be the amount of group-specific capital provided by the firm per unit of effective labor, the marginal product of a worker with ability a_i at firm j is $MP_i = a_i \gamma k_j^{1-\gamma}$. A given firm may provide different levels of k_j for members of each group.

For tractability, we assume that worker ability is distributed log-normally: $\ln a_i \sim N\left(\mu_{a,j}, \sigma^2_{a,j}\right)$. Firms receive a noisy but unbiased signal θ_i about each worker's productivity. Specifically, $\ln \theta_i = \ln a_i + \ln \varepsilon_i$

[24] The logic underlying the analysis here is fundamentally similar to Kahn's (2013) test for asymmetry in employer information.

where $\ln \varepsilon_i \sim N\left(0, \sigma^2_{\varepsilon,j}\right)$. If workers are paid their expected marginal product, the wage paid by a firm to a worker with ability a_i at a firm with $k_{j,F}$ can be shown to be as follows.

$$\ln w_i = \left(\frac{\sigma^2_{a,j}}{\sigma^2_{\varepsilon,j}+\sigma^2_{a,j}}\right)\ln a_i + \left(\frac{\sigma^2_{\varepsilon,j}}{\sigma^2_{\varepsilon,j}+\sigma^2_{a,j}}\right)\mu_{a,j} + \ln\gamma + (1-\gamma)\ln k_{j,F} + \frac{1}{2}\left(\frac{\sigma^2_{\varepsilon,j}\sigma^2_{a,j}}{\sigma^2_{\varepsilon,j}+\sigma^2_{a,j}}\right) + \left(\frac{\sigma^2_{a,j}}{\sigma^2_{\varepsilon,j}+\sigma^2_{a,j}}\right)\ln\varepsilon_i$$

Purely for pedagogical purposes, we now temporarily adopt a strong assumption about employer learning, which we will subsequently relax.

Assumption. *For a worker of long enough tenure at her previous employer, her past wage exactly reflects her ability at a that firm and is not observable to a new firm. Ability at the old and new firms are equivalent.*

This is restrictive for two reasons. First, we are assuming that learning is complete with long enough tenure. Secondly, we are assuming that a worker's ability at a new firm is equivalent to her ability at her old firm. These assumptions may both be reasonable in some contexts, but it is easy to imagine violations.

The assumption above allows us to write the wage (w_i) offered to an experienced worker who moves to a new firm as a particularly simple function of her wage at her previous firm (w_i^{OLD}), group-specific fixed effects for the source and destination firms ($\alpha_{j,f^{NEW}}$ and $\alpha_{j,f^{OLD}}$), and an error term ν_i:

$$\ln(w_i) = \beta_j \ln\left(w_i^{OLD}\right) + \alpha_{j,f^{OLD}} + \alpha_{j,f^{NEW}} + \nu_i \tag{7}$$

where $\beta_j = \frac{\sigma^2_{a,j}}{\sigma^2_{\varepsilon,j}+\sigma^2_{a,j}}$, $\nu_i = \left(\frac{\sigma^2_{a,j}}{\sigma^2_{\varepsilon,j}+\sigma^2_{a,j}}\right)\ln\varepsilon_i$ and the fixed effects are functions of the model's parameters.

$$\begin{aligned}
\alpha_{j,f^{NEW}} &= \left(\frac{\sigma^2_{\varepsilon,j}}{\sigma^2_{\varepsilon,j}+\sigma^2_{a}}\right)\mu_{a,j} + (1-\gamma)\,k_{j,F^{NEW}} + \frac{1}{2}\left(\frac{\sigma^2_{\varepsilon,j}\sigma^2_{a,j}}{\sigma^2_{\varepsilon,j}+\sigma^2_{a,j}}\right)\\
\alpha_{j,f^{OLD}} &= -(1-\gamma)\ln k_{j,F^{OLD}}
\end{aligned}$$

The coefficient on a worker's previous wage, β_j is the elasticity of the wage with respect to individual ability. This is a measure of the return to productivity-enhancing investment, with $\beta_j = 1$ corresponding to a worker always receiving her marginal product. The two fixed effect terms allow for potentially different levels of group-specific capital at the new and old firms.

To assess the impact of statistical discrimination, we propose a test of whether the return to ability is lower for blacks. Since β_j is exactly the degree to which statistical discrimination lowers the return to productivity, this amounts to the following statistical test.

$$\begin{aligned}
H_0\text{: } \Gamma &= \beta_W - \beta_B \leq 0\\
H_1\text{: } \Gamma &= \beta_W - \beta_B > 0
\end{aligned}$$

With data on past and present wages, and adequate movement between firms, it is straightforward to estimate equation (7) for each group and calculate an estimate of the racial difference in returns $\hat{\Gamma} = \hat{\beta}_W - \hat{\beta}_B$.

A potential complication for the attainment of a consistent estimate of Γ in a regression is that movement between firms may be non-random. Yet, selective movement does not necessarily affect the estimated relationship between current and past wages ($\hat{\beta}_j$), *conditional on including firm fixed effects*. For example, there may be correlation between the investments made by a worker's current and previous firm but this is accounted for by including fixed effects for both firms. Alternatively, idiosyncratic match effects or a connection between firm-wide shocks and mobility would bias estimates of the firm fixed effects themselves (see Card et al., 2016) but do not necessarily affect $\hat{\Gamma}$.

We next relax the assumption that past wages fully reflect productivity. First, we allow for the possibility that the previous employer also has imperfect information about a worker's ability. However, we continue to assume that this information is better than the new firm in the sense that $\sigma^2_{\varepsilon,j,OLD} < \sigma^2_{\varepsilon,j}$, since some private learning has occurred over the worker's tenure. Second, we allow for the possibility that ability at the new and old firms are correlated but not equivalent.

Under these alternative assumptions, we argue that our proposed test for statistical discrimination is conservative in the sense that differences in returns to ability are understated. Specifically, we argue that $\hat{\Gamma}$ is a downward-biased measure of the difference in returns to ability induced by statistical discrimination.

Proposition 10. *Assume that ability at the new and old firms are correlated:* $\ln a_i = c_j + \rho \ln a_i^{OLD} + \ln \eta_i$ *where* $0 < \rho \leq 1$. *Then the difference in coefficients from equation (7) is* $\hat{\Gamma}$, *where:*

$$\hat{\Gamma} = \rho \left[\underbrace{\left(\frac{\sigma^2_{a,W}}{\sigma^2_{\varepsilon,W} + \sigma^2_{a,W}} \right) - \left(\frac{\sigma^2_{a,B}}{\sigma^2_{\varepsilon,B} + \sigma^2_{a,B}} \right)}_{\Gamma \text{ (true difference in returns)}} + \underbrace{\left(\frac{\sigma^2_{\varepsilon,W,OLD}}{\sigma^2_{\varepsilon,W} + \sigma^2_{a,W}} \right) - \left(\frac{\sigma^2_{\varepsilon,B,OLD}}{\sigma^2_{\varepsilon,B} + \sigma^2_{a,B}} \right)}_{\text{Bias term}} \right].$$

It is evident from Proposition 10 that imperfect correlation between ability at the old and new firms ($0 < \rho < 1$) biases $\hat{\Gamma}$ toward zero. Second, any statistical discrimination against black relative to white workers reduces the return to investment, compressing the productivity distribution and leading to a tighter employer prior ($\sigma^2_{a,B} \leq \sigma^2_{a,W}$). This further pushes toward a downward-biased estimate of Γ. Finally, as long as workers of similar tenure are compared across races, we would expect a similar amount of learning to have occured so that $\frac{\sigma^2_{\varepsilon,B}}{\sigma^2_{\varepsilon,B,OLD}} \approx \frac{\sigma^2_{\varepsilon,W}}{\sigma^2_{\varepsilon,W,OLD}}$. This is enough to conclude that our proposed test remains a conservative measure of the disparate impact of statistical discrimination.

7 Conclusion

Statistical discrimination is a foundational concept in the economic analysis of discrimination. Intuitively, the information problem inherent in such models seems two-sided. Yet, current models do not take this into account. Two-sided belief formation makes the interpretation of any empirical data on group differences more complicated – though not impossible – because for any disparity, differences can be driven by either side of the market: workers or firms; universities or applicants; police or civilians.

Furthermore, policies designed to break equilibria with negative beliefs about certain groups are complicated by complementarity between the beliefs and actions of workers and firms. Affirmative action, employment quotas, wage subsidies, and unemployment insurance all perform quite poorly relative to traditional statistical discrimination models. Indeed, we demonstrate that any one-sided policy fails to reliably ensure homogeneous beliefs.

We posit a new policy – two-sided investment insurance – as a solution to statistical discrimination. Investment insurance is a method for the government or another entity to guarantee returns for workers it deems as investors, while rewarding firms for making their workplaces productive for all types. We demonstrate that this policy, or one equivalent to it, weakly dominates any alternative. *Year Up* is a strikingly successful example of this type of opportunity for urban youth. Similar policies might be envisoned for broader classes of workers.

References

[1] Aigner, and G. Cain. "Statistical Theories of Discrimination in Labor Markets." *Industrial and Labor Relations Review* 30 (1977): 175-187.

[2] Altonji, Joseph G. and Rebecca M. Blank. "Race and Gender in the Labor Market." In *Handbook of Labor Economics*, ed. O. Ashenfelter and D. Card. Elsevier Science, 1999.

[3] Altonji, Joseph G. and Charles R. Pierret. "Employer Learning and Statistical Discrimination." *The Quarterly Journal of Economics* 116, no.1 (2001): 313-350.

[4] Anderson, Simon P. and Stephen Coate. "Market Provision of Broadcasting: A Welfare Analysis." *Review of Economic Studies* 72, no.4 (2005): 947-972.

[5] Armstrong, Mark. "Competition in Two-sided Markets." *The RAND Journal of Economics* 37, no.3 (2006): 668-691.

[6] Arrow, Kenneth J. "The Theory of Discrimination." In *Discrimination in Labor Markets*, ed. O. Ashenfelter and A. Rees. New Jersey: Princeton University Press, 1973.

[7] Becker, Gary S. *Human Capital.* 2nd ed. New York: Columbia University Press, 1964.

[8] Becker, Gary S. *The Economics of Discrimination.* 2nd ed. Chicago: The University of Chicago Press, 1957.

[9] Black, Dan A. "Discrimination in an Equilibrium Search Model." *The American Economic Review* 13, no.2 (1995): 309-334.

[10] Bulow, Jeremy I., John D. Geanakoplos, and Paul D. Klemperer. "Multimarket Oligopoly: Strategic Substitutes and Complements." *Journal of Political Economy* 93, no.3 (1985): 488-511.

[11] Caillaud, Bernard, and Bruno Jullien. "Chicken-and-Egg: Competition among Intermediation Service Providers." *The RAND Journal of Economics* 34, no.2 (2003): 309-328.

[12] Card, David, Ana Rute Cardoso, and Patrick Kline. "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women." *Quarterly Journal of Economics* 131, no.2 (2016): 633-686

[13] Cave, George, Hans Bos, Fred Doolittle, and Cyril Toussaint. "JOBSTART: Final Report on a Program for School Dropouts." Manpower Demonstration Research Corporation, October 1993.

[14] Chung, Kim-Sau. "Affirmative Action as an Implementation Problem." Unpublished paper, University of Minnesota, 1999.

[15] Coate, Stephen and Glenn C. Loury. "Will Affirmative-Action Policies Eliminate Negative Stereotypes?" *The American Economic Review* 83, no.5 (1993): 1220-1240.

[16] Cooper, Russell and Andew John. "Coordinating Coordination Failures in Keynesian Models" *Quarterly Journal of Economics* 103, no.3 (1988): 441-463.

[17] Fang, Hanming. "Social Culture and Economic Performance." *The American Economic Review* 91, no. 4 (2001): 924-937.

[18] Fang, Hanming and Andrea Moro. "Theories of Statistical Discrimination and Affirmative Action: A Survey." In *Handbook of Social Economics Volume 1A*, ed. Jess Benhabib, Matthew O. Jackson, and Alberta Bisin (2011): 133-200.

[19] Filippin, Antonio. "Can Workers' Expectations Account for the Persistence of Discrimination?" IZA Discussion Paper 4490. Institute for the Study of Labor (IZA), 2009.

[20] Fryer, Roland G. "Belief Flipping in a Dynamic Model of Statistical Discrimination." *Journal of Public Economics* 91 5-6 (2007): 1151-1166.

[21] Fryer, Roland G., Devah Pager, and Jörg L. Spenkuch. "Racial Disparities in Job Finding and Offered Wages." *Journal of Law and Economics* 56 no.3 (2013): 633-689.

[22] Goldin, Claudia and Lawrence Katz. "A Most Egalitarian Profession: Pharmacy and the Evolution of a Family-Friendly Occupation." *Journal of Labor Economics* vol. 34 no.3 (2016): 705-746.

[23] Hendra, Richard, David H. Greenberg, Gayle Hamilton, et al. *Encouraging Evidence on a Sector-Focused Advancement Strategy: Two-Year Impacts from the WorkAdvance Demonstration.* MDRC Report, August 2016.

[24] Hossain, Farhana and Dan Bloom. *Toward a Better Future: Evidence on Improving Employment Outcomes for Disadvantaged Youth in the United States.* MDRC Report, February 2015.

[25] "Job Training Partnership Act; Final Rule," Employment and Training Administration (Department of Labor), 59 Federal Register 170 (2 September 1994): 20 CFR Part 626, et al.

[26] Kahn, Lisa B. "Asymmetric Information between Employers." *American Economic Journal: Applied Economics* 5, no.4 (2013): 165-205.

[27] Knowles, John, Persico, Nicola, and Todd, Petra. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy* 109, no.1 (2001): 203-229.

[28] Lang, Kevin. "A Language Theory of Discrimination." *Quarterly Journal of Economics* 101, May (1986): 363-382.

[29] Lang, Kevin, and Lehmann, Jee-Yeon K. "Racial Discrimination in the Labor Market: Theory and Empirics." *Journal of Economic Literature* 50, no.4 (2012): 959-1006.

[30] Lang, Kevin, Manove, Michael, and Dickens, William T. "Racial Discrimination in Labor Markets with Posted Wage Offers." *American Economic Review* 95, no.4, (2005): 1327-1340.

[31] Milgrom, Paul and Sharon Oster. "Job Discrimination, Market Forces, and the Invisibility Hypothesis." *The Quarterly Journal of Economics* 102, no.3 (1987): 453-476.

[32] Milgrom Paul, and John Roberts. "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities." *Econometrica* 58, no.6 (1990): 1255-1277.

[33] Moro, Andrea and Peter Norman. "A General Equilibrium Model of Statistical Description." *Journal of Economic Theory* 114, (2004): 1-30.

[34] Neusser, Klaus. "Difference Equations for Economists." Unpublished manuscript, October 3 2016.

[35] Nöldeke, Georg and Samuelson, Larry. "Investment and Competitive Matching." *Econometrica* 83, no.3 (2015): 835-896.

[36] Phelps, Edmund S. "The Statistical Theory of Racism and Sexism." *The American Economic Review* 62, no.4 (1972): 659-661.

[37] "Revised Philadelphia Plan." Department of Labor Order (27 June 1969).

[38] Rochet, Jean-Charles, and Jean Tirole. "Platform Competition in Two-Sided Markets." *The European Economic Journal* 1 no.4 (2003): 990-1029.

[39] Roder, Anne and Mark Elliott. *Sustained Gains: Year Up's Continued Impact on Young Adults' Earnings.* Economic Mobility Corporation Report, May 2014.

[40] Schochet, Peter Z., John Burghardt, and Sheena McConnell. "Does Job Corps Work? Impact Findings from the National Job Corps Study." *American Economic Review* 98 no. 5 (2008): 1864-86.

[41] Topkis, Donald M. "Equilibrium Points in Nonzero-Sum n-person Submodular Games." *SIAM Journal on Control and Optimization* 17 no.6 (1979): 773-787.

[42] U.S. Congress. Commitee on Education and Labor. *Equal Employment Opportunity.* U.S. Government Printing Office, 1963.

[43] Vives, Xavier. "Nash Equilibrium with Strategic Complementarities." *Journal of Mathematical Economics* 19, no.3 (1990): 305-321.

[44] Weyl, E. Glen. "A Price Theory of Multi-sided Platforms." *American Economic Review* 100, no.4 (2010): 1642-1672.

[45] "Workforce Investment Act; Final Rule," Employment and Training Administration (Department of Labor), 65 Federal Register 156 (11 August 2000): 20 CFR Part 652 et al.

[46] "Workforce Innovation and Opportunity Act; Final Rule," Employment and Training Administration (Department of Labor), 81 Federal Register 161 (19 August 2016): 20 CFR Part 603, 651, 652, et al.

8 Online Appendix (Not For Publication)

8.1 Technical Proofs

Proof of Proposition 1. Given the assumptions, the worker and employer EE curves lie above their WW curves for s and t near zero and one respectively. The conditions that $G^W(\beta_W(s,\delta)) > \phi(s)/[\chi_q/\chi_u + \phi(s)]$ and $G^E(\beta_E(t,\pi)|\lambda) > \tau(t)/[\omega_q/\omega_u + \tau(t)]$ guarantee that the EE curves and WW curves cross at least once, implying at least two non-zero solutions to each of (3) and (4). Since $\overline{\omega}(\delta)$ and $\overline{\chi}(\pi)$ are increasing, the same is true for any $\delta > \underline{\delta}$ and $\pi > \underline{\pi}$. Assume, then, that $\underline{\delta}$ and $\underline{\pi}$ are the lowest values for which these conditions hold. Below $\underline{\delta}$ and $\underline{\pi}$, there is no non-zero solution to (3) and (4) respectively. Observe that the non-zero solutions in $\pi^*(\delta)$ are bounded strictly between zero and one. This is because: (i) our assumptions on $\phi(\theta)$ guarantee that $s^*(\pi) = 1$ for any $\pi < \underline{\pi}$; and (ii) there exists a threshold $\pi < 1$ above which $s^*(\pi) = 0$. Both $s = 0$ and $s = 1$ imply zero worker investment. Equivalent arguments imply that the non-zero solutions in $\delta^*(\pi)$ are bounded strictly between zero and one.

For any value of δ, let $\overline{\pi}^*(\delta) = \max\{\pi^*(\delta)\}$. Similarly define $\overline{\delta}^*(\pi) = \max\{\delta^*(\pi)\}$. Both $\overline{\pi}^*(\delta)$ and $\overline{\delta}^*(\pi)$ are obviously defined on $[0,1]$. Both are also increasing in their arguments. To see why, start at $\overline{\pi}(\delta_1)$, at which $s = s_1$. Consider increasing δ to $\delta_2 > \delta_1$. For any given s, $G([F_u(s) - F_q(s)]\overline{\omega}(\delta))$ increases since $\overline{\omega}(\delta)$ is increasing in δ. This means that $G([F_u(s_1) - F_q(s_1)]\overline{\omega}(\delta_2)) > \overline{\pi}(\delta_1)$. In other words, the WW curve is above the EE curve at s_1. Thus, since the EE curve is strictly decreasing and $G([F_u(0) - F_q(0)]\overline{\omega}(\delta_2)) = 0$, there must be at least one solution to the left of s_1, which implies a value of $\overline{\pi}(\delta_2)$ greater than $\overline{\pi}(\delta_1)$. An analogous argument can be used to show that $\overline{\delta}^*(\pi)$ is increasing in π.

We directly assume that is some $\{\pi, \delta\}$ such that $\delta \in \delta^*(\pi)$ and $\pi < \overline{\pi}^*(\delta)$. Combined with the monotonicity of $\overline{\pi}^*(\delta)$ and $\overline{\delta}^*(\pi)$, this implies that there is some π such that $\overline{\pi}^*\left(\overline{\delta}^*(\pi)\right) > \pi$. We also know that $\overline{\pi}^*\left(\overline{\delta}^*(1)\right) < 1$ since $\overline{\pi}^*(\delta)$ is bounded below 1. There must therefore be a $\tilde{\pi}$ such that $\overline{\pi}^*\left(\overline{\delta}^*(\tilde{\pi})\right) = \tilde{\pi}$. To see why, suppose that there is not. Then there must be a downward discontinuity in $\overline{\pi}^*\left(\overline{\delta}^*(\pi)\right) - \pi \leq 0$. This is impossible since π is continuous and $\overline{\pi}^*\left(\overline{\delta}^*(\pi)\right)$ is positive monotonic. Since $\tilde{\pi}$ is a non-zero solution and $\delta = \pi = 0$ always satisfies both (3) and (4), there are multiple solutions to the two-sided problem. □

Proof of Proposition 2. Assume that there exists an equilibrium without affirmative action in which there is positive investment in the A market: $\pi_A = \pi_A^* > 0$ and $\delta_A = \delta_A^* > 0$. In the B market, there is always an equilibrium with zero investment: $\pi_B = \delta_B = 0$. Now suppose that, under affirmative action, $\pi_A = \pi_A^*$, $\delta_A = \delta_A^*$, $\pi_B = \delta_B = 0$ and $t_B = t_B^*(0) = 1$. Then affirmative action is non-binding since, with zero workers applying, an employer's profits are independent of s_B. It therefore optimally sets $s_B = s_B^{AA}$ such that the

affirmative action constraint holds.

$$\rho\left(s^*\left(\pi_A^*\right),\pi_A^*\right)=\pi_A^*\left[1-F_q^W\left(s^*\left(\pi_A^*\right)\right)\right]+\left(1-\pi_A^*\right)\left[1-F_u^W\left(s^*\left(\pi_A^*\right)\right)\right]=1-F_u^W\left(s_B^{AA}\right)=\rho\left(s_B^{AA},0\right)$$

Regardless of s_B^{AA}, $\pi_B=\delta_B=0$ is obviously still an equilibrium, since both firms and workers have zero investment returns. □

Proof of Proposition 3. First consider an equilibrium without affirmative action. Suppose that $\pi_A=\pi_B=\pi>0$. Then employers' unique optimal signal threshold is $s_A=s_B=s^*\left(\pi\right)$. Worker beliefs δ cannot be homogeneous since $G^E\left(\beta_E\left(t^*\left(\delta\right),\pi\right)|\lambda\right)$ must be strictly lower for the minority for a given any threshold t. Combined with $s_A=s_B$ this is incompatible with $\pi_j=G^W\left(\beta_W\left(s_j,\delta_j\right)\right)$ being the same for both groups. This is a contradiction. Finally, consider imposing an affirmative action target. Since the constraint does not bind if employers have homogeneous beliefs, they still set $s_A=s_B=s^*\left(\pi\right)$. By the same logic as above, worker beliefs cannot be homogenous, which is incompatible with $\pi_A=\pi_B$. □

Proof of Proposition 4. If δ_B is low enough, affirmative action lowers ever firm's threshold for Bs to some $s<s_B$ if beliefs are held constant.[25] Now consider the firm's payoff conditional on a worker application.

$$\pi_B\left(1-F_q^W\left(s\right)\right)\chi_q-\left(1-\pi_B\right)\left(1-F_u^W\left(s\right)\right)\chi_u$$

Under our assumptions, low enough π_B ensures that $s^*\left(\pi_B\right)=1$. Suppose that π_B is above but close to this critical value such that $s^*\left(\pi_B\right)\approx 1$. The firm's total payoff is then arbitrarily close to zero. The imposition of $s<s_B$ adds a strictly positive mass of workers with negative expected payoffs to the firm, ensuring that the total firm payoff from hiring type B workers is negative. As a result, zero firms subsequently invest in the B amenity. In turn, this ensures that no B workers have an incentive to invest. □

Lemma. *Assume that $\phi\left(\theta\right)$ and $\tau\left(\psi\right)$ are continuous, strictly decreasing and strictly positive on $[0,1]$. For fixed current beliefs with $\pi_A>\pi_B$, $\delta_A>\delta_B$ and t_B close enough to one, affirmative action lowers s_B for all firms with s_A approximately unchanged.*

Proof. The Lagrangean for an affirmative action target is as follows.

$$\mathcal{L}\left(s_A,s_B,i_A,i_B,\gamma|\pi_A,\pi_B\right)=\lambda_A P\left(s_A,\pi_A,i_A\right)+\lambda_B P\left(s_B,\pi_B,i_B\right)+\gamma\left[\rho\left(s_B,\pi_B\right)-\rho\left(s_A,\pi_A\right)\right]$$

[25] Low enough δ_B ensures $t_B\approx 1$. In Appendix 8.1, we show that affirmative action amounts to setting a lower s_B in this case.

where $\rho(s_j, \pi_j)$ is the probability that the employer assigns to hiring a randomly drawn worker from group $j \in \{A, B\}$ and $P(s_j, \pi_j, i_j)$ is the expected payoff from making an offer to said worker (which depends on whether the firm has invested – $i_j \in \{q, u\}$).

$$\rho(s_j, \pi_j) = \pi_j \left[1 - F_q^W(s_j)\right] + (1 - \pi_j)\left[1 - F_u^W(s_j)\right]$$
$$P(s_j, \pi_j) = \pi_j \left[1 - F_q^W(s_j)\right]\left[1 - F_{i_j}^E(t_j)\right]\chi_q - (1 - \pi_j)\left[1 - F_u^W(s_j)\right]\left[1 - F_{i_j}^E(t_j)\right]\chi_u$$

This is enough for us to write down the expressions for the key FOCs.

$$\gamma\left[\pi_A f_A^W(s_A) + (1 - \pi_A) f_u^W(s_A)\right] = \lambda_A\left[1 - F_{i_A}^E(t_A)\right]\left[\pi_A f_q^W(s_A)\chi_q - (1 - \pi_A) f_u^W(s_A)\chi_u\right]$$
$$-\gamma\left[\pi_B f_q^E(s_B) + (1 - \pi_B) f_u^E(s_B)\right] = \lambda_B\left[1 - F_{i_B}^E(t_B)\right]\left[\pi_B f_q^E(s_B)\chi_q - (1 - \pi_B) f_u^E(s_B)\chi_u\right]$$

These can be re-arranged as follows.

$$\left(\frac{1 - \pi_A}{\pi_A}\right)\frac{f_u^W(s_A)}{f_q^W(s_A)} = \frac{\chi_q - \frac{\gamma}{\lambda_A\left[1 - F_{i_A}^E(t_A)\right]}}{\chi_u + \frac{\gamma}{\lambda_A\left[1 - F_{i_A}^E(t_A)\right]}} = r_A(\gamma)$$
$$\left(\frac{1 - \pi_B}{\pi_B}\right)\frac{f_u^W(s_B)}{f_q^W(s_B)} = \frac{\chi_q + \frac{\gamma}{\lambda_B\left[1 - F_{i_B}^E(t_B)\right]}}{\chi_u - \frac{\gamma}{\lambda_B\left[1 - F_{i_B}^E(t_B)\right]}} = r_B(\gamma)$$

These FOCs characterize the firm's signal thresholds for any given investment decision. The threshold t_B being close to one means that $\left[1 - F_{i_B}^E(t_B)\right]$ is close to zero and all the adjustment occurs on the B side: the multiplier approaches zero in this case. Intuitively, if very few Bs apply then it is nearly costless to adjust on their margin relative to adjustment on the A side.

More formally, we know that the two signals must change in the following proportion to satisfy the affirmative action constraint.

$$\frac{\partial s_B}{\partial s_A} = -\frac{\pi_A f_q^W(s_A) + (1 - \pi_A)\pi_A f_q^W(s_A)}{\pi_B f_q^W(s_B) + (1 - \pi_B)\pi_B f_q^W(s_B)}$$

This implies that the change in profits from an increase in s_A is proportional to:

$$-\lambda_A\left[\frac{\pi_A f_q^W(s_A)\chi_q - (1 - \pi_A) f_u^W(s_A)\chi_u}{\pi_A f_q^W(s_A) + (1 - \pi_A)\pi_A f_q^W(s_A)}\right]\left[1 - F_{i_A}^E(t_A)\right] + \lambda_B\left[\frac{\pi_B f_q^W(s_B)\chi_q - (1 - \pi_B) f_u^W(s_B)\chi_u}{\pi_B f_q^W(s_B) + (1 - \pi_B)\pi_B f_q^W(s_B)}\right]\left[1 - F_{i_B}^E(t_B)\right].$$

Our assumptions on $\phi(\theta)$ and $\tau(\psi)$ imply that as $t_B \to 1$, the firm's optimal s_A approaches $s^*(\pi_A)$. Since the affirmative action constraint implies that s_B is strictly less than s_A, $s_B > s_A$ without affirmative action and s_A is approximately unchanged, s_B is lower for all firms with the additional constraint. □

Proof of Proposition 5. First note that the firm investment rate is bounded strictly below $\overline{\delta} = G^E(\omega_q) < 1$. With close-to-perfect observability of firm investment, $f_q^E(\psi) \to 0$ for all $\psi < 1$ and $f_u^E(\psi) \to 0$ for all $\psi > 0$. This implies that for any $\delta_j \in \left(0, \overline{\delta}\right]$ workers can and will optimally set $t = t^*(\delta_j)$ such that $F_u^E(t^*(\delta_j)) \to 1$ and $F_q^E(t^*(\delta_j)) \to 0$. The mass of workers *hired* by any firm that only makes one of the two investments must therefore be approximately zero (for both types), which implies that the return to making a single investment is approximately zero. Nearly all firms therefore make both investments or neither, which means that $\delta_A \approx \delta_B$. Finally, firms that do not invest at all hire approximately zero workers as well.

Combined, this means that as firm investment becomes near-perfectly observable, there is only one type of firm for which the employment quota could lead to different signal thresholds that impact worker investment: firms that make both investments. Specifically, the fraction of workers who invest approaches: $\pi_j \to G^W\left(\delta_j\left[F_u^W\left(s_j^{q,q}\right) - F_q^W\left(s_j^{q,q}\right)\right]\omega_q\right)$.

Next, for firms that made both investments, the affirmative action constraint amounts to:

$$\pi_A\left[1 - F_q^W\left(s_A^{q,q}\right)\right] + (1-\pi_A)\left[1 - F_u^W\left(s_A^{q,q}\right)\right] = \pi_B\left[1 - F_q^W\left(s_B^{q,q}\right)\right] + (1-\pi_B)\left[1 - F_u^W\left(s_B^{q,q}\right)\right]$$

Using $\pi = G^W(\beta_W(s,\delta))$, define $\hat{\rho}(s|\delta)$ as the probability of employment for any given s (fixing δ).

$$\hat{\rho}(s|\delta) = G^W\left(\delta\left[F_u^W(s) - F_q^W(s)\right]\omega_q\right)\left[1 - F_q^W(s)\right] + \left(1 - G^W\left(\delta\left[F_u^W(s) - F_q^W(s)\right]\omega_q\right)\right)\left[1 - F_u^W(s)\right]$$

The slope of $\hat{\rho}(s|\delta)$ with respect to s is as follows.

$$\hat{\rho}'(s|\delta) = g^W \cdot \delta\left[f_u^W(s) - f_q^W(s)\right]\omega_q\left[F_u^W(s) - F_q^W(s)\right] - \pi f_q^W(s) - (1-\pi) f_u^W(s)$$

This is always strictly negative if $\phi(s) \leq 1$, so a sufficient condition for strict monotonicity is that $\hat{\rho}'(s|\delta) < 0$ for all $\delta \in [0,1]$ and $s : \phi(s) > 1$. Re-arranging the expression for $\hat{\rho}'(s|\delta)$, this requirement amounts to the following condition.

$$\eta\left(\overline{\beta}(s)\right) < \frac{\phi(s)}{\phi(s) - 1}$$

for all $s : \phi(s) > 1$ where $\eta(c) = \frac{d[c \cdot G(c)]}{dc}$ and $\overline{\beta}(s) = \left[F_u^W(s) - F_q^W(s)\right]\omega_q$.

Finally, as $\delta_A \to \delta_B$, non-homogeneous beliefs requires that $\hat{\rho}(s_A|\delta_A) = \hat{\rho}(s_A|\delta_B)$ for some $s_A \neq s_B$. But with $\delta_A \approx \delta_B$, a necessary condition for this is that $\hat{\rho}(s|\delta)$ is non-monotonic. If the above condition is satisfied, this is not possible. □

Proof of Proposition 6. Consistent with our proposed dynamic adjustment process, fix beliefs at their original values. Given these beliefs, there are two actions that a firm can take to boost employment of minorities:

(a) it can make the B investment if it was not doing so already; and (b) it can lower its hiring standard for B workers. However, if few enough minority workers are applying (i.e., low enough δ_B), a standard of $s_B = 0$ still does not allow the firm to satisfy the employment quota, regardless of its investment decision. Thus, the equality constraint must entail an immediate reduction in employment of type A workers.

Next, consider the firm's choices of s_A and s_B. The firm's problem can be written as follows.

$$\begin{aligned} \max_{s_A, s_B, i_A, i_B} \; & \lambda_A P(s_A, \pi_A, i_A) + \lambda_B P(s_B, \pi_B, i_B) \\ \text{s.t.} \; & [\rho(s_B, \pi_B, i_B) = \rho(s_A, \pi_A, i_A)] \\ & s_B \geq 0 \end{aligned}$$

In principle, there are also constraints that $s_B \leq 1$, $s_A \geq 0$ and $s_A \leq 1$ but these will never bind.

We now proceed to prove that an internal solution does not exist for some values of the parameters. To do so, assume that the inequality constraint does not bind and let γ be the multiplier on the equality constraint. Differentiating with respect to s_A and s_B, we obtain expressions for the effect of increasing each threshold.

$$-\lambda_A \left[\pi_A f_q^W(s_A)\chi_q - (1-\pi_A) f_u^W(s_A)\chi_u\right] + \gamma\left[\pi_A f_A^W(s_A) + (1-\pi_A) f_u^W(s_A)\right] \tag{8}$$

$$-\lambda_B \left[\pi_B f_q^E(s_B)\chi_q - (1-\pi_B) f_u^E(s_B)\chi_u\right] - \gamma\left[\pi_B f_q^E(s_B) + (1-\pi_B) f_u^E(s_B)\right] \tag{9}$$

For low enough δ_B, $t_B \to 1$ and the maximum hiring probability that a firm can achieve for B worker is $\rho_H(0, \pi_j, q) = \left[1 - F_q^E(t_j)\right] \to 0$. We therefore also require that $\rho(s_A, \pi_A, i_A) \to 0$, which in turn implies that $s_A \to 1$. Setting (8) to zero allows us to obtain the limit for the multiplier γ in this scenario.

$$\gamma \to \gamma^* = \lambda_A \left[\frac{\pi_A \chi_q - (1-\pi_A)\phi(1)\chi_u}{\pi_A + (1-\pi_A)\phi(1)}\right]$$

A sufficient condition for (9) to be strictly less than zero for any value of s_B is that $\lambda_B \chi_u > \gamma^*$. Fixing π_A, this condition must hold for an open set of parameters and ensures that we obtain a boundary solution with $s_B = 0$. This contradicts the assumption that there is an internal solution to the firm's problem, proving that the firm sets $s_B = 0$. This in turn ensures that no workers have an incentive to invest and that $\pi_B = 0$ subsequently. □

Proof of Proposition 7. Consider the effectiveness of setting $\{s, t\}$, $\{t, \delta\}$ or $\{s, \pi\}$ when $\pi_0 = \delta_0 = 0$. Even if $s \in [0, 1]$ and $t \in [0, 1]$, the investment returns of workers (β^W) and firms (β^E) are weakly negative as long as $\delta = 0$ and $\pi = 0$ respectively. Similarly, $s = 1$ and $\pi = 0$ ensures that $\beta^W = 0$ and $\beta^E = 0$. The same

applies whenever $t = 1$ and $\delta = 0$. The failure of these pairs also implies that no intervention on a single margin can succeed.

Next, it is obvious that targeting both investment decisions is effective. If the planner sets $\delta = \delta^*$ and $\pi = \pi^*$, workers and firms immediately set $s = s^*$ and $t = t^*$. A second effective policy is to set $t = t^*$ and $\pi = \pi^*$, which ensures that firms set $s = s^*$ immediately and $\delta = \delta^*$ in the following period. Analogous logic applies to a policy that sets $s = s^*$ and $\delta = \delta^*$.

Finally, a $\{\delta, \pi\}$ policy can achieve π^* and δ^* in exactly one period, while any $\{t, \pi\}$ policy must take more than one period if $\pi_0 \neq \pi^*$ and $t_0 \neq t^*$. Intuitively, firms do not change their investment immediately because they have seen no evidence of a change in worker behavior. Similar logic again applies to $\{s, \delta\}$. □

Proof of Proposition 8. First, set government signal thresholds $s^g \in (0,1)$ and $t^g \in (0,1)$. Next, set worker and firm incentive payments ω_g and χ_g. Let $\hat{F}_i^W(\theta^g)$ and $\hat{F}_i^E(\theta^g)$, $i \in \{q, u\}$, be the distributions of θ^g and ψ^g respectively, the increase in investment returns for workers and firms are:

$$\left[\hat{F}_u^W(s^g) - \hat{F}_q^W(s^g)\right] \cdot \omega^g$$
$$\left[\hat{F}_u^E(t^g) - \hat{F}_q^E(t^g)\right] \cdot \chi^g$$

Providing that the government signals are strictly informative, these payments can be set such that the expected returns to investment for B workers and firms investing in the B amenity are equal to those that would prevail at $\{s_A, t_A, \pi_A, \delta_A\}$. In response, the fraction of B workers who invest is $\pi_B = \pi_A$ and the fraction of firms who invest in the B amenity is $\delta_B = \delta_A$. Then B workers set $t = t_A$ and firms $s = s_A$. Once $\pi_B = \pi_A$ and $\delta_B = \delta_A$ have been achieved, they can be retained with only the following permanent investment subsidy.

$$\left[\hat{F}_u^E(t^g) - \hat{F}_q^E(t^g)\right] \cdot \chi^g = (\lambda_A - \lambda_B) \cdot \left[F_u^E(t) - F_q^E(t)\right] \overline{\chi}(\pi_A)$$

Clearly $\lambda_A = \lambda_B$ ensures that the required permanent investment subsidy is zero. □

Proof of Proposition 9. Assume that G^W and G^E are strictly increasing with $G^E(0) = G^W(0) = 0$, and take any $\pi_B \in [0,1)$ and $\pi_A \in (0,1)$ with $\pi_B < \pi_A$. These worker investment levels, combined with signal distributions and threshold rules t^* and s^* pin down firm investment returns for any δ_j. The fractions of firms that invest in each amenity are:

$$\delta_j = G^E\left(\left[F_u^E(t^*(\delta_j)) - F_q^E(t^*(\delta_j))\right] \overline{\chi}(\pi_j)\right)$$

If $\pi_B = 0$ then $\delta_B = 0$ for any G^E satisfying our assumptions, which ensures equilibrium in the B market. For any $\pi_j > 0$, there always exists a set of worker payoffs (ω_q and ω_u) and a distribution function G^E such that this equation is solved by any $\delta_A \in (0,1)$ and $\delta_B \in (0,1)$ with $\delta_B < \delta_A$, given π_A and π_B. Combined with the worker and firm threshold rules, δ_A and δ_B pin down worker investment. The fractions of workers who invest are:

$$\pi_j = G^W\left(\left[F_u^W\left(s^*\left(\pi_j\right)\right) - F_q^W\left(s^*\left(\pi_j\right)\right)\right]\overline{\omega}\left(\delta_j\right)\right)$$

Since $\delta_A > \delta_B > 0$, there always exists a function G^W that satisfies our assumptions and for which π_A and π_B satisfy this equation given firm investment rates δ_A and δ_B respectively.

Next, the maximum level of worker investment that can be achieved with firm investment incentives alone is as follows.

$$\pi_{B,t} = G^W\left(\left[F_u^W\left(s^*\left(\pi_{B,t-1}\right)\right) - F_q^W\left(s^*\left(\pi_{B,t-1}\right)\right)\right]\omega_q\right)$$

Clearly the maximum difference between worker investment rates occurs as $\pi_{B,t-1} \to 0$ and $\pi_{A,t-1} \to 1$. Moreover, equilibrium worker investment $\pi_{B,t-1}$ close to zero implies, for strictly positive ω_q, that $F_u^W\left(s^*\left(\pi_{B,t-1}\right)\right) - F_q^W\left(s^*\left(\pi_{B,t-1}\right)\right) \approx 0$. This ensures that, for any finite ω_q, $\pi_{B,t}$ is also arbitrarily close to zero. The same logic implies that $\pi_{B,t+1} \approx 0$, given that $\pi_{B,t} \approx 0$. Fixing a finite time horizon T, a low enough $\pi_{B,t}$ therefore ensures that no one-sided investment incentive can achieve equality by time T.

Similar but more complex logic applies when affirmative action is allowed. As $\pi_B \to 0$, firm investment returns approach zero and thus $\delta_B \to 0$. In turn, this implies that $t^*\left(\delta_B\right) \to 1$. Firms therefore respond to AA by changing s_B to $\hat{s}_B < s_A$, with s_A unchanged (see Lemma in Appendix 8.1). Now assume a signal distribution $F_i\left(\theta\right)$ such that $\phi\left(s_A\right) > 1$. This implies that $F_u^W\left(\hat{s}_B\right) - F_q^W\left(\hat{s}_B\right) < F_u^W\left(s_A\right) - F_q^W\left(s_A\right)$ for any $\hat{s}_B < s_A$. If we then assume that $G^W\left(x\right) \approx 0$ for any $x < F_u^W\left(s_A\right) - F_q^W\left(s_A\right)$, this ensures that $\pi_{B,t} \approx 0$. The same logic implies that $\pi_{B,t+1} \approx 0$, given that $\pi_{B,t} \approx 0$. Fixing a finite time horizon T, a low enough $\pi_{B,t}$ therefore ensures that no one-sided investment incentive can achieve equality by time T, even if combined with affirmative action. □

Proof of Proposition 10. The log wage of a worker at her new firm is given by the following equation.

$$\begin{aligned}\ln w_i = &\left(\frac{\sigma_{a,j}^2}{\sigma_{\varepsilon,j}^2 + \sigma_{a,j}^2}\right)\ln a_i + \left(\frac{\sigma_{\varepsilon,j}^2}{\sigma_{\varepsilon,j}^2 + \sigma_{a,j}^2}\right)\mu_{a,j} + \ln\gamma + (1-\gamma)\ln k_{j,F^{NEW}} \\ &+ \frac{1}{2}\left(\frac{\sigma_{\varepsilon,j}^2\sigma_{a,j}^2}{\sigma_{\varepsilon,j}^2 + \sigma_{a,j}^2}\right) + \left(\frac{\sigma_{a,j}^2}{\sigma_{\varepsilon,j}^2 + \sigma_{a,j}^2}\right)\ln\varepsilon_i\end{aligned}$$

Similarly, her wage at her past firm is as follows.

$$\ln w_i^{OLD} = \left(\frac{\sigma_{a,j}^2}{\sigma_{\varepsilon,j,OLD}^2 + \sigma_{a,j}^2}\right) \ln a_i + \left(\frac{\sigma_{\varepsilon,j,OLD}^2}{\sigma_{\varepsilon,j,OLD}^2 + \sigma_{a,j}^2}\right) \mu_{a,j} + \ln \gamma + (1-\gamma) \ln k_{j,F^{OLD}} + \frac{1}{2}\left(\frac{\sigma_{\varepsilon,j,OLD}^2 \sigma_{a,j}^2}{\sigma_{\varepsilon,j,OLD}^2 + \sigma_{a,j}^2}\right) + \left(\frac{\sigma_{a,j}^2}{\sigma_{\varepsilon,j,OLD}^2 + \sigma_{a,j}^2}\right) \ln \varepsilon_i$$

This can be re-arranged to isolate ability.

$$\ln a_i^{OLD} = \left(\frac{\sigma_{\varepsilon,j,OLD}^2 + \sigma_{a,j}^2}{\sigma_{a,j}^2}\right) \ln w_i^{OLD} - \left(\frac{\sigma_{\varepsilon,j,OLD}^2}{\sigma_{a,j}^2}\right) \mu_{a,j} - \left(\frac{\sigma_{\varepsilon,j,OLD}^2 + \sigma_{a,j}^2}{\sigma_{a,j}^2}\right) \ln \gamma - (1-\gamma)\left(\frac{\sigma_{\varepsilon,j,OLD}^2 + \sigma_{a,j}^2}{\sigma_{a,j}^2}\right) \ln k_{j,F^{OLD}} - \frac{1}{2}\sigma_{\varepsilon,j,old}^2 - \ln \varepsilon_i^{OLD}$$

Since $\ln a_i = c_j + \rho \ln a_i^{OLD} + \ln \eta_i$, this allows us to write current ability as a function of the past wage.

$$\ln a_i = c_j + \rho \left[\left(\frac{\sigma_{\varepsilon,j,OLD}^2 + \sigma_{a,j}^2}{\sigma_{a,j}^2}\right) \ln w_i^{OLD} - \left(\frac{\sigma_{\varepsilon,j,OLD}^2}{\sigma_{a,j}^2}\right) \mu_{a,j} - \left(\frac{\sigma_{\varepsilon,j,OLD}^2 + \sigma_{a,j}^2}{\sigma_{a,j}^2}\right) \ln \gamma - (1-\gamma)\left(\frac{\sigma_{\varepsilon,j,OLD}^2 + \sigma_{a,j}^2}{\sigma_{a,j}^2}\right) \ln k_{j,F^{OLD}} - \frac{1}{2}\sigma_{\varepsilon,j,old}^2 - \ln \varepsilon_i^{OLD}\right] + \ln \eta_i$$

Finally, we can substitute this measure of ability into the equation for the wage at the current firm to obtain the following regression equation:

$$\ln w_i = \rho \left(\frac{\sigma_{\varepsilon,j,OLD}^2 + \sigma_{a,j}^2}{\sigma_{\varepsilon,j}^2 + \sigma_{a,j}^2}\right) \ln w_i^{OLD} + \alpha_{j,F^{OLD},} + \alpha_{j,F^{NEW},} + \nu_i$$

where:

$$\begin{aligned}
\alpha_{j,f^{NEW}} &= \rho \left[\left(\frac{\sigma_{\varepsilon,j}^2 - \sigma_{\varepsilon,j,OLD}^2}{\sigma_{\varepsilon,j}^2 + \sigma_a^2}\right) \mu_{a,j} + \left[1 - \left(\frac{\sigma_{\varepsilon,j,OLD}^2 + \sigma_{a,j}^2}{\sigma_{\varepsilon,j}^2 + \sigma_{a,j}^2}\right)\right] \ln \gamma + (1-\gamma)\, k_{j,F^{NEW}} \right. \\
&\quad \left. + \frac{1}{2}\left(\frac{\left(\sigma_{\varepsilon,j}^2 - \sigma_{\varepsilon,j,OLD}^2\right)\sigma_{a,j}^2}{\sigma_{\varepsilon,j}^2 + \sigma_{a,j}^2}\right)\right] + \left(\frac{\sigma_{a,j}^2}{\sigma_{\varepsilon,j}^2 + \sigma_{a,j}^2}\right) c_j \\
\alpha_{j,f^{OLD}} &= -\rho\,(1-\gamma)\left(\frac{\sigma_{\varepsilon,j,OLD}^2 + \sigma_{a,j}^2}{\sigma_{\varepsilon,j}^2 + \sigma_{a,j}^2}\right) \ln k_{j,F^{OLD}} \\
\nu_i &= \rho \left(\frac{\sigma_{a,j}^2}{\sigma_{\varepsilon,j}^2 + \sigma_{a,j}^2}\right) \left(\ln \varepsilon_i - \ln \varepsilon_i^{OLD}\right) - \left(\frac{\sigma_{a,j}^2}{\sigma_{\varepsilon,j}^2 + \sigma_{a,j}^2}\right) \ln \eta_i
\end{aligned}$$

It follows directly that the difference in coefficients will be as shown in the proposition. □

8.2 Derivations

8.2.1 Normalization of Worker Payoffs

A worker will apply for a job if the expected benefit is better than her outside option.

$$\begin{aligned}\left(1-F_i^W(s)\right)\xi(\delta,\psi)\,w_q+\left(1-F_i^W(s)\right)(1-\xi(\delta,\psi))\,w_u+F_i^W(s)\,\bar{U}>\bar{U}\\ \xi(\delta,\psi)\left(w_q-\bar{U}\right)-(1-\xi(\delta,\psi))\left(\bar{U}-w_u\right)>0\end{aligned}$$

Providing that $w_q-\bar{U}>0$ and $\bar{U}-w_u>0$, this amounts to the following condition.

$$\frac{w_q-\bar{U}}{\bar{U}-w_u}>\left(\frac{1-\delta}{\delta}\right)\tau(\psi)$$

The utility that the individual expects to get from investment is as follows.

$$\left(1-F_q^W(s)\right)\left[\delta\left(1-F_q^E\right)w_q+(1-\delta)\left(1-F_u^E\right)w_u\right]+\left[F_q^W(s)+\left(1-F_q^W(s)\right)\left[\delta F_q^E+(1-\delta)F_u^E\right]\right]\bar{U}-c$$

The utility from not investing is:

$$\left(1-F_u^W(s)\right)\left[\delta\left(1-F_q^E\right)w_q+(1-\delta)\left(1-F_u^E\right)w_u\right]+\left[F_u^W(s)+\left(1-F_u^W(s)\right)\left[\delta F_q^E+(1-\delta)F_u^E\right]\right]\bar{U}\Big].$$

The worker will invest if and only if the following condition holds.

$$\left(F_u^W-F_q^W\right)\left[\delta\left(1-F_q^E\right)\left(w_q-\bar{U}\right)+(1-\delta)\left(1-F_u^E\right)\left(w_u-\bar{U}\right)\right]>c$$

If we normalize the payoffs in this example by defining $\omega_q=w_q-\bar{U}$ and $\omega_u=\bar{U}-w_u$, these conditions exactly match those discussed in section 3.

8.2.2 Returns in Our Example

The probabilities that a worker sends an unclear signal if he did or did not invest are respectively $p_q=\frac{\theta_u-\theta_q}{1-\theta_q}$. Similarly, the probabilities that an employer sends an unclear signal if he did or did not invest are respectively $q_q=\frac{\psi_u-\psi_q}{1-\psi_q}$ and $q_u=\frac{\psi_u-\psi_q}{\psi_u}$. These can be used to derive return to investment for workers and employers.

With the parameter values we provide, these then collapse to the returns we discuss in section 2.1.

$$\beta_W = \begin{cases} \left(\frac{\theta_q}{\theta_u}\right) \cdot \left[\delta_j \omega_q - \left(\frac{\psi_u - \psi_q}{\psi_u}\right)(1-\delta_j)\,\omega_u\right] & \text{if } \delta_j \geq \hat{\delta}_j \text{ and } \pi_j \geq \hat{\pi}_j \\ \left(\frac{1-\theta_u}{1-\theta_q}\right) \cdot \left[\delta_j \omega_q - \left(\frac{\psi_u - \psi_q}{\psi_u}\right)(1-\delta_j)\,\omega_u\right] & \text{if } \delta_j \geq \hat{\delta}_j \text{ and } \pi_j < \hat{\pi}_j \\ \left(\frac{\theta_q}{\theta_u}\right) \cdot \left[\left(\frac{1-\psi_u}{1-\psi_q}\right)\delta_j \omega_q\right] & \text{if } \delta_j < \hat{\delta}_j \text{ and } \pi_j \geq \hat{\pi}_j \\ \left(\frac{1-\theta_u}{1-\theta_q}\right) \cdot \left[\left(\frac{1-\psi_u}{1-\psi_q}\right)\delta_j \omega_q\right] & \text{if } \delta_j < \hat{\delta}_j \text{ and } \pi_j < \hat{\pi}_j \end{cases}$$

$$\beta_E = \begin{cases} \lambda_j \left(\frac{\psi_q}{\psi_u}\right) \cdot \left[\pi_j \chi_q - \left(\frac{\theta_u - \theta_q}{\theta_u}\right)(1-\pi_j)\,\chi_u\right] & \text{if } \delta_j \geq \hat{\delta}_j \text{ and } \pi_j \geq \hat{\pi}_j \\ \lambda_j \left(\frac{\psi_q}{\psi_u}\right) \cdot \left[\left(\frac{1-\theta_u}{1-\theta_q}\right)\pi_j \chi_q\right] & \text{if } \delta_j \geq \hat{\delta}_j \text{ and } \pi_j < \hat{\pi}_j \\ \lambda_j \left(\frac{1-\psi_u}{1-\psi_q}\right) \cdot \left[\pi_j \chi_q - \left(\frac{\theta_u - \theta_q}{\theta_u}\right)(1-\pi_j)\,\chi_u\right] & \text{if } \delta_j < \hat{\delta}_j \text{ and } \pi_j \geq \hat{\pi}_j \\ \lambda_j \left(\frac{1-\psi_u}{1-\psi_q}\right) \cdot \left[\left(\frac{1-\theta_u}{1-\theta_q}\right)\pi_j \chi_q\right] & \text{if } \delta_j < \hat{\delta}_j \text{ and } \pi_j < \hat{\pi}_j \end{cases}$$

8.3 Further Extensions of the Model

8.3.1 Endogenous Wages

A. Ex-Post Bargaining

Consider the following modification to the model described in Section 2. Rather than payoffs from a match being fixed at $\{\omega_q, \omega_n, \chi_q, \chi_n\}$, we can add a third stage at which worker and firm investment decisions become common knowledge. To model the bargaining process, we assume for simplicity that total worker and firm payoffs are linear in monetary transfers that can be made between the two parties. The firm and worker investment decisions $i^W, i^E \in \{q, u\}$ determine the total surplus to be split, $x^{i^W i^E}$.[26] Workers receive a fixed fraction $\alpha \in (0,1)$ of this surplus.

Workers can, at the time of *application*, exercise a more valuable outside option $w_0^{i^W}$ if they invested than if they did not, with an equivalent assumption regarding the outside option for firms $x_0^{i^E}$. However, at the time of bargaining, the outside options of both parties are zero. To exactly replicate the payoff structure of our baseline model, further assume that the benefit to workers (resp. firms) from being matched to a good

[26] For example, a firm and worker that both invested split a surplus x^{qq}.

firm (resp. worker) is independent of their own investment decision. This allows us to define ω_q, ω_u, χ_q, χ_u.

$$\begin{aligned}
\omega_q &= \alpha x^{qq} - w_0^q = \alpha x^{uq} - w_0^u \geq 0 \\
\omega_u &= w_0^q - \alpha x^{qu} = w_0^u - \alpha x^{uu} \geq 0 \\
\chi_q &= (1-\alpha)\, x^{qq} - x_0^q = (1-\alpha)\, x^{uq} - x_0^u \geq 0 \\
\chi_u &= x_0^u - (1-\alpha)\, x^{qu} = x_0^u - (1-\alpha)\, x^{uu} \geq 0
\end{aligned}$$

Finally, if the lowest worker cost is $\underline{c} \geq w_0^q - w_0^u$ and the lowest firm cost is $\underline{k} \geq x_0^q - x_0^u$ then this structure exactly replicates our baseline model.

It is possible to relax some of these assumptions without any qualitative changes to the model. For example, if $(1-\alpha)\, x^{qq} - x_0^q \neq (1-\alpha)\, x^{uq} - x_0^u$ or $x_0^q - (1-\alpha)\, x^{qu} \neq x_0^u - (1-\alpha)\, x^{uu}$ then the firm hiring threshold would depend on whether the firm invested. This does not introduce any substantive change to our results. The restriction that firm and worker costs are bounded above zero is more important, but also sensible: without it, a worker would have an incentive to invest even if doing so never increased the probability of being hired.

B. Wage Equals Marginal Product

We assumed throughout the analysis that net worker and firm payoffs are exogenous parameters. We show above that this payoff structure can be rationalized by ex-post bargaining. Here, we instead explore the possibility of variable wage offers at the hiring stage. Specifically, we consider a simple benchmark model in which workers are paid their marginal product, although investment remains binary. The resulting policy implications are qualitatively the same as those of our baseline model.

Model

Begin by assuming that output (with a price normalized to one) is produced at constant returns to scale using a mass of qualified workers Q_j, combined with group-specific 'capital's provided by the firm K_j. For convenience, we adopt a Cobb-Douglas specification for each group.

$$Y_j = K_j^{1-\gamma} Q_j^{\gamma}$$

We can derive the average product of a worker by dividing by the total labor force $L_j = Q_j + U_j$ where Q_j is the mass of qualified workers, U_j is the mass of unqualified workers and $S_i = Q_j / L_j$ is the *share* of qualified workers.

$$\frac{Y_j}{L_j} = \left(\frac{K_j}{L_j}\right)^{1-\gamma} S_j^{\gamma}$$

Mirroring the assumptions of our baseline model, the firm can choose to provide exactly one unit or zero units of capital per worker for each group $j \in \{A, B\}$ so that $K_j/L_j \in \{0, 1\}$ depending on which investments the firm chooses to make. If the firm doesn't invest, no output is produced for that group.

Next, we can derive the marginal product of a worker. For a qualified group j worker at a firm that made the group j investment, the marginal product is:

$$MP_j = \gamma \left(\frac{Y_j}{Q_j} \right) = \gamma S_j^{\gamma - 1}.$$

The marginal product for a worker who did not invest is zero, since such workers never add value to production. This implies, given the same signal structure as in our baseline model, that a random worker's expected marginal product – which we assume is also the wage that a firm offers – is as follows.

$$\kappa \left(\pi_j, \theta \right) \gamma S_j^{\gamma - 1} = w \left(\pi_j, \theta \right) \geq 0$$

Assuming that workers have no outside option, they are always willing to accept this offer, since it is always weakly positive.

Aggregating up, the share of qualified workers is just π_j. This means that the average wage payment for group j is $\gamma \pi_j^{\gamma}$. Thus, the revenue that the firm earns per worker, net of wage payments, is $\lambda_j \left[(1 - \gamma) \pi_j^{\gamma} \right]$. The fraction of firms who invest is therefore given by:

$$\delta_j = G^E \left(\lambda_j \left[(1 - \gamma) \pi_j^{\gamma} \right] \right).$$

Since a fraction δ_j of firms made the group j investment, the return to investment for group j workers is simply δ_j multiplied by impact that worker investment has on the average wage offer. Thus, the fraction of workers of group j who invest is as follows.

$$\pi_j = G^W \left(\delta_j \int_0^1 w \left(\pi_j, \theta \right) \left[f_u^W \left(\theta \right) - f_q^W \left(\theta \right) \right] d\theta \right)$$

Clearly if $\delta_j = 0$, then the return to investment is zero. Similarly, if $\pi_j = 0$, then there is no return to investment because the wage is zero for any signal.

Discriminatory Equilibria

Since this model can have multiple equilibria, there is potential for an equilibrium in which there is zero investment by Bs and positive investment by As. If $\pi_B = 0$, firms would never offer a positive wage to B workers here, since $\kappa \left(\pi_j, \theta \right) = 0$. In turn, this means that there is never an incentive for workers to invest. Since hiring a B worker never adds to output, the return to B investment for firms is zero as well. Thus

$\delta_B = 0$. Turning to the A market, if $\pi_A > 0$, wages are positive and the fraction of A workers who invest is as follows.

$$G^W\left(\delta_A \int_0^1 w\left(\pi_A, \theta\right)\left[f_u^W\left(\theta\right) - f_q^W\left(\theta\right)\right] d\theta\right)$$

The return to investment for firms is also positive, and a fraction δ_A invest.

$$\delta_A = G^E\left(\lambda_A\left[\left(1-\gamma\right)\pi_A^{\beta}\right]\right)$$

We can prove that there can be such an equilibrium by positing a value of π_A, and a function G^E such that $\delta_A > 0$. For any such π_A and δ_A, there is some function G^W that that satisfies our assumptions and which yields the required worker investment levels.

$$\pi_A = G^W\left(\delta_A \int_0^1 w\left(\pi_A, \theta\right)\left[f_u^W\left(\theta\right) - f_q^W\left(\theta\right)\right] d\theta\right)$$

Symmetric Investment

We next examine the conditions under which a non-discriminatory equilibrium exists. Assume that $\pi_A = \pi_B = \pi$. This means that the fraction of workers who invest (for both types) is:

$$G^W\left(\delta \int_0^1 w\left(\pi, \theta\right)\left[f_u^W\left(\theta\right) - f_q^W\left(\theta\right)\right] d\theta\right).$$

If G^W is strictly increasing and $\delta_j > 0$, than this fraction can only be the same for both groups if $\delta_A = \delta_B = \delta$. However, the return to investment for firms is $\delta = G^E\left(\lambda_j\left[\left(1-\gamma\right)\pi^{\gamma}\right]\right)$. If G^E is strictly increasing, then firm investment levels cannot be the same for both groups unless $\lambda_A = \lambda_B$. This precludes $\pi_A = \pi_B = \pi$ if $\lambda_A \neq \lambda_B$, implying that an equilibrium with positive investment but no discrimination is impossible.

Affirmative Action

One definition of affirmative action in this model is a requirement that the average wage paid to workers, conditional on their being hired, is equal across groups.

$$\int_0^1 w\left(\pi_A, \theta\right) = \int_0^1 w\left(\pi_B, \theta\right)$$

This has many of the same problems as affirmative action in our baseline model. First, it does not eliminate the possibility of zero investment by Bs but positive investment by As, with no B workers receiving *any* wage offer. Under affirmative action, there is an equilibrium with $\pi_B = \delta_B = 0$ combined with any set of beliefs $\{\pi_A, \delta_A\}$ that constituted an equilibrium in the A market without affirmative action.

The second question we can ask is whether it is possible for this type of AA to lead to homogeneous

beliefs. First, note that $\pi_A = \pi_B = \pi$ implies that wages are identical across groups for every θ and that affirmative action does not bind.

$$w(\pi, \theta) = \kappa(\pi, \theta)\gamma\pi^{\gamma-1}$$

Assuming again that G^W and G^E are strictly increasing, a requirement for positive and equal rates of worker investment is again that $\delta_A = \delta_B = \delta$, which is only possible if $\lambda_A = \lambda_B$. Otherwise, affirmative action again has no prospect of eliminating discrimination in equilibrium.

Investment Insurance

Our main policy prescription, two-sided investment insurance, is similarly effective in the model with variable wages. Assume initially that Bs are a numerical minority ($\lambda_B \leq \lambda_A$) and that they are in an inferior equilibrium compared to As: $\pi_A > \pi_B$. This implies that wages are lower for this group.

$$\begin{aligned} w(\pi_A, \theta) &= \kappa(\pi_A, \theta)\gamma\pi_A^{\gamma-1} \\ &< \kappa(\pi_B, \theta)\gamma\pi_B^{\gamma-1} = w(\pi_B, \theta) \end{aligned}$$

Assuming that G^E is strictly increasing, it also implies that firm investment is lower.

$$\delta_A = G^E\left(\lambda_A\left[(1-\gamma)\pi_A^{\gamma}\right]\right) < G^E\left(\lambda_B\left[(1-\gamma)\pi_B^{\gamma}\right]\right) = \delta_B$$

As we did in our baseline model, imagine that the government has access to its signals of worker and firm investment: θ^g and ψ^g, which satisfy the same assumptions as θ and ψ. The government can use these signals to target potentially variable "wage" payments to workers, with similar incentive payments for firms. This must be effective here as well, because large enough wage payments can achieve any investment return for both workers and firms.

Consider the following policy, which will lead to immediate elimination of discrimination. First, set government wages $w^g(\theta)$ such that the fraction of Bs who invest is π_A.

$$\pi_{B,t} = G^W\left(\delta\int_0^1 \left[w(\pi_{B,t-1}, \theta) + w^g(\theta)\right]\left[f_u^W(\theta) - f_q^W(\theta)\right]d\theta\right) = \pi_A$$

Similarly, set weakly positive payments $p^g(\psi)$ firms such that the fraction who invest is δ_A.

$$\delta_{B,t} = G^E\left(\lambda_B\left[(1-\gamma)\pi_{B,t-1}^{\gamma}\right] + \int_0^1 p^g(\psi)\left[f_u^E(\psi) - f_q^E(\psi)\right]d\psi\right) = \delta_A$$

Once this has been achieved, firms will set $w(\pi_A, \theta) = w(\pi_B, \theta)$. This will ensure that $\pi_{B,t+1} = \pi_A$ with no subsidy, and it can be removed. The aggregate firm investment subsidy that is still required to maintain

equal firm investment returns is as follows.

$$\int_0^1 p^g(\psi)\left[f_u^E(\psi) - f_q^E(\psi)\right] d\psi = (\lambda_A - \lambda_B)\left[(1-\gamma)\pi^\gamma\right]$$

Thus, if $\lambda_A = \lambda_B$ then $p^g(\psi) = 0$ for all ψ: i.e., no investment subsidy is needed. Otherwise, some level of firm investment subsidy must be maintained to preserve an equilibrium without homogenous beliefs.

8.3.2 Marginal Firm Investment Costs

In our baseline model, we assumed that firms paid a fixed cost k_j for each investment that they chose to make. Here we provide intuition for an alternative case in which the cost of investing in group j is proportional to the number of workers from group j who end up being hired. With this change, there is no longer an inherent disadvantage to being a minority, but the model is otherwise qualitatively unchanged in most respects.

The investment cost is only paid for workers who apply and receive offers from the firm, which implies that the expected investment cost for a given group is as follows.

$$\lambda\left[1 - F_q^E(t^*(\delta))\right] \cdot \left[\pi\left(1 - F_q^W(s^*(\pi))\right) + (1-\pi)\left(1 - F_u^W(s^*(\pi))\right)\right] \cdot k$$

The gross returns to investment are unaffected. For any equilibrium without zero investment, the fraction of firms who invest is therefore as follows.

$$\delta = G^E\left(\frac{\lambda\left[F_u^E(t^*(\delta)) - F_q^E(t^*(\delta))\right] \cdot \left[\pi\left(1 - F_q^W(s^*(\pi))\right)\chi_q - (1-\pi)\left(1 - F_u^W(s^*(\pi))\right)\chi_u\right]}{\lambda\left[1 - F_q^E(t^*(\delta))\right] \cdot \left[\pi\left(1 - F_q^W(s^*(\pi))\right) + (1-\pi)\left(1 - F_u^W(s^*(\pi))\right)\right]}\right) \quad (10)$$

Compared to our baseline model, the net return is simply scaled up by the proportion of workers hired.

It is obvious from equation (10) that the returns to firm investment, and the fraction of firms who invest, are both independent of the population fraction λ. This is enough to conclude that for otherwise identical groups, the set of equilibria no longer depends on population size. In this sense, if investment costs are marginal, there is no inherent disadvantage to being a member of a minority group.

Aside from this point, the change in assumptions does not substantively alter the model, although this version is much less convenient to analyze. Compared to the case with fixed investment costs, firm investment returns are scaled up by a factor that varies with δ and π. As δ and π both approach one, the denominator in equation (10) is simply λ. At lower levels, returns are further scaled up, since costs are only paid for individuals who are hired.

The zero investment equilibrium clearly still exists and is stable. With the same regularity assumptions as we adopt for our existence proposition, the firm will set its signal threshold to one if π falls below some low but positive level. Similarly, if δ falls below some positive critical value, no workers apply. As long as

π and δ are low enough, there is therefore no incentive for any firm or worker to invest. Similarly, other equilibria may still exist, although stability and existence are much more complex to verify. This is already enough to conclude that modified versions of Propositions 1 and 2 continue to hold.

The logic behind the proof of Proposition 4 also remains intact, since the numerator of equation (10) is negative for a low enough value of value of π, while investment costs remain strictly positive. Perhaps most importantly, the logic of investment insurance (Proposition 8) is fundamentally unchanged.