The Effect of Occupational Choice and Stereotypes on Labor Market Sorting*

Oleg Muratov and Marc Möller

January 2025

Abstract

We incorporate competition for jobs into an assignment model to investigate the implications of occupational choice for the matching between heterogeneous workers and jobs of differing quality. When occupational choice is without frictions, more able workers choose more (costly) education and workers sort across occupations in a way that induces positive assortative matching. We characterize the distortions that arise when entry into an occupation is costly for a group of workers, e.g. due to the existence of stereotypes. The associated utility-loss is increasing with a worker's ability because, although high-ability workers obtain jobs of similar quality as in the absence of stereotypes, competition for those jobs turns out to be stronger.

Keywords: Labor market sorting; Positive assortative matching; Occupational choice; Stereotypes; Occupational segregation.
JEL *classification*: D31, D47, J21, J24.

1 Introduction

In the presence of complementarities between capital and labor, surplus is maximized when the most able workers are matched with the highest quality jobs, that is, when there is positive assortative matching (PAM). Becker's (1973) insight, that PAM emerges as an equilibrium outcome in a matching market without frictions can therefore be interpreted as a welfare theorem, ensuring the efficient allocation of productive resources. An implicit assumption of Becker's (1973) seminal work and most of the literature that followed is that workers compete in a *common* labor market for jobs belonging to a *single* "sector" or "occupation". In reality, workers' choice how much to invest into their careers is preceded by a decision which kind of career to pursue. The choice of a field of

*For valuable comments, we thank Ghazala Azmat, Igor Letina, Georg Nöldeke, Ernesto Rivera Mora, and Ron Siegel. Part of the research was carried out while Oleg was supported by Swiss National Science Foundation (Grant number 100018_185202). All errors are our own.

study or occupation is taken relatively early in life and is likely to depend on a worker's abilities as well as the opportunities workers expect in different sectors of the labor market.

Accounting for occupational choice, PAM might be more difficult to achieve, due to the existence of a coordination problem: A more able worker might obtain more education than a less able worker and yet fail to obtain the better job at a hospital due to the worker's choice to become an engineer and not a doctor. As occupational choice can be interpreted as the selection between alternative labor contests, workers of differing abilities may fail to sort across sectors that differ in *opportunity* (Azmat and Möller, 2018; Morgan et al., 2018). In fact, the view that "competitive contests do not automatically sort workers in ways that yield an efficient allocation of resources" dates back to Lazear and Rosen's (1981) seminal interpretation of careers as rank-order tournaments.[1]

In this paper we investigate the influence of occupational choice on labor market sorting. Our approach follows Olszewski and Siegel (2016) by modeling workers' competition for jobs within each sector as a "large contest". Sectors differ with respect to the distribution of "prizes" consisting of jobs of heterogeneous quality. Following the matching literature with non-transferable utility, we assume that a worker experiences an increase in utility from obtaining a better job and that this increase is larger for more able workers. Workers choose between two possible occupations and compete for the jobs in the corresponding sector by deciding how much of an occupation-specific education to acquire. Education is equally costly for both occupations but occupations may differ with respect to their distribution of job-qualities. Within each sector, higher education is rewarded with a better job, i.e. the assignment of jobs is assumed assortative with respect to education. Our interest lies on the equilibrium allocation of workers across the full set of jobs comprising both occupations, with a particular focus on whether this allocation is assortative with respect to workers' abilities.

We first show that Becker's (1973) welfare theorem remains valid, in that PAM constitutes the unique equilibrium allocation in a labor market *with* occupational choice. That is, occupational choice itself does not constitute a friction and the market resolves the workers' coordination problem leading to an efficient allocation of workers across jobs. This result is surprising in light of the aforementioned literature, casting doubt on the existence of sorting. In order to clarify the relation between PAM and sorting we offer a precise definition of *opportunity*, under which sorting with respect to opportunity becomes necessary and sufficient for PAM. More specifically, we show that when the distributions of jobs within each occupation are absolutely continuous, then sorting and PAM become equivalent if the distribution of jobs in one occupation dominates the distribution of jobs in the other in the sense of *Likelihood Ratio Ordering* (Shaked and Shanthikumar, 2007). Identification of this condition bridges a gap between the literature on matching in labor-markets (Becker, 1973) and sorting across labor-tournaments (Azmat and Möller, 2018; Morgan et al., 2018) by explaining the link between the related notions of PAM and sorting.

[1]Lazear and Rosen (1981) investigate the possibility of sorting across two endogenously formed wage structures offered by competitive but otherwise identical firms.

In the second part of our analysis we introduce the possibility that a subset of workers experiences a cost from participating in one of the two occupations. Such a cost may arise from the disutility certain workers experience from working in a job that fails to match their "identity" (Akerlof and Kranton, 2005), e.g. due to the existence of stereotypes. For instance, there exists evidence that women choose occupations that are oriented towards working with people, whereas men choose occupations oriented towards working with things (Kuhn and Wolter, 2022), explaining high degrees of occupational gender segregation (Charles and Grusky, 2005) in spite of considerable public efforts towards its reduction. Understanding the causes and implications of occupational segregation is of both academic and policy interest, especially when certain sectors, such as health care, experience a shortage in labor supply (Delfino, 2024).

Assuming that the distribution of abilities within the group of "constrained" workers is the same as within the group of "mobile" workers—who do not experience any costs of participation—we first show that when the distributions of jobs in the two occupations can be ranked in terms of the likelihood-ratio ordering, then PAM can be maintained as an equilibrium outcome if and only if the group of constrained workers is sufficiently small. Remarkably, the threshold above which PAM fails to emerge turns out to be independent of the *size* of the disutility that constrained workers experience from stereotypes. This finding is worrying because it implies that policies aiming at the elimination of job-market distortions via a reduction of stereotypes may prove ineffective unless they eliminate those stereotypes altogether.

That stereotypes constitute an obstacle for PAM is not surprising, as they induces workers to choose occupations not only on the basis of their individual abilities and the opportunities offered. Instead, our contribution lies in characterizing the precise nature of the induced distortions, and our results inform policy about the group of workers that needs to be targeted if occupational segregation is aimed at being reduced.

An important insight that emerges from our analysis is that the utility losses that constrained workers experience in the presence of stereotypes are increasing with their abilities. In particular, although assimilation may help constrained workers with high abilities to obtain jobs of identical quality as in the absence of stereotypes, the increase in competition for those jobs is sufficient to make high-ability workers the biggest losers.

2 Setup

We consider a labor market consisting of a continuum of workers with heterogeneous abilities and a continuum of jobs with heterogeneous quality. There are two sectors or *occupations*, $\mathcal{A}$ and $\mathcal{B}$. Within each occupation, workers compete by acquisition of costly education and the most educated workers get assigned to the highest quality jobs.[2] Firms play a passive role in our model, and the distribution of jobs across sectors is assumed exogenous. Instead, the focus of our model lies on the workers' choice of occupation and its implications for the overall allocation of workers across jobs.

More specifically, let the mass of workers be normalized to one and assume that, within each occupation, there exists a mass one of jobs. Assuming the number of jobs within each occupation to be sufficiently large to serve the overall number of workers in the market is innocuous because our model allows for a non-zero mass of jobs with zero quality. Workers who become matched with zero-quality jobs can be interpreted as remaining unemployed.

The ability of a worker is a random variable $X \in [0,1]$, absolutely continuously distributed. X is characterized by cumulative distribution $\mathcal{F}(x)$ and density $f(x) > 0$. We denote by σ_X the sigma-algebra generated by X. The quality of a job is a random variable $Y \in [0,1]$ and $\mathcal{H}_{\mathcal{A}}(y)$ and $\mathcal{H}_{\mathcal{B}}(y)$ are the cumulative distributions of jobs within each sector. $\mathcal{F}$, $\mathcal{H}_{\mathcal{A}}$, and $\mathcal{H}_{\mathcal{B}}$ are exogenous and constitute common knowledge.

Payoffs. A worker with ability x, who chooses occupation $i \in \{\mathcal{A}, \mathcal{B}\}$ and education level $e \geq 0$, obtaining a job with quality y in sector $j \in \{\mathcal{A}, \mathcal{B}\}$ has utility

$$U(x, y, e, i, j) = \mathbb{I}(i = j) \cdot x \cdot w(y) - e. \tag{1}$$

Here $\mathbb{I}(.)$ denotes the indicator function taking the value one when its argument is true and zero otherwise. It accounts for the assumption that education is *occupation-specific*, in that choosing the right *kind* of education is necessary to make a worker "employable" in a given sector. The function $w(.)$ is assumed to be continuously differentiable, non-negative, and strictly increasing. An important assumption, implicit in (1) is that an increase in job-quality is more beneficial to workers with higher abilities. In contrast, a worker's cost of education is given by $-e$, and is thus assumed to be independent of the worker's ability and occupational choice.[3] A worker's utility from choosing no education and remaining without employment is normalized to zero

Timing. In stage 1, all workers simultaneously choose an occupation $i \in \{\mathcal{A}, \mathcal{B}\}$ and a level $e \geq 0$ of occupation-specific education.[4] In stage 2, workers are assigned to the jobs in the sector corresponding to their occupational choice. Within each sector, the matching between workers and jobs is assumed to be assortative with respect to workers' education. That is, workers who chose more education are matched to jobs of higher quality.

[2] Such an assignment arises when firms have a preference for highly educated workers or when education serves as a signal for workers' abilities as in Hoppe et al. (2009) and Hopkins (2012).

[3] Results remain qualitatively unchanged when a worker's marginal cost of education is decreasing in his ability x.

[4] Assuming that workers choose their level of education after observing every other worker's occupational choice would have no effect on our results.

2.1 Strategies and equilibrium

A worker's strategy combines a choice of occupation with a level of education. Formally, let a measurable function

$$p_{\mathcal{A}} : [0,1] \to [0,1] \tag{2}$$

denote a worker's likelihood of choosing occupation $\mathcal{A}$ in dependence of his ability. As a worker's utility from being unemployed is normalized to zero and as workers are free to choose zero education, a worker who chooses occupation $\mathcal{A}$ with probability $p_{\mathcal{A}}(x)$ will choose occupation $\mathcal{B}$ with probability $p_{\mathcal{B}} = 1 - p_{\mathcal{A}}$. Note that because we consider a continuum of workers, an alternative interpretation is that any specific worker chooses *either* $\mathcal{A}$ *or* $\mathcal{B}$ and $p_{\mathcal{A}}(x)$ and $p_{\mathcal{B}}(x)$ denote the fraction of workers with type x choosing $\mathcal{A}$ or $\mathcal{B}$, respectively. To be consistent with either interpretation, we define next a worker's choice of education, *conditional* on his choice of occupation. For this purpose, let

$$\varepsilon_{\mathcal{A}} : [0,1] \to \mathbb{R}_+, \qquad \varepsilon_{\mathcal{B}} : [0,1] \to \mathbb{R}_+ \tag{3}$$

be a pair of measurable functions mapping a worker's ability into non-negative levels of occupation-specific education. Restricting attention to strategies that randomize occupational choice but not education (conditional on occupation) rather than considering joint probability distributions over $\{\mathcal{A}, \mathcal{B}\} \times [0, \infty)$ turns out to be inconsequential for the determination of equilibrium.

We focus on situations where workers strategies are symmetric, in that workers' of identical type, x, follow the same strategy. Then given workers' (symmetric) strategies $(p_{\mathcal{A}}, p_{\mathcal{B}}, \varepsilon_{\mathcal{A}}, \varepsilon_{\mathcal{B}})$ we can define for $i \in \{\mathcal{A}, \mathcal{B}\}$:

$$\Psi_i(e) = \int_0^1 \mathbb{I}\{\varepsilon_i(x) \leqslant e\} p_i(x) d\mathcal{F}(x). \tag{4}$$

Ψ_i is the cumulative distribution of education levels in occupation $i \in \{\mathcal{A}, \mathcal{B}\}$ that arises from workers' strategies. Now define, for each occupation $i \in \{\mathcal{A}, \mathcal{B}\}$, the *quantile function*

$$\mathcal{H}_i^{-1}(q) \equiv \inf\{y : \mathcal{H}_i(y) \geq q\}, \tag{5}$$

which returns, for any $q \in [0, 1]$, the minimal quality of a job in the highest quantile q of the distribution of jobs in sector i. The quantile function allows us to determine the quality $y_i(e)$ of the job a worker with education e gets matched with in sector i. Because within each sector matching is assumed assortative with respect to education, the mass of jobs with quality at least as large as

$y_i(e)$ has to be just enough to cover the mass of workers whose education is at least e, that is

$$y_i(e) = \mathcal{H}_i^{-1}\left(1 - \int \mathbb{I}\{e \leqslant \tilde{e}\} d\Psi_i(\tilde{e})\right). \tag{6}$$

Note that when the distribution of job-qualities in sector i has a strictly positive density then $\mathcal{H}_i$ is strictly increasing and the quantile function $\mathcal{H}_i^{-1}$ is simply the inverse of the CDF (justifying the slight abuse of notation). Finally, for our subsequent definition of equilibrium it is useful to define a worker's occupation-specific payoff function in dependence of his ability-type x:

$$V_i(x) = \max_e \left\{x \cdot w\left(y_i(e)\right) - e\right\}. \tag{7}$$

Equilibrium. With the help of these concepts, we can now define a symmetric equilibrium as a collection of mappings from workers' ability-types into occupational choice and occupation-specific education levels $(p_{\mathcal{A}}, p_{\mathcal{B}}, \varepsilon_{\mathcal{A}}, \varepsilon_{\mathcal{B}})$ that satisfy the following conditions:

$$p_{\mathcal{A}}(x) \in \arg\max_{p\in[0,1]} \{pV_{\mathcal{A}}(x) + (1-p)V_{\mathcal{B}}(x)\} \quad \text{and} \quad p_{\mathcal{B}}(x) = 1 - p_{\mathcal{A}}(x), \tag{8}$$

$$\varepsilon_i(x) = \arg\max_e \left\{x \cdot w\left(y_i(e)\right) - e\right\} \quad \text{for all} \quad i \in \{\mathcal{A}, \mathcal{B}\}. \tag{9}$$

Under the assumption that workers with equal ability behave identically, conditions (8) and (9) ensure that: (i) a worker enters an occupation only if he cannot obtain a strictly larger payoff from entering the alternative; and (ii) conditional on entering a specific occupation, a worker chooses education to maximize the difference between the benefit of obtaining a job with quality $y_i(e)$ and the cost of education e.

2.2 Benchmarking: Positive assortative matching

Given our objective to understand the effects of occupational choice on the allocation of workers across jobs, we now consider, as a benchmark, a setting where occupational choice is absent. In particular, we consider a combined labor market with only *one* occupation, where workers compete for *all* jobs in $\mathcal{A} \cup \mathcal{B}$. We maintain our assumption that in this combined market, workers are matched to jobs assortatively, with respect to their endogenously chosen education, but now education is to be seen as *universal*, equally relevant for all jobs in $\mathcal{A} \cup \mathcal{B}$. In the combined market, $\mathcal{A} \cup \mathcal{B}$, a mass one of jobs with the lowest quality will remain unfilled, which makes the CDF

$$\mathcal{H}_{\mathcal{A}\cup\mathcal{B}}(y) = \max\{\mathcal{H}_{\mathcal{A}}(y) + \mathcal{H}_{\mathcal{B}}(y) - 1, 0\} \tag{10}$$

of the jobs with qualities above

$$\underline{y} = \inf\{y : \mathcal{H}_{\mathcal{A}}(y) + \mathcal{H}_{\mathcal{B}}(y) = 1\} \tag{11}$$

our primitive. On the basis of the distribution of jobs in the combined labor market, we can now define the allocation of workers across jobs that satisfies *positive assortative matching* (PAM):

$$y^*(x) \equiv \mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\mathcal{F}(x)). \tag{12}$$

The allocation $y^*(x)$ is assortative in that it matches a worker's ability-quantile with the quality-quantile of the job he gets assigned to. Our first result shows that, when in the combined market the assignment of jobs is assortative with respect to education, then positive assortative matching with respect to ability arises as an equilibrium—*almost uniquely*—when workers compete for jobs by choosing education:

Proposition 1. *In the absence of occupational choice, the following mapping between a worker's ability and (universal) education constitutes an equilibrium of the workers' competition for jobs:*

$$\varepsilon_{\mathcal{A}\cup\mathcal{B}}(x) = x \cdot w(y^*(x)) - \int_0^x w(y^*(z))dz. \tag{13}$$

Because $\varepsilon_{\mathcal{A}\cup\mathcal{B}}$ is monotone increasing, the resulting allocation of workers across jobs satisfies PAM, i.e. $y_{\mathcal{A}\cup\mathcal{B}}(\varepsilon_{\mathcal{A}\cup\mathcal{B}}(x)) = y^*(x)$, and equilibrium payoffs are

$$V_{\mathcal{A}\cup\mathcal{B}}(x) = xw(y^*(x)) - \varepsilon_{\mathcal{A}\cup\mathcal{B}}(x) = \int_0^x w(y^*(z))dz. \tag{14}$$

Moreover, the equilibrium in (13) is essentially unique, i.e. for any symmetric but potentially mixed strategy equilibrium, represented by a family of CDFs $\{G(e;x)\}_{x\in[0,1]}$ it holds that $G(e;x) = \mathbb{I}\{\varepsilon_{\mathcal{A}\cup\mathcal{B}}(x) \leqslant e\}$ for almost all $x \in (0,1)$.

While the formal proof is relegated to the Appendix, the following "derivation" of the equilibrium in (13) is useful to build intuition. For this purpose, let $\varepsilon(.)$ be any (pure-strategy) equilibrium mapping between ability and education that is monotonically increasing (and thus implements PAM). Consider a worker with ability x who contemplates obtaining the job allocated in equilibrium to a worker of higher ability $\hat{x} > x$. Doing so requires higher education and for $\varepsilon(.)$ to be an equilibrium, the corresponding increase in the costs of education, $\varepsilon(\hat{x}) - \varepsilon(x)$, should exceed the worker's gain from being matched with the better job:

$$\varepsilon(\hat{x}) - \varepsilon(x) \geq x[w(y^*(\hat{x})) - w(y^*(x))]. \tag{15}$$

Reversing the argument, for a worker with ability $\hat{x}$ the cost-savings from a reduction in education,

$\varepsilon(\hat{x}) - \varepsilon(x)$, should not suffice to compensate for the worker's loss from being matched with the worse job:

$$\varepsilon(\hat{x}) - \varepsilon(x) \leq \hat{x}[w(y^*(\hat{x})) - w(y^*(x))]. \tag{16}$$

In combination, these two conditions provide upper and lower bounds on how much a worker's education can vary with his ability, for $\varepsilon(.)$ to be an equilibrium. Dividing both conditions by $\hat{x} - x$ and taking the limit $\hat{x} \to x$ they imply that

$$\frac{d\varepsilon(x)}{dx} = x \frac{dw(y^*(x))}{dx}. \tag{17}$$

Because the utility a worker derives from a job is normalized to zero when the worker has zero ability, such a worker must choose zero education in equilibrium, i.e. $\varepsilon(0) = 0$. It thus follows from the fundamental theorem of calculus and integration by parts that for $\varepsilon(.)$ to be an equilibrium it has to satisfy

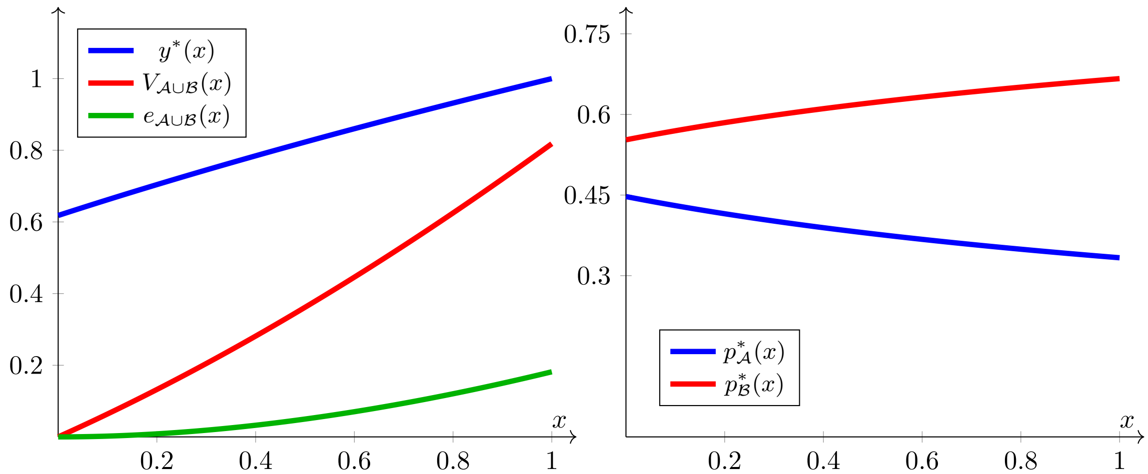
$$\varepsilon(x) = \varepsilon(0) + \int_0^x z \frac{dw(y^*(z))}{dx} dz = x \cdot w(y^*(x)) - \int_0^x w(y^*(z)) dz = \varepsilon_{\mathcal{A} \cup \mathcal{B}}(x). \tag{18}$$

In summary, this argument therefore shows that a symmetric, monotone mapping between ability and education must follow (13) to be an equilibrium. For an illustration of this equilibrium, see Figure below.

3 Occupational choice

In this section, we prove a "welfare theorem" guaranteeing the efficient allocation of workers across jobs. Our analysis in the previous section has shown that PAM emerges as the unique equilibrium outcome, when workers compete in a common labor market that rewards higher education with better jobs. However, when the labor market is segmented and workers must choose between different occupations, it is not clear whether the "invisible hand" induces workers to sort across occupations in a way that allows the most able workers to become matched with the highest quality jobs. The workers' self-selection into an occupation then resembles the choice between alternative "contests" and workers may fail to sort, in that more able workers are not necessarily more likely to enter the contest offering the better "opportunities" (Azmat and Möller, 2018, Morgan et al., 2018). Our subsequent analysis sheds light on the relation between PAM and sorting by offering a precise definition of opportunity for which sorting becomes necessary for PAM.

To formulate this section's main result, denote by $\overline{x}$ the lowest ability type among those workers who get matched to a job of highest quality $y = 1$ in the PAM allocation $y^*(x)$. Correspondingly,



(a) No occupational choice

(b) Occupational choice with free entry

Figure 1: **Benchmarks.** In the absence of occupational choice a worker with ability x obtains education $\varepsilon_{\mathcal{A}\cup\mathcal{B}}(x)$ and utility $V_{\mathcal{A}\cup\mathcal{B}}(x)$. Competition for jobs induces PAM between workers and jobs, i.e. $y_{\mathcal{A}\cup\mathcal{B}}(\varepsilon_{\mathcal{A}\cup\mathcal{B}}(x)) = y^*(x)$. The same outcome emerges when workers choose occupations but entry is free. Workers then enter occupation $\mathcal{A}$ with probability $p^*_{\mathcal{A}}(x)$ and occupation $\mathcal{B}$ with probability $p^*_{\mathcal{B}} = 1 - p_{\mathcal{A}}$. Plots assume $\mathcal{F}(x) = x$, $\mathcal{H}_{\mathcal{A}}(y) = y^2$, $\mathcal{H}_{\mathcal{B}}(y) = y$, and $w(y) = y$ which implies that $y^*(x) = \frac{1}{2}(\sqrt{5+4x} - 1)$ and $p^*_{\mathcal{A}}(x) = \frac{\sqrt{5+4x}-1}{\sqrt{5+4x}} = 1 - p^*_{\mathcal{B}}(x)$.

denote by $\underline{x}$ the highest type among those workers who get matched to a job featuring the lowest quality $y = \underline{y}$ of all jobs that get matched. Formally, let $\overline{x}$ and $\underline{x}$ be defined implicitly by:

$$\mathcal{F}(\overline{x}) = \lim_{y\uparrow 1} \mathcal{H}_{\mathcal{A}\cup\mathcal{B}}(y) \quad \text{and} \quad \mathcal{F}(\underline{x}) = \mathcal{H}_{\mathcal{A}\cup\mathcal{B}}(\underline{y}). \tag{19}$$

We then have the following:

Proposition 2. *Occupational choice induces no loss to allocative efficiency. PAM arises as an equilibrium in which workers choose occupation $\mathcal{A}$ with probability*

$$p^*_{\mathcal{A}}(x) = \begin{cases} \frac{1-\lim_{y\uparrow 1}\mathcal{H}_{\mathcal{A}}(y)}{2-\lim_{y\uparrow 1}(\mathcal{H}_{\mathcal{A}}(y)+\mathcal{H}_{\mathcal{B}}(y))}, & \text{if } x \geqslant \overline{x}, \\ \frac{h_{\mathcal{A}}(y^*(x))}{h_{\mathcal{A}}(y^*(x))+h_{\mathcal{B}}(y^*(x))}, & \text{if } x \in [\underline{x}, \overline{x}), \\ \frac{\mathcal{H}_{\mathcal{A}}(0)}{\mathcal{H}_{\mathcal{A}}(0)+\mathcal{H}_{\mathcal{B}}(0)}, & \text{if } x \leqslant \underline{x}, \end{cases} \tag{20}$$

and occupation $\mathcal{B}$ with probability $p^*_{\mathcal{B}} = 1 - p^*_{\mathcal{A}}$ and spend the same resources on education as in the benchmark without occupational choice, i.e. $\varepsilon_{\mathcal{A}}(x) = \varepsilon_{\mathcal{B}}(x) = \varepsilon_{\mathcal{A}\cup\mathcal{B}}(x)$. Moreover, PAM is essentially the unique equilibrium allocation of workers across jobs, i.e. in every equilibrium

$y_{\mathcal{A}}(x) = y_{\mathcal{B}}(x) = y^*(x)$ *almost everywhere.*[5]

The proof of Proposition 2 is constructive and is therefore presented here. The proof proceeds in two steps. Step 1 shows that when matching *within* each occupation is *assumed* to be assortative with respect to ability, then we can choose the workers' entry-probabilities $p_{\mathcal{A}}(x)$ and $p_{\mathcal{B}}(x)$ in such a way that in each of the two sectors, $y^*(x)$ emerges as the allocation of workers across jobs. Step 1 is "mechanical" in that it abstracts from workers' choices. Step 2 mirrors our benchmark analysis by showing that, within each occupation $i \in \{\mathcal{A}, \mathcal{B}\}$, workers' equilibrium choice of education is monotone increasing and thus implements assortative matching jobs with respect to workers' abilities. As a worker's equilibrium payoff is entirely determined by the allocation of jobs to workers of lower ability, and because allocations are identical across occupations by our choice of entry-probabilities in Step 1, i.e. $y_{\mathcal{A}} = y_{\mathcal{B}} = y^*$, a worker with ability x receives the same payoff, $V_{\mathcal{A}}(x) = V_{\mathcal{B}}(x) = \int_0^x w(y^*(z))dz$, in each occupation. This justifies why, in Step 1, we could choose entry-probabilities from the entire range $[0, 1]$ necessary for the construction of $p_{\mathcal{A}}(x)$.

Step 1: Assume that within each occupation, matching is assortative with respect to workers' abilities and replicate the PAM-allocation for each occupation, i.e. $y_{\mathcal{A}}(.) = y_{\mathcal{B}}(.) = y^*(.)$, by appropriate choice of entry-probabilities $p_{\mathcal{A}}(.)$ and $p_{\mathcal{B}}(.) = 1 - p_{\mathcal{A}}(.)$. Starting with the most talented workers, $x \geqslant \overline{x}$, let

$$p_{\mathcal{A}}(x) = \frac{1 - \lim_{y\uparrow 1} \mathcal{H}_{\mathcal{A}}(y)}{2 - \lim_{y\uparrow 1} (\mathcal{H}_{\mathcal{A}}(y) + \mathcal{H}_{\mathcal{B}}(y))} \quad \text{and} \quad p_{\mathcal{B}}(x) = \frac{1 - \lim_{y\uparrow 1} \mathcal{H}_{\mathcal{B}}(y)}{2 - \lim_{y\uparrow 1} (\mathcal{H}_{\mathcal{A}}(y) + \mathcal{H}_{\mathcal{B}}(y))}. \tag{21}$$

This way, the share of top-workers joining $\mathcal{A}$ is proportional to the size of the atom of $\mathcal{H}_{\mathcal{A}}$ at $y = 1$ relative to the total mass of such jobs across the two occupations, and the same holds for top-workers joining $\mathcal{B}$.

Next, for workers with ability $x \in (\underline{x}, \overline{x})$, recall that $y^*(x) = \mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\mathcal{F}(x))$, and let

$$p_{\mathcal{A}}(x) = \frac{h_{\mathcal{A}}(y^*(x))}{h_{\mathcal{A}}(y^*(x)) + h_{\mathcal{B}}(y^*(x))} \quad \text{and} \quad p_{\mathcal{B}}(x) = \frac{h_{\mathcal{B}}(y^*(x))}{h_{\mathcal{A}}(y^*(x)) + h_{\mathcal{B}}(y^*(x))}. \tag{22}$$

This way, a worker with ability x, who, according to PAM, should obtain a job with quality $y^*(x)$, joins each occupation with a probability that equals the relative frequency with which such jobs can be found in the corresponding occupation. It is as if each worker tacitly agrees to be matched to a job whose quality-rank in a ranking formed by all jobs in both sectors corresponds to the worker's ability-rank, and workers coordinate their occupation choice in such a way that the fraction of workers with ability x matches the fraction of jobs with quality $y^*(x)$ in each occupation.

[5] Multiplicity of equilibrium is due to potential atoms of the distribution of jobs at $y = 0$ and $y = 1$. For example, workers with abilities $x \geqslant \overline{x}$ may join occupation $\mathcal{A}$ with varying probabilities in an asymmetric equilibrium, as long as *on average*, they join with probability $p_{\mathcal{A}}$ given by (20).

Finally, when $\underline{y} = 0$ so that $\underline{x} > 0$, we can choose entry probabilities for the lowest types $x \leqslant \underline{x}$ arbitrarily because those workers will be matched with jobs of lowest quality $y = 0$ and thus obtain zero utility in either occupation. For the sake of being concrete, let types $x \leqslant \underline{x}$ mix their occupational choice in accordance with the size of the corresponding atoms at $y = 0$:

$$p_{\mathcal{A}}(x) = \frac{\mathcal{H}_{\mathcal{A}}(0)}{\mathcal{H}_{\mathcal{A}}(0) + \mathcal{H}_{\mathcal{B}}(0)} \quad \text{and} \quad p_{\mathcal{B}}(x) = \frac{\mathcal{H}_{\mathcal{B}}(0)}{\mathcal{H}_{\mathcal{A}}(0) + \mathcal{H}_{\mathcal{B}}(0)}. \tag{23}$$

On the basis of the entry-probabilities specified above, we are now ready to determine the expected assortative allocations of abilities within each occupation, $y_{\mathcal{A}}(x) = \mathcal{H}_{\mathcal{A}}^{-1}(\int_x^1 p_{\mathcal{A}}(z)d\mathcal{F}(z))$ and $y_{\mathcal{B}}(x) = \mathcal{H}_{\mathcal{B}}^{-1}(\int_x^1 p_{\mathcal{B}}(z)d\mathcal{F}(z))$. The following analysis considers occupation $\mathcal{A}$. For occupation $\mathcal{B}$ an analogous argument applies.

High types $x \geqslant \overline{x}$ get assigned to jobs with quality $y_{\mathcal{A}}(x) = 1$ because the number of jobs with highest quality in occupation $\mathcal{A}$, $1 - \lim_{y\uparrow 1} \mathcal{H}_{\mathcal{A}}(y)$, is sufficient to serve the measure of workers in $\mathcal{A}$ whose ability is at least x:

$$\begin{aligned} \int_x^1 p_{\mathcal{A}}(z)d\mathcal{F}(z) &= \frac{1 - \lim_{y\uparrow 1} \mathcal{H}_{\mathcal{A}}(y)}{2 - \lim_{y\uparrow 1} \left(\mathcal{H}_{\mathcal{A}}(y) + \mathcal{H}_{\mathcal{B}}(\tilde{y})\right)} \left[1 - \mathcal{F}(x)\right] \\ &\leq \frac{1 - \lim_{y\uparrow 1} \mathcal{H}_{\mathcal{A}}(y)}{1 - \lim_{y\uparrow 1} \mathcal{H}_{\mathcal{A}\cup\mathcal{B}}(y)} \left[1 - \mathcal{F}(\overline{x})\right] = 1 - \lim_{y\uparrow 1} \mathcal{H}_{\mathcal{A}}(y). \end{aligned} \tag{24}$$

Here the last equality follows from the definition of $\overline{x}$ in (19), and we have thus established that $y_{\mathcal{A}}(x) = y^*(x)$ for $x \geq \overline{x}$. Next consider workers with types $x \in (\underline{x}, \overline{x})$. The measure of workers in $\mathcal{A}$ with higher abilities than x is:

$$\begin{aligned} \int_x^1 p_{\mathcal{A}}(z)d\mathcal{F}(z) &= \int_x^{\overline{x}} \frac{h_{\mathcal{A}}(y^*(z))f(z)dz}{h_{\mathcal{A}}(y^*(z)) + h_{\mathcal{B}}(y^*(z))} + (1 - \mathcal{F}(\overline{x}))\frac{1 - \lim_{y\uparrow 1} \mathcal{H}_{\mathcal{A}}(y)}{1 - \lim_{y\uparrow 1} \mathcal{H}_{\mathcal{A}\cup\mathcal{B}}(y)} \\ &= \lim_{y\uparrow 1} \int_{y^*(x)}^{y} h_{\mathcal{A}}(\tilde{y})d\tilde{y} + 1 - \lim_{y\uparrow 1} \mathcal{H}_{\mathcal{A}}(y) = 1 - \mathcal{H}_{\mathcal{A}}(y^*(x)), \end{aligned} \tag{25}$$

where for the second equality, we changed the variable of integration:

$$dy^*(x) = d\mathcal{H}_{\mathcal{A}\cup\mathcal{B}}^{-1}(\mathcal{F}(x)) = \frac{f(x)dx}{h_{\mathcal{A}\cup\mathcal{B}}(y^*(x))}. \tag{26}$$

By the assortativeness of the allocation $y_{\mathcal{A}}$, the measure of workers with abilities above x in $\mathcal{A}$ equals the measure $1 - \mathcal{H}_{\mathcal{A}}(y_{\mathcal{A}}(x))$ of jobs with qualities above $y_{\mathcal{A}}(x)$ and we have thus established that $y_{\mathcal{A}}(x) = y^*(x)$ for $x \in (\underline{x}, \overline{x})$. Finally, if $\underline{y} = 0$ so that $\underline{x} > 0$, then by construction of $p_{\mathcal{A}}$ it holds that

$$\int_{\underline{x}}^1 p_{\mathcal{A}}(z)d\mathcal{F}(z) = 1 - \mathcal{H}_{\mathcal{A}}(y^*(\underline{x})) = 1 - \mathcal{H}_{\mathcal{A}}\left(0\right), \tag{27}$$

which means that there are not enough jobs with positive quality left in the assortative allocation $y_{\mathcal{A}}$ for types $x \leqslant \underline{x}$ to be matched with so that for those types $y_{\mathcal{A}}(x) = 0 = y^*(x)$ as required.

In summary, we thus have that $y_{\mathcal{A}}(x) = y^*(x)$, for all $x \in [0, 1]$, and replicating the argument for occupation $\mathcal{B}$ we can conclude that our choice of entry-probabilities $p^*_{\mathcal{A}}$ and $p^*_{\mathcal{B}}$ in implements PAM of workers across jobs by inducing PAM within *each* occupation. Figure 2 illustrates the construction of the entry-probabilities inducing PAM for a simple example, where the distributions of jobs in each sector has no atoms. Note from this figure the simple fact that PAM requires the occupation offering the better jobs to also attract more entry.

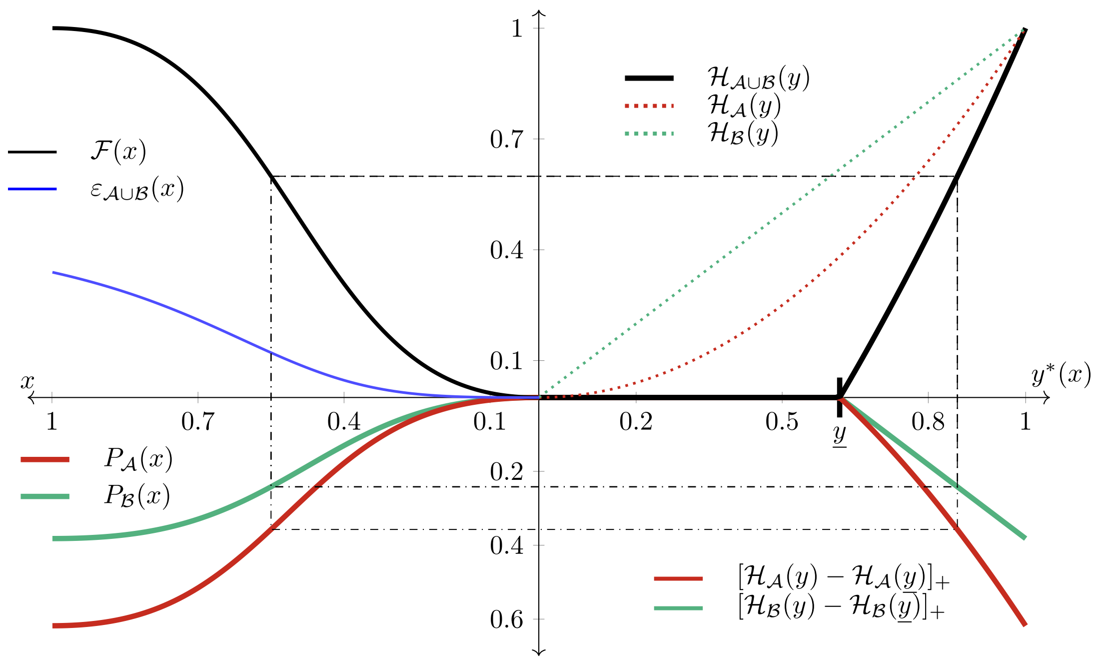


Figure 2: **Construction of equilibrium with occupational choice.** For any worker-type x select the job $y^*(x)$ that matches the mass of workers choosing higher education than $\varepsilon_{\mathcal{A}\cup\mathcal{B}}(x)$ (upper-left panel) with the overall mass of jobs with higher quality than $y^*(x)$ (upper-right panel). "Good jobs" ($y \geq \underline{y}$) get filled, "bad jobs ($y < \underline{y}$) remain unfilled. Then, for each occupation $i \in \{\mathcal{A}, \mathcal{B}\}$, match the mass of good jobs with qualities below $y^*(x)$ (lower-right panel) with the mass of workers with education lower than $\varepsilon_{\mathcal{A}\cup\mathcal{B}}(x)$ in sector i by selecting the corresponding value of $P_i(x) = \int_0^x p_i(z)f(z)dz$ (lower-left panel). Repeating this process for all x determines the entry-probabilities $p^*_{\mathcal{A}}$ and $p^*_{\mathcal{B}}$ inducing PAM. The plot assumes $\mathcal{F}(x) = \frac{x^2}{x^2+(1-x)^2}$, $\mathcal{H}_{\mathcal{A}} = y^2$, $\mathcal{H}_{\mathcal{B}} = y$, and $w(y) = y$.

Step 2: It remains to show that in each occupation the assortative allocation with respect to ability arises endogenously, from workers' equilibrium choice of education being monotonically increasing with their type, and that workers are willing to make the occupational choices prescribed by our selection of $p^*_{\mathcal{A}}$ and $p^*_{\mathcal{B}}$ in Step 1. For this purpose we can use the main insight from our benchmark analysis which has shown that the workers' competition for jobs induces an equilibrium mapping from abilities to education that is almost unique and that can be characterized completely via the corresponding assortative allocation of jobs. Given that by construction, these allocations are identical across sectors, i.e. $y_{\mathcal{A}}(.) = y_{\mathcal{B}}(.) = y^*(.)$, it holds that

$$\varepsilon_{\mathcal{A}}(x) = \varepsilon_{\mathcal{B}}(x) = xw(y^*(x)) - \int_0^x w(y^*(z))dz = \varepsilon_{\mathcal{A}\cup\mathcal{B}}(x), \tag{28}$$

that is, in each sector workers choose the same education as in the benchmark without occupational choice. Hence, even though sectors may differ with respect to the quality of their jobs, *workers' competition for jobs makes each occupation look identical from the viewpoint of an individual worker.* In each occupation, a worker with ability x expects to choose the level of education $\varepsilon_{\mathcal{A}\cup\mathcal{B}}(x)$ and to be assigned to a job with quality $y^*(x)$. This shows that, each worker, independently of the worker's ability, is indifferent between occupations $\mathcal{A}$ and $\mathcal{B}$, i.e.

$$V_{\mathcal{A}}(x) = V_{\mathcal{B}}(x) = \int_0^x w(y^*(z))dz = V(x), \tag{29}$$

and is therefore willing to choose among occupations as specified in Step 1.

3.1 Opportunity

The above analysis sheds light on a question that has attracted considerable interest in the literature on contests-selection. In this literature, models capturing the choice between alternative contests have focused on the issue of sorting, trying to understand the conditions under which contestants with higher ability will enter contests offering better "opportunity". A lack of tractable models capable of combining heterogeneous contestants with heterogeneous prizes has been a major obstacle. However, Olszewski and Siegel (2016) have shown that the behavior of contestants in "large contests" can be approximated by a model with a continuum of contestants. In fact by interpreting a worker's ability x as the inverse of a contestant's (constant) marginal cost of effort and the function $w(.)$ as a contestant's value for money y, our setting becomes a special case of their model. Our results thus have more general implications for contest-choice in "large markets". In particular, our analysis allows us to derive a precise definition of "opportunity" that induces sorting in the sense used by this literature when the number of potential contestants is large.

For this purpose, let us focus on the case of pairs of distributions which are absolutely continuous everywhere above $\underline{y}$. Remember that $\underline{y}$ denotes the lowest quality job that gets assigned to a worker under PAM. Let workers who choose positive education be denoted as *active* workers. For active workers, the probability of joining occupation $\mathcal{A}$ under the equilibrium described in

Proposition 2 can then be rewritten as

$$p_{\mathcal{A}}^*(x) = \frac{h_{\mathcal{A}}(y^*(x))}{h_{\mathcal{A}}(y^*(x)) + h_{\mathcal{B}}(y^*(x))} = \frac{\frac{h_{\mathcal{A}}(y^*(x))}{h_{\mathcal{B}}(y^*(x))}}{1 + \frac{h_{\mathcal{A}}(y^*(x))}{h_{\mathcal{B}}(y^*(x))}}. \tag{30}$$

Note that $y^*(x)$ is increasing in x and the expression $q/(1+q)$ is increasing in q for $q > 0$. Thus, $p_{\mathcal{A}}^*(x)$ is increasing in x if and only if the ratio $\frac{h_{\mathcal{A}}(y)}{h_{\mathcal{B}}(y)}$ is increasing in y for $y > \underline{y}$. Let $Y_{\mathcal{A}}|Y_{\mathcal{A}} > \underline{y}$ and $Y_{\mathcal{B}}|Y_{\mathcal{B}} > \underline{y}$ be the random variables characterizing the quality of jobs assigned to active workers in sectors $\mathcal{A}$ and $\mathcal{B}$, respectively. Then, using the definition of *Likelihood Ratio Order* (Shaked and Shanthikumar, 2007) we can state the following result:

Corollary 1. *Suppose that the distributions of jobs, $\mathcal{H}_{\mathcal{A}}$ and $\mathcal{H}_{\mathcal{B}}$, are absolutely continuous above $\underline{y}$. Then, for active workers, the probability of choosing occupation $\mathcal{A}$ is monotonically increasing in the worker's ability x if and only if $Y_{\mathcal{A}}|Y_{\mathcal{A}} > \underline{y}$ dominates $Y_{\mathcal{B}}|Y_{\mathcal{B}} > \underline{y}$ in the likelihood ratio order.*

Proposition 2 has shown that PAM arises in our model of occupational choice quite generically, i.e. without restrictions on the distribution of jobs within each occupation. In contrast, Corollary 1 assumes those distributions to be absolutely continuous and describes the conditions under which workers sort across the two occupations, in the sense that workers with higher ability are more likely to enter the occupation with more "opportunity". The corollary thereby clarifies the relation between PAM and sorting in the sense studied by the contest-selection literature. It shows that when occupations cannot be compared, in the sense of the likelihood-ratio ordering of their distributions of jobs, then sorting is not necessary for PAM, and, in fact, would prevent its implementation. Note, in particular, that when occupation $\mathcal{A}$ dominates occupation $\mathcal{B}$ in a weaker sense— for example, in the sense of first-order stochastic dominance—then there exist situations where PAM cannot emerge when the likelihood with which workers enter sector $\mathcal{A}$ increases monotonically with their ability. Corrollary 1 thus offers an implicit definition of "opportunity" for which workers' "sorting in accordance with opportunity" becomes equivalent to positive assortative matching.

4 Stereotypes

Our analysis so far has focused on the relation between ability and opportunity as the main driving force of workers' occupational choice. Abstracting from other factors that potentially influence this choice allowed for the conclusion that workers' competition for opportunity is powerful enough to induce an efficient allocation of workers across jobs. In this section, we consider *stereotypes* as a potential friction. Stereotypes are frequently named as a cause of occupational segregation, referring to an unbalanced composition of labor-market sectors with respect to a worker-characteristic, such as race or gender, that is orthogonal to ability. For example, Charles and Grusky (2005) document large degrees of occupational gender segregation that persist even in countries with considerable efforts towards their reduction. Occupational segregation is a worrying issue in light of evidence showing that race- or gender-differences in abilities—such as STEM-performance—are negligible.

Stereotypes are norms, beliefs, or expectations about how individuals belonging to a specific group ought to behave. It is, by now, a widely accepted view among economists that individuals derive utility or disutility from adhering to or deviating from behavior associated with their identity (Akerlof and Kranton, 2005). In this section we thus investigate the relevance of *stereotypes* as a friction affecting occupational choice, by assuming that some workers, belonging to a specific group, suffer a disutility from entering one of the two occupations.

Specifically, we now extend our model by assuming that, besides their direct utility U from entering occupation $i \in \{\mathcal{A}, \mathcal{B}\}$, given by (1), a share $s \in (0,1)$ of workers experience a disutility from choosing one of the two occupations, say occupation $\mathcal{A}$. For simplicity, we assume that the disutility from entering occupation $\mathcal{A}$ is the same for all workers who experience it. More specifically, we denote this disutility by $\zeta_{\mathcal{A}} \geq 0$ and assume that it is independent of the worker's ability. We will denote workers who experience the disutility as "constrained" and all remaining workers as "mobile". Throughout this section, every worker is thus characterized by a *two*-dimensional type: ability x, and whether the worker belongs to group C ("constrained") or M ("mobile"). Correspondingly, functions that depend on a worker's type will now have two arguments. For example, $p_{\mathcal{A}}(x; M)$ now denotes the probability of choosing occupation $\mathcal{A}$ for a worker with ability x from group M. Whether a worker is constrained or mobile is assumed to be independent of the worker's ability, or, in other words, the distribution of abilities within each group is assumed identical and equal to the overall distribution $\mathcal{F}(x)$ of abilities across workers.

To simplify exposition, our analysis in this section assumes that abilities are distributed uniformly, i.e. $\mathcal{F}(x) = x$, and that the distributions of jobs across sectors $\mathcal{H}_{\mathcal{A}}$ and $\mathcal{H}_{\mathcal{B}}$ are absolutely continuous everywhere, except, possibly, at $y = 0$, where atoms are admitted to allow for the possibility of "unemployment". For all $y \in (0,1)$, we will denote by $h_{\mathcal{A}}$ and $h_{\mathcal{B}}$ the corresponding PDFs. Finally, we will focus on situations where PAM emerges from the sorting of abilities with respect to opportunity as characterized in the previous section. Formally, we assume that above $y = 0$, the two distributions $\mathcal{H}_{\mathcal{A}}$ and $\mathcal{H}_{\mathcal{B}}$ can be ranked in the Likelihood Ratio Order.

Our first result shows that if the share of workers experiencing stereotypes is not significant enough, then the presence of stereotypes does not alter the labor-market's outcome, i.e. PAM arises, just as in the absence of stereotypes:

Lemma 1. *Let* $\mathcal{H}_{\mathcal{A}} \succsim_{LR} \mathcal{H}_{\mathcal{B}}$ *(*$\mathcal{H}_{\mathcal{B}} \succsim_{LR} \mathcal{H}_{\mathcal{A}}$*). If* $s \leqslant \frac{h_{\mathcal{B}}(1)}{h_{\mathcal{A}}(1)+h_{\mathcal{B}}(1)}$ *(*$s \leqslant \lim_{x \downarrow 0} \frac{h_{\mathcal{B}}(y^*(x))}{h_{\mathcal{A}}(y^*(x))+h_{\mathcal{B}}(y^*(x))}$*), then PAM emerges as an equilibrium allocation. In the corresponding equilibrium, workers choose education independently of occupation and as in the benchmark,* $\varepsilon(x; M) = \varepsilon(x; C) = \varepsilon_{\mathcal{A} \cup \mathcal{B}}(x)$*. Constrained workers choose occupation* $\mathcal{B}$ *in accordance with stereotypes, i.e.* $p_{\mathcal{B}}(x; C) = 1$*, and* *mobile workers mix between occupations, by choosing $\mathcal{A}$ with probability*

$$p_{\mathcal{A}}(x; M) = \frac{h_{\mathcal{A}}(y^*(x))}{[h_{\mathcal{A}}(y^*(x)) + h_{\mathcal{B}}(y^*(x))](1-s)} \in (0,1), \tag{31}$$

and $\mathcal{B}$ with probability $p_{\mathcal{B}}(x; M) = 1 - p_{\mathcal{A}}(x; M)$.

Note that in the equilibrium described in Lemma 1, the occupational choice of mobile workers is sufficient to preserve PAM because mobile workers can adjust their behavior in such a way that, for any given ability-type x, the overall share of workers in each sector coincides with the equilibrium allocation of abilities in the absence of stereotypes, $p^*_{\mathcal{A}}$ and $p^*_{\mathcal{B}}$, given by (30):

$$(1-s)p_{\mathcal{A}}(x; M) = p^*_{\mathcal{A}}(x) \quad \text{and} \quad s + (1-s)p_{\mathcal{B}}(x; M) = p^*_{\mathcal{B}}(x). \tag{32}$$

The upper bound on s in Lemma 1 arises from the requirement that $p_{\mathcal{A}}(x; M) \in [0,1]$ and the observation that due to likelihood ratio dominance, $p^*_{\mathcal{A}}(.)$ is strictly monotone. Note that when stereotypes affect the dominating (dominated) occupation, the upper bound on s is determined by the relative frequency of the most (least) productive jobs $\frac{h_{\mathcal{A}}(1)}{h_{\mathcal{B}}(1)}$ $(\lim_{x\to 0} \frac{h_{\mathcal{A}}(y^*(x))}{h_{\mathcal{B}}(y^*(x))})$.

Lemma 1 identifies conditions under which the existence of stereotypes has no effect on allocative efficiency. We will now turn our attention to the distortions that arise when these conditions fail to be satisfied. In doing so, we focus on the more interesting case, where stereotype-induced entry costs exist for the occupation offering better opportunities. This case is the more interesting, because it induces constrained workers to trade off better opportunities against the disutility from choosing an occupation subject to a stereotype. Formally, we assume for the remainder that $\mathcal{H}_{\mathcal{A}} \succsim_{\text{LR}} \mathcal{H}_{\mathcal{B}}$ and $s > \frac{h_{\mathcal{B}}(1)}{h_{\mathcal{A}}(1)+h_{\mathcal{B}}(1)}$.

In preparing for this section's main result it turns out to be useful to think about the allocation of jobs that would arise if there was *full segregation (FS)* of workers—with all mobile workers entering sector $\mathcal{A}$ and all constrained workers entering sector $\mathcal{B}$—and within each sector, matching was assortative. Denoting by $y_i^{FS}(x; s)$ the job obtained by a worker with ability x in sector $i \in \{\mathcal{A}, \mathcal{B}\}$ under full segregation, and setting equal the mass of workers with ability above x with the mass of jobs with qualities above $y_i^{FS}(x; s)$ in each sector then leads to

$$(1-s)(1-x) = 1 - \mathcal{H}_{\mathcal{A}}(y_{\mathcal{A}}^{FS}(x; s)) \quad \text{and} \quad s(1-x) = 1 - \mathcal{H}_{\mathcal{B}}(y_{\mathcal{B}}^{FS}(x; s)) \tag{33}$$

or, equivalently

$$y_{\mathcal{A}}^{FS}(x; s) \equiv \mathcal{H}_{\mathcal{A}}^{-1}(s + x - sx) \quad \text{and} \quad y_{\mathcal{B}}^{FS}(x; s) \equiv \mathcal{H}_{\mathcal{B}}^{-1}(1 - s + sx). \tag{34}$$

We now argue that under full segregation, high-ability workers get better jobs in sector $\mathcal{A}$ whereas low-ability workers get better jobs in sector $\mathcal{B}$. Formally, there exists a unique $x^{FS}(s) \in [0, 1)$, possibly equal to zero, such that $y_{\mathcal{A}}^{FS}(x;s) < y_{\mathcal{B}}^{FS}(x;s)$ for all $x \in [0, x^{FS})$ and $y_{\mathcal{A}}^{FS}(x;s) > y_{\mathcal{B}}^{FS}(x;s)$ for all $x \in (x^{FS}, 1)$.

To see this, note first that the job that a worker with the lowest type obtains under full separation in $\mathcal{A}$ is worse than $\underline{y}$ if and only if $\mathcal{H}_{\mathcal{A}}(y_{\mathcal{A}}^{FS}(0;s)) = s < \mathcal{H}_{\mathcal{A}}(\underline{y})$. Similarly, the job this worker obtains in $\mathcal{B}$ is better than $\underline{y}$ under an equivalent condition, i.e. if and only if $\mathcal{H}_{\mathcal{B}}(y_{\mathcal{B}}^{FS}(0;s)) = 1 - s > \mathcal{H}_{\mathcal{B}}(\underline{y}) = 1 - \mathcal{H}_{\mathcal{A}}(\underline{y})$. It follows that under full segregation, a worker with the lowest type obtains a better job in $\mathcal{B}$ than in $\mathcal{A}$, i.e. $y_{\mathcal{A}}^{FS}(0;s) < y_{\mathcal{B}}^{FS}(0;s)$ if and only if $s < \mathcal{H}_{\mathcal{A}}(\underline{y})$. In contrast, workers with sufficiently high ability must obtain (weakly) better jobs in $\mathcal{A}$ than in $\mathcal{B}$. This holds trivially for the highest types, who obtain the best possible job, $y_{\mathcal{A}}^{FS}(1;s) = y_{\mathcal{A}}^{FS}(1;s) = 1$, independently of their occupation, and it remains true for workers with sufficiently high abilities because

$$\frac{d}{dx}[y_{\mathcal{A}}^{FS}(x;s) - y_{\mathcal{B}}^{FS}(x;s)]|_{x=1} = \frac{1-s}{h_{\mathcal{A}}(1)} - \frac{s}{h_{\mathcal{B}}(1)} < 0 \Leftrightarrow s > \frac{h_{\mathcal{B}}(1)}{h_{\mathcal{A}}(1) + h_{\mathcal{B}}(1)}. \tag{35}$$

Finally, it follows from the fact that $\mathcal{A}$ dominates $\mathcal{B}$ in the likelihood ratio order that the derivative of $y_{\mathcal{A}}^{FS}(x;s) - y_{\mathcal{B}}^{FS}(x;s)$ can change sign at most once, which shows that in our setting, the assumption of likelihood order dominance amounts to a single-crossing property. It implies that there must exist a unique critical type of worker $x^{FS}(s) \in [0,1)$ with the properties stated above and our analysis has shown that $x^{FS}(s) > 0$ if and only if $s < \mathcal{H}_{\mathcal{A}}(\underline{y})$. The existence of such a unique $x^{FS}(s) \in [0,1)$ then allows us to define, for any given $s > \frac{h_{\mathcal{B}}(1)}{h_{\mathcal{A}}(1)+h_{\mathcal{B}}(1)}$, a threshold $\overline{\zeta}_{\mathcal{A}}(s)$ on the entry-cost of occupation $\mathcal{A}$ that will play a critical role for our subsequent characterization of equilibrium:

$$\overline{\zeta}_{\mathcal{A}}(s) \equiv \int_{x^{FS}(s)}^{1} \left(w(y_{\mathcal{A}}^{FS}(z;s)) - w(y_{\mathcal{B}}^{FS}(z;s)) \right) dz. \tag{36}$$

Proposition 3. *Suppose that occupation $\mathcal{A}$, where constrained workers experience disutility $\zeta_{\mathcal{A}}$ due to stereotypes, offers better opportunities, i.e. $\mathcal{H}_{\mathcal{A}} \succsim_{LR} \mathcal{H}_{\mathcal{B}}$. Further assume that the fraction of constrained workers is sufficiently large, i.e. $s > \frac{h_{\mathcal{B}}(1)}{h_{\mathcal{A}}(1)+h_{\mathcal{B}}(1)}$. Then PAM cannot be an equilibrium outcome and distortions to allocative efficiency can be characterized as follows:*

1. ***Segregation and Diversification**: For $\zeta_{\mathcal{A}} \geq \overline{\zeta}_{\mathcal{A}}(s)$, all constrained workers segregate by choosing occupation $\mathcal{B}$ in accordance with stereotypes. Mobile workers with high ability, $x \geq x^{FS}(s)$, choose occupation $\mathcal{A}$ offering better opportunity whereas mobile workers with lower ability diversify by mixing between occupations with*

$$p_{\mathcal{A}}(x;M) = \frac{h_{\mathcal{A}}(y^*(x))}{h_{\mathcal{A}}(y^*(x)) + h_{\mathcal{B}}(y^*(x))} \frac{1}{1-s} \in (0,1). \tag{37}$$

Diversification induces PAM for low-ability workers, i.e. $y(x;C) = y(x;M) = y^*(x)$ for $x \in [0, x^{FS}(s)]$, whereas high-ability workers, $x \in (x^{FS}(s), 1]$, obtain different jobs than under

PAM: better jobs for the mobile; $y(x;M) = \mathcal{H}_{\mathcal{A}}^{-1}(s + x - sx) > y^*(x)$*, worse jobs for the constrained,* $y(x;C) = \mathcal{H}_{\mathcal{B}}^{-1}(1 - s + sx) < y^*(x)$*. Distortions to PAM are experienced by all workers, i.e.* $x^{FS} = 0$*, if and only if* $s > \mathcal{H}_{\mathcal{A}}(\underline{y})$*.*

2. ***Assimilation:*** *For* $\zeta_{\mathcal{A}} < \overline{\zeta}_{\mathcal{A}}(s)$*, constrained workers with low abilities* $x \leq \overline{x} \in (0,1)$ *choose occupation* $\mathcal{B}$ *whereas constrained workers with higher ability assimilate, by mixing between occupations with*

$$p_{\mathcal{A}}(x;C) = 1 - \frac{h_{\mathcal{B}}(y^*(x))}{h_{\mathcal{A}}(y^*(x)) + h_{\mathcal{B}}(y^*(x))}\frac{1}{s} \in (0,1). \tag{38}$$

Mobile workers with high abilities, $x \geq \underline{x} \in [0, \overline{x})$*, choose occupation* $\mathcal{A}$ *whereas mobile workers with lower abilities diversify by mixing between occupations with* $p_{\mathcal{A}}(x;M)$ *given by (37). Assimilation and diversification induce PAM for high-ability and low-ability workers, i.e.* $y(x;C) = y(x;M) = y^*(x)$ *for* $x \in [0, \underline{x}] \cup [\overline{x}, 1]$*, but distortions arise for workers with abilities in an intermediate range,* $x \in (\underline{x}, \overline{x})$*, where mobile workers obtain better jobs and constrained workers obtain worse jobs than under PAM:*

$$y(x;M) = \mathcal{H}_{\mathcal{A}}^{-1}\left(\mathcal{H}_{\mathcal{A}}(y^*(\overline{x})) - (1-s)(\overline{x} - x)\right) > y^*(x) \tag{39}$$

$$y(x;C) = \mathcal{H}_{\mathcal{B}}^{-1}\left(\mathcal{H}_{\mathcal{B}}(y^*(\overline{x})) - s(\overline{x} - x)\right) < y^*(x). \tag{40}$$

Under both, segregation and assimilation, the allocations of jobs $y_i(x)$ *within each sector* $i \in \{\mathcal{A}, \mathcal{B}\}$ *are assortative and supported by the educational choices*

$$\varepsilon_i(x;j) = xw(y_i(x)) - \int_0^x w(y_i(z))dz, \quad j \in \{C, M\}. \tag{41}$$

Finally, the threshold $\overline{\zeta}_{\mathcal{A}}(s)$ *is strictly increasing and converges to zero for* $s \to \frac{h_{\mathcal{B}}(1)}{h_{\mathcal{A}}(1)+h_{\mathcal{B}}(1)}$ *implying that, qualitatively, the labor-market outcome is as depicted in the example in Figure 3.*

Proposition 3 characterizes the allocational distortions to PAM that arise from the presence of stereotypes in the dominating occupation. When the disutility $\zeta_{\mathcal{A}}$ associated with stereotypes in occupation $\mathcal{A}$ is above the threshold $\overline{\zeta}_{\mathcal{A}}(s)$, constrained workers segregate by entering occupation $\mathcal{B}$. Below that threshold, constrained workers with high abilities assimilate by mixing between both occupations. Their assimilation has an important effect on the allocation of jobs. In particular, assimilation enables PAM between workers with high abilities and jobs with high quality. Such "matching at the top" is not possible when stereotypes are too strong, because for $\zeta_{\mathcal{A}} > \overline{\zeta}_{\mathcal{A}}(s)$ only mobile workers are willing to mix between occupations but mobile workers mix only when their ability is low. Under segregation, either every worker gets matched with a different job than under PAM, or there is only "matching at the bottom". In summary, while diversification of mobile workers may help to induce matching at the bottom, it is only through the assimilation of constrained workers that matching at the top becomes a possibility. Note that this feature is a direct consequence of the fact that the likelihood ratio dominance of sector $\mathcal{A}$ over sector $\mathcal{B}$ induces workers' preference for occupation $\mathcal{A}$ over occupation $\mathcal{B}$ to be monotonically increasing in a worker's ability.

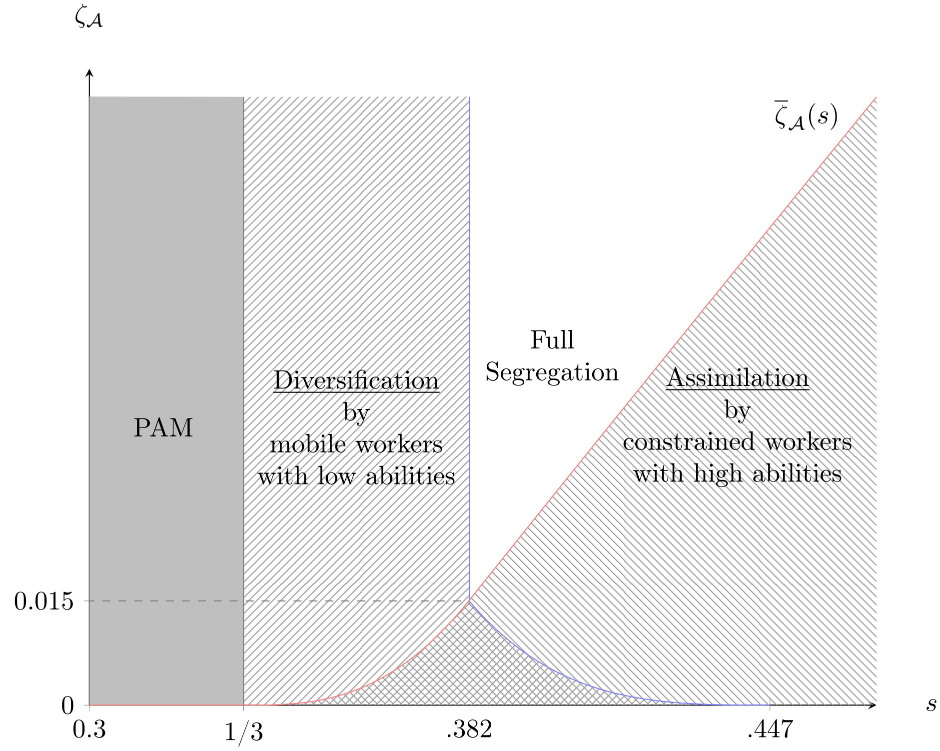


Figure 3: **Characterization of occupational choice with stereotypes**. The plot assumes $F(x) = x$, $w(y) = y$, $\mathcal{H}_{\mathcal{A}} = y^2$, and $\mathcal{H}_{\mathcal{B}} = y$. Distortions to PAM exist when the mass of constrained workers, s, is larger than $\frac{h_{\mathcal{B}}(1)}{h_{\mathcal{A}}(1)+h_{\mathcal{B}}(1)} = \frac{1}{3}$. Diversification of mobile workers induces low ability workers to obtain the same jobs as under PAM. Assimilation of constrained workers induces high ability workers to obtain the same jobs as under PAM. When stereotypes are sufficiently strong, i.e. $\zeta_{\mathcal{A}} > 0.015$, there exists a parameter range $[\underline{s}, \overline{s}]$ where neither diversification nor assimilation occur and the labor market is fully segregated: all mobile workers choose the dominating occupation $\mathcal{A}$ and obtain better jobs than under PAM whereas all constrained workers choose the dominated occupation $\mathcal{B}$ and obtain worse jobs.

Proposition 3 also shows that full segregation between mobile and constrained workers can arise as an equilibrium although only *one* of these groups of workers experiences a disutility from stereotypes. When the fraction of constrained workers is sufficiently large, i.e. $s > \mathcal{H}_{\mathcal{A}}(\underline{y})$, and stereotypes are strong, i.e. $\zeta_{\mathcal{A}} > \overline{\zeta}_{\mathcal{A}}$, all mobile workers are driven towards occupation $\mathcal{A}$ due to the exclusive presence of constrained workers in occupation $\mathcal{B}$. Note that in this case, the distortions that occupational choice induces to PAM are extreme, in the sense that not a single worker obtains the job that he would be allocated in the benchmark without occupational choice. This makes precise our statement from the introduction, that, in the presence of stereotypes, occupational choice constitutes an obstacle for the efficient allocation of workers across jobs, and it shows that the associated distortions can be significant. Under full segregation *all* mobile workers obtain better jobs than under PAM whereas *all* constrained workers obtain worse jobs.

So what are the consequences of the allocational distortions characterized by Proposition 3 for workers' utilities, in particular, for those experiencing stereotypes? One might expect and it has been argued that those workers with the highest abilities are immune to the effects of stereotypes, because jobs of the highest quality exist in both occupations. However, a straightforward consequence of Proposition 3 is the following:

Corollary 2. *Under the conditions of Proposition 3, stereotypes make constrained workers worse off, no matter whether they assimilate or not, and the associated loss in utility,*

$$\Delta V(x, C) = \int_0^x [w(y^*(z)) - w(y(z, C))]dz, \tag{42}$$

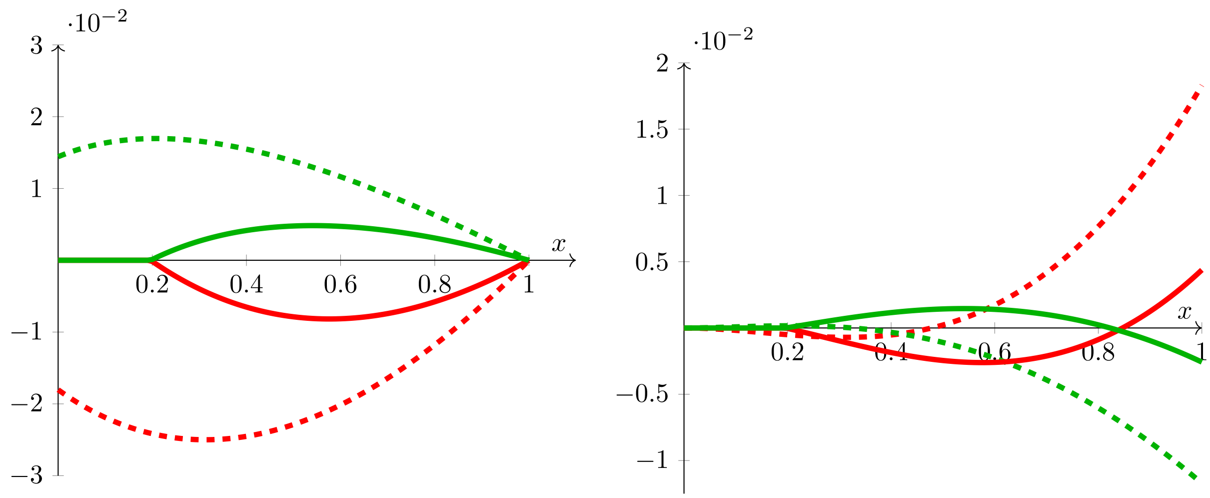
is increasing with a worker's ability x. In particular, although constrained workers with high abilities obtain jobs of similar or even equal quality as under PAM they suffer most from the presence of stereotypes, because fiercer competition for these jobs induces them to spend more resources on education, i.e. there exists $\hat{x} \in [0, \overline{x})$ such that $\epsilon_i(x; C) > \epsilon^*(x)$ for all $x > \hat{x}$.

Corollary 2 is a direct consequence of the fact that in the presence of stereotypes, there always exists a non-empty subset of constrained workers who obtain jobs of lower quality than under PAM. The corollary shows that while allocational effects might be restricted to constrained workers of lower abilities, the consequences for utilities are universal, and, in fact, strongest for the workers with the highest abilities. The reason is that constrained workers with high abilities face stronger competition for high quality jobs which are relatively rare in the dominated sector and respond by increasing their educational efforts. We further show in the proof of Corollary 2 that, under the additional assumption that $w(.)$ is concave, this increase in competition becomes amplified towards the top, in that $\epsilon_i(x; C) - \epsilon^*(x)$ is monotone increasing above a certain ability-level. However, we also provide an example where $\epsilon_i(x; C) - \epsilon^*(x) > 0$ for *all* $x \in [0, 1]$, showing that the increase in competition amongst constrained workers due to stereotypes may affect even workers of the lowest type.

Illustrative example

To illustrate the results in this section, we return to the example used in Figure 1 where $\mathcal{F}(x) = x$, $\mathcal{H}_\mathcal{A} = y^2$, $\mathcal{H}_\mathcal{B} = y$, and $w(y) = y$. Remember that in the absence of stereotypes, competition for jobs induces PAM (Proposition 2), occupational choice is given by $(p^*_\mathcal{A}(x), p^*_\mathcal{B}(x))$, and workers choose education and obtain payoffs as depicted in Figure 1. Because $\frac{h_\mathcal{A}(y)}{h_\mathcal{B}(y)} = 2y$ is monotone increasing, $\mathcal{H}_\mathcal{A}$ dominates $\mathcal{H}_\mathcal{B}$ in the likelihood ratio order, and it thus follows from Lemma 1, that the same outcome emerges when the mass of workers constrained by stereotypes is low, i.e. $s \leq \frac{h_\mathcal{B}(1)}{h_\mathcal{A}(1)+h_\mathcal{B}(1)} = \frac{1}{3}$. For the remaining cases, $s > \frac{1}{3}$, the characterization of equilibrium as described by Proposition 3 is depicted in Figure 3.

Segregation and Diversification. When stereotypes are strong, i.e. $\zeta_\mathcal{A} > \overline{\zeta}_\mathcal{A}(s)$, all constrained workers choose occupation in accordance with stereotypes by entering sector $\mathcal{B}$. While diversification of mobile workers with low ability may induce PAM at the bottom of the ability distribution, i.e. $y(x; M) = y(x; C) = y^*(x)$ for all $x < \underline{x} = \frac{1-3s+s^2}{s^2}$, all workers with abilities higher than $\underline{x}$ get different jobs than under PAM: better jobs for the mobile, $y(x; M) = \sqrt{s + (1-s)x} > y^*(x)$; worse jobs for the constrained, $y(x; C) = 1 - s + sx < y^*(x)$.



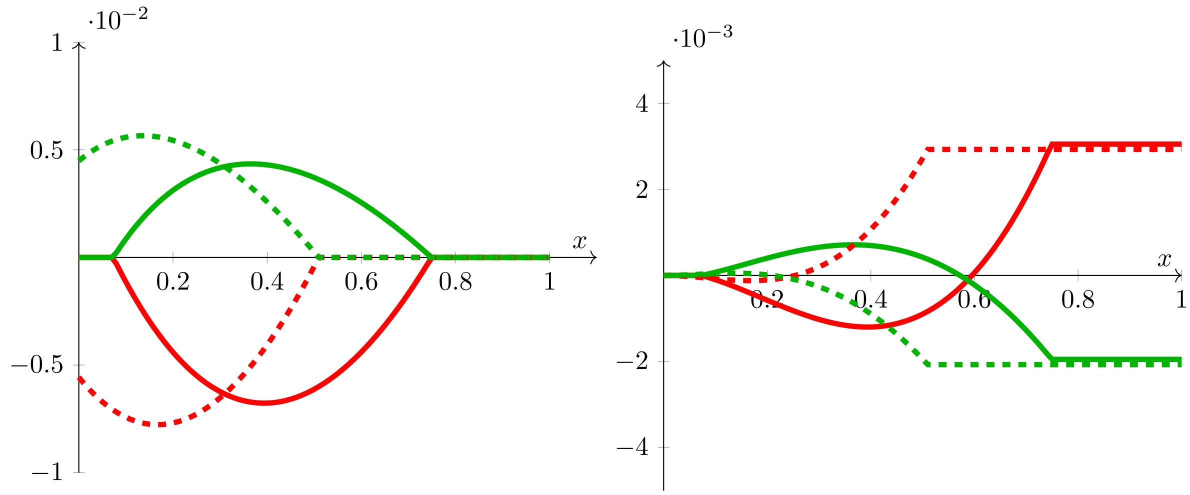
(a) Job quality relative to PAM

(b) Education relative to PAM

Figure 4: **Strong stereotypes: Segregation and diversification.** Changes in job quality (left panel) and education (right panel) relative to the PAM-benchmark for mobile workers (green) and constrained workers (red). Parameter values are $s = 0.37 < \mathcal{H}_\mathcal{A}(\underline{y}^*) = \frac{1}{4}(\sqrt{5}-1)^2 \approx 0.382$ (solid) and $s = 0.4$ (dashed) and depicted is the case where $\zeta_\mathcal{A} > \overline{\zeta}_\mathcal{A}(s)$. Note that for $s = 0.37$ diversification of mobile workers induces PAM at the bottom of the ability distribution whereas for $s = 0.4$ mobile workers fully segregate from constrained workers who choose the dominated occupation $\mathcal{B}$ in accordance with stereotypes in both cases.

Because workers with the highest ability obtain similar jobs as under PAM, i.e. $\lim_{x\to 1}[y(x;C) - y^*(x)] = \lim_{x\to 1}[y(x;M) - y^*(x)] = 0$ and the associated utility gains or losses are monotonically increasing with workers' abilities (c.f. Corrollary 2), there exists an ability-threshold above which constrained workers spend more on education, whereas mobile workers spend less (see Figure 4).

Assimilation. When stereotypes are weak, i.e. $\zeta_{\mathcal{A}} < \overline{\zeta}_{\mathcal{A}}(s)$, constrained workers with high abilities $x \geq \overline{x}$ assimilate by entering both occupations $\mathcal{A}$ and $\mathcal{B}$. These workers are willing to incur the disutility $\zeta_{\mathcal{A}}$ associated with stereotypes in exchange for the better opportunities offered by occupation $\mathcal{A}$. Assimilation induces PAM at the top of the ability distribution, i.e. all workers with abilities $x \geq \overline{x}$, obtain the same jobs as under PAM, no matter whether they are constrained or mobile.



(a) Job quality relative to PAM

(b) Education relative to PAM

Figure 5: **Weak stereotypes: Assimilation.** Changes in job quality (left panel) and education (right panel) relative to the PAM-benchmark for mobile workers (green) and constrained workers (red). Parameter values are $\zeta_{\mathcal{A}} = 0.005$ and $s = 0.39$ (solid) or $s = 0.42$ (dashed), respectively. Note that, in contrast to the case of strong stereotypes depicted in Figure 4, assimilation of constrained workers induces PAM at the top of the ability-distribution. Constrained workers with high abilities are made indifferent between occupations because the disutility $\zeta_{\mathcal{A}}$ from experiencing stereotypes in sector $\mathcal{A}$ is exactly compensated by higher educational expenses in sector $\mathcal{B}$.

Similar to the case before there may or may not exist diversification of mobile workers with low abilities, leading to PAM at the bottom, i.e. for all workers with abilities $x < \underline{x}$. Importantly, $\underline{x} \in [0, \overline{x})$, i.e. even when diversification and assimilation work together, there always remains a non-empty interval of ability-types $x \in (\underline{x}, \overline{x})$ who obtain different jobs than under PAM, with $y(x;C) < y^*(x)$ for the constrained workers and $y(x;M) > y^*(x)$ for the mobile. The thresholds $\underline{x}$ and $\overline{x}$ are determined by the requirement that $y(\underline{x};C) = y(\underline{x};M)$, $y(\overline{x};C) = y(\overline{x};M)$, and

$\zeta_{\mathcal{A}} = \int_{\underline{x}}^{\overline{x}} w(y(z,M)) - w(y(z,C))dz$, the latter guaranteeing that constrained workers with $x \geq \overline{x}$ are indifferent between $\mathcal{A}$ and $\mathcal{B}$ and thus willing to enter both occupations. While $\overline{x}$ is increasing in $\zeta_{\mathcal{A}}$, $\underline{x}$ is decreasing, and both thresholds are decreasing in s. Figure 5 depicts how workers' job quality and education under assimilation compare to the PAM-benchmark. The exact allocations expressions and the conditions, from which they were derived, are relegated to the appendix

5 Conclusion

This article uses a model of "contest-choice" to investigate the implications of workers' competition for jobs in a labor market consisting of two occupational sectors. A distinguishing feature of our approach is that it poses minimal restrictions on the distribution of workers' abilities and the qualities of jobs in the two sectors, allowing for an in-depth investigation of the resulting allocation of workers across occupations and jobs, an issue that is central for our understanding of labor markets.

We have first argued that when workers make occupational choices solely based on the "opportunities" offered by the two sectors, then workers' competition for jobs is powerful enough to induce an efficient allocation of workers, resulting in positive assortative matching between workers and jobs. Our theory offers a definition of "opportunity" that makes positive assortative matching within a segmented labor market associated with sorting of workers across the corresponding occupations. More specifically, we show that when sectors can be ranked—in the sense of likelihood-ratio dominance—with respect to the quality-distribution of the jobs offered, then positive assortative matching emerges because workers sort across occupations, in that a worker's likelihood of choosing the dominating sector is increasing in the worker's ability.

In the second part of our analysis we have identified occupation-specific entry costs as a potential obstacle for a labor-market's allocative efficiency. Motivated by the idea that the high degrees of occupational segregation found even in the most progressive labor markets could arise from the existence of stereotypes, we have extended our model by allowing for workers to differ in a characteristic that is orthogonal to their abilities but associated with stereotypes. We have focused on the case where some workers are "constrained" in that they incur a disutility from choosing the dominant occupation, while all remaining workers are "mobile". Our theory allows for a complete characterization of the distortions arising from the existence of stereotypes, in dependence of the size of the group experiencing them and the size of the associated disutility.

When the group of constrained workers is sufficiently small, then independently of the disutility constrained workers would incur, the mobile workers competition for jobs alone suffices to induce PAM. Mobile workers then constitute the lube in the economy, and if there is enough lube, efficiency is achieved by mobile workers mixing between occupations, and constrained workers avoiding the disutility from stereotypes by exclusive choice of the alternative occupation. In contrast, when stereotypes are sufficiently strong and experienced by a sufficiently large group of workers, mobile worker's competition for jobs is not enough to guarantee the efficient allocation of jobs across all workers.

A Appendix: Proofs

A.1 Proof of Proposition 1

In what follows below, we proceed in several steps with the proof of Proposition 1.

Consider an effort profile

$$\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(x) = x \cdot w(y_{\mathcal{A}\cup\mathcal{B}}(x)) - \int_0^x w(y_{\mathcal{A}\cup\mathcal{B}}(z))dz. \tag{43}$$

This effort profile follows from the Envelope Condition.[6] We have the following result about this effort profile.

Claim 1. $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}$ *is an equilibrium of the effort choice game of the Grand Contest.*

Proof. Define $\underline{x}$ as $\underline{x} : \mathcal{F}(\underline{x}) = \mathcal{H}_{\mathcal{A}\cup\mathcal{B}}(\underline{y})$, i.e. the highest among the types, that get $\underline{y}$ in the assortative allocation. Similarly, let $\overline{x}$ be such that $\mathcal{F}(\overline{x}) = \lim_{\tilde{y}\uparrow 1} \mathcal{H}_{\mathcal{A}\cup\mathcal{B}}(\tilde{y})$, so that $\overline{x}$ is the lowest among the types, that get 1 in the assortative allocation. Rewrite $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}$ as

$$\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(x) = \int_0^x \left(w(y^*(x)) - w(y_{\mathcal{A}\cup\mathcal{B}}(z))\right) dz. \tag{44}$$

Notice that because of our assumptions on $\mathcal{H}_{\mathcal{A}}$ and $\mathcal{H}_{\mathcal{B}}$ (continuous and strictly increasing in $(0,1)$), $\mathcal{H}_{\mathcal{A}\cup\mathcal{B}}$ is continuous and strictly increasing for $y \in \left(\underline{y}, 1\right)$. This means that $y^*(x) = \mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\mathcal{F}(x))$ is globally continuous and, moreover, it is strictly increasing in x for $x \in (\underline{x}, \overline{x})$ and stays constant for $x \in [0, \underline{x})$ and $x \in (\overline{x}, 1]$. Thus, $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(x)$ is also globally continuous, strictly increasing in x for $x \in (\underline{x}, \overline{x})$ and is constant for $x \in [0, \underline{x})$ and $x \in (\overline{x}, 1]$. Notice also that if $\underline{x} > 0$ and $\overline{x} < 1$, the range of efforts that occur with probability one in candidate equilibrium is $[\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\underline{x}), \hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\overline{x})] = [0, \overline{x} \cdot w(1) - \int_0^{\overline{x}} w(y^*(z))dz]$. If it is the case that $\underline{x} = 0$ ($\overline{x} = 1$), the point $e = 0$ ($e = \overline{x} \cdot w(1) - \int_0^{\overline{x}} w(y^*(z))dz$) would be excluded from the support of all possible efforts.

The payoff from following the candidate equilibrium is $V(x) = \int_0^x w(y^*(z))dz$, which is strictly positive for $x > \underline{x}$. Let us check that for all types $x \in (0,1)$, there does not exist profitable deviations from $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(x)$ when everyone else is following $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(x)$. Start with types $x \in (\underline{x}, \overline{x})$. Consider the deviation level of effort $\tilde{e} \in \mathbb{R}_+$. Deviations to effort levels outside of the candidate range, i.e. $\tilde{e} \notin [\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\underline{x}), \hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\overline{x})]$ are strictly dominated. $\tilde{e} \leqslant \hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\underline{x})$ leads to zero prize and weakly negative payoff, whereas the candidate payoff is strictly positive. $\tilde{e} > \hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\overline{x})$ leads to prize 1, but so does the effort level $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\overline{x})$, but for a strictly smaller cost. For a deviation $\tilde{e} \in [\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\underline{x}), \hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\overline{x})]$, there exists a type $\tilde{x} \in [\underline{x}, \overline{x}]$ such that $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\tilde{x}) = \tilde{e}$. That is, computing the payoff from following the candidate effort function but mimicking $\tilde{x}$ is the same as computing the payoff from the deviation $\tilde{e}$. Let us write down the net gain from such deviation (below, $V^D(x, \tilde{x})$ denotes the payoff from

[6] See, for example, Milgrom and Segal (2002).

following the effort profile $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}$ but "mimicking" type $\tilde{x}$ and everyone else playing $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}$ truthfully):

$$V^D(x,\tilde{x}) - V(x) = x \cdot w(y^*(\tilde{x})) - \tilde{x} \cdot w(y^*(\tilde{x})) + \int_0^{\tilde{x}} w(y^*(z))dz - \int_0^x w(y^*(z))dz = \tag{45}$$

$$= \int_x^{\tilde{x}} \left(w(y^*(z)) - w(y^*(\tilde{x}))\right) dz, \tag{46}$$

which is negative, regardless of whether $\tilde{x} > x$ or $\tilde{x} < x$, since y^* is increasing. Thus, there are no profitable deviations from $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}$ for types $x \in (\underline{x}, \overline{x})$.

Consider now a type $x \in (0, \underline{x}]$.[7] If x follows $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}$, they receive the prize 0. In order to receive a strictly positive prize, they would need to deviate to effort $\tilde{e} > \hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\underline{x})$ which amounts to emulating some type $\tilde{x} > \underline{x}$ such that $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\tilde{x}) = \tilde{e}$. We have that

$$V^D(x,\tilde{x}) - V(x) = x \cdot w(y^*(\tilde{x})) - \hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\tilde{x}) - 0 \leqslant \underline{x} \cdot w(y^*(\tilde{x})) - \hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\tilde{x}) = \tag{47}$$

$$= \underline{x}w(y^*(\tilde{x})) - \tilde{x}w(y^*(\tilde{x})) + \int_{\underline{x}}^{\tilde{x}} w(y^*(z))dz = \tag{48}$$

$$= \int_{\underline{x}}^{\tilde{x}} \left(w(y^*(z)) - w(y^*(\tilde{x}))\right) < 0, \tag{49}$$

which implies there are no profitable deviations for $x \in (0, \underline{x}]$.

Finally, consider a type $x \in [\underline{x}, 1]$. Any such type receives the highest-valued prize of 1 if they follow $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}$. Their payoff is then

$$V(x) = (x - \overline{x}) \cdot w(1) + \int_{\underline{x}}^{\overline{x}} w(y^*(z))dz. \tag{50}$$

Exerting efforts higher than $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\overline{x})$ leads to the same prize but at a higher cost. Thus, only deviations to $\tilde{e} < \hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\overline{x})$ need to be considered. They are equivalent to emulating some $\tilde{x} < \overline{x}$ such that $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\tilde{x}) = \tilde{e}$. As before, we have

$$V^D(x,\tilde{x}) - V(x) = x \cdot w(y^*(\tilde{x})) - \hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(\tilde{x}) - (x - \overline{x}) \cdot w(1) - \int_{\underline{x}}^{\overline{x}} w(y^*(z))dz = \tag{51}$$

$$= x \cdot w(y^*(\tilde{x})) - \tilde{x} \cdot w(y^*(\tilde{x})) - \int_{\tilde{x}}^{x} w(y^*(z))dz \tag{52}$$

$$= \int_{\tilde{x}}^{x} (w(y^*(\tilde{x})) - w(y^*(z)))dz < 0. \tag{53}$$

This completes the argument that there are no profitable deviations from $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}$ for any type $x \in (0, 1)$. □

[7] If it is the case that $\underline{x} = 0$ so that $(0, \underline{x}]$ is empty, it is vacuously true that $x \in (0, \underline{x}]$ have no profitable deviations.

Our next step will be to argue that almost all types $x \in (0,1)$ follow the effort profile $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}$ in all equilibria. To do that, we first need to generalize the notion of equilibria to account for the possibility of Mixed Strategy Equilibria. Let the Mixed Strategy $\sigma\varepsilon$ be a mapping from the types to random variables with non-negative supports: $\sigma\varepsilon : (0,1) \to \Delta([0,+\infty))$. Working with $\sigma\varepsilon$ is equivalent to working with a famility of CDFs of efforts for each type, $\{G_x(e)\}_{x\in(0,1)}$ such that

$$G_x(e) = \mathbb{P}\{\sigma\varepsilon(x) \leqslant e\}. \tag{54}$$

Given a family of such CDFs, we can define a global distribution of effort through its CDF as

$$\Gamma_{\sigma\varepsilon}(e) \doteq \mathbb{P}\{\sigma\varepsilon(X) \leqslant e\} = \int_0^1 G_x(e)d\mathcal{F}(x). \tag{55}$$

This way, $\Gamma_{\sigma\varepsilon}(e)$ is the probability that the effort level of a randomly drawn contestant will not be greater than some effort level e, given that everybody follows the profile $\sigma\varepsilon$.

Then, Mixed Equilibrium of the effort choice game in the Grand Contest is a profile $\sigma\varepsilon$ (equivalently, a family of CDFs $\{G_x(e)\}_{x\in(0,1)}$) such that for every type $x \in (0,1)$, the support of the random effort, given by $G_x(e)$, is contained within the set

$$\arg\max_e \{x \cdot w(\mathcal{H}^{-1}(\Gamma_{\sigma\varepsilon}(e))) - e\}. \tag{56}$$

The fact that the set of maximizers of the above objective is non-empty for every x follows from the fact that there exists an equilibrium that we have established above. In that equilibrium, $G_x(e) = \mathbb{I}\{\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(x) \leqslant e\}$. We are now ready to state and prove the result regarding the essential uniqueness of the equilibrium we have previously considered:

Claim 2. *In all Mixed Equilibria, $G_x(e) = \mathbb{I}\{\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(x) \leqslant e\}$ for almost all $x \in (0,1)$.*

So, effectively, almost everyone's equilibrium strategy is pure.

Proof. Consider an arbitrary Mixed Equilibrium given by some $\{G_x(e)\}_{x\in(0,1)}$ and the correspondent $\Gamma_{\sigma\varepsilon}(e)$.[8] Let us look at the objective function of a contestant with type x, given that everybody else is following $\{G_x\}$:

$$u(x,e) = x \cdot w\left(\mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(e))\right) - e. \tag{57}$$

We are interested in the set of maximizers of the above expression with respect to e and how such

[8]From here until the end of the proof, suppress the subscript $\sigma\varepsilon$.

sets evolve with x. Formally, we are interested in $E(x)$ such that

$$V(x) = \sup_{e \in \mathbb{R}_+} u(x,e) \tag{58}$$

$$E(x) = \{e : u(x,e) = V(x)\}. \tag{59}$$

That $E(x)$ is not empty follows from the facts that we study the Best Response to equilibrium, and an equilibrium exists. Let us check that the conditions of Theorem 4 of Milgrom and Shannon (1994) are satisfied and we can apply the Monotone Comparative Statics result. Indeed, $u(x,e)$ is trivially quasi-supermodular in e, as a function of e, $u(x,.)$ is univariate. Moreover, $u(x,e)$ has the single crossing property in (x,e). To check that, consider $x' > x''$ and $e' > e''$. If it is the case that $u(x'',e') > u(x'',e'')$, which can be rewritten as

$$x''w(\mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(e'))) - e' > x''w(\mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(e''))) - e'' \tag{60}$$

$$\iff \tag{61}$$

$$x''(w(\mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(e'))) - w(\mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(e'')))) > e' - e'', \tag{62}$$

then, we have that $u(x',e') > u(x',e'')$, as

$$x'(w(\mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(e'))) - w(\mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(e'')))) > x''(w(\mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(e'))) - w(\mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(e'')))) > e' - e'' \Rightarrow \tag{63}$$

$$x'w(\mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(e'))) - e' > x'w(\mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(e''))) - e''. \tag{64}$$

Furthermore, $u(x'',e') \geqslant u(x'',e'')$ implies $u(x',e') \geqslant u(x',e'')$, as under $e' > e''$ it cannot be that $u(x'',e') = u(x'',e'')$. Thus, using the results of Milgrom and Shannon (1994), we have that $E(x)$ is monotone non-decreasing in x in the set sense. This means that for $x' \geqslant x''$, the largest element of the union $E(x') \cup E(x'')$ is contained in $E(x')$, and the smallest element of that union is contained in $E(x'')$. Thus, the set of maximizers is increasing with respect to type.

Now let us argue that $E(x)$ is a singleton for almost all $x \in (0,1)$. Check that conditions of Theorem 2 of Milgrom and Segal (2002) are satisfied. Indeed, $f(.,e)$ is a linear function (as a function of x) for all effort levels e, which makes it absolutely continuous. Moreover, $\frac{\partial u(x,e)}{\partial x} = w(\mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(e)))$, hence, $|u(x,e)'_x| \leqslant w(1)$ for all x and e, with $w(1)$ being a constant, thus, an integrable function. Furthermore, $u(.,e)$ is differentiable (in type x) for all e and $E(x)$ is non-empty for all x. Altogether, we can apply the results of Milgrom and Segal (2002) and state that given equilibrium $\{G_x\}$, for any selection $\varepsilon^*(x) \in E(x)$,

$$V(x) = V(0) + \int_0^x \frac{\partial u(z,\varepsilon^*(z))}{\partial x} dz = \int_0^x w(\mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(\varepsilon^*(z))))dz. \tag{65}$$

Suppose now that for a subset $\tilde{X} \subset (0,1)$ of talent types, it holds that $E(x)$ consists of at least two distinct elements, $\varepsilon^1(x)$ and $\varepsilon^2(x)$, $\varepsilon^1(x) > \varepsilon^2(x)$. Then, there is an interval $(x^I, x^{II}) \subset \tilde{X}$ such that for every type $\hat{x}$ inside (x^I, x^{II}), it also holds that $\varepsilon^1(\hat{x}) > \varepsilon^2(\hat{x})$. From the envelope condition above, it should hold that $\mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(\varepsilon^1(\hat{x}))) = \mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(\varepsilon^2(\hat{x})))$, $x \in (x^I, x^I I)$. Let us write down the payoffs of type $\hat{x}$ from $\varepsilon^1(\hat{x})$ and $\varepsilon^2(\hat{x})$ and remember that they should coincide:

$$u(\hat{x}, \varepsilon^1(\hat{x})) = \hat{x} \cdot w(\mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(\varepsilon^1(\hat{x})))) - \varepsilon^1(\hat{x}), \tag{66}$$

$$u(\hat{x}, \varepsilon^2(\hat{x})) = \hat{x} \cdot w(\mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\Gamma(\varepsilon^2(\hat{x})))) - \varepsilon^2(\hat{x}), \tag{67}$$

from which we get the contradiction, since we agreed that $\varepsilon^1(\hat{x}) > \varepsilon^2(\hat{x})$. Thus, $E(x)$ is a singleton for almost every x, so we can write $E(x) = \{\varepsilon(x)\}$ where $\varepsilon(x)$ is a weakly increasing real-valued function of x.

So far we have established that in any equilibrium, $\{G_x(e)\} = \{\mathbb{I}\{\varepsilon(x) \leqslant e\}\}$ for almost any x. Since $\varepsilon(x)$ is weakly increasing, any equilibrium should be assortative. So, in any equilibrium, the prize that type x gets is $y^*(x) = \mathcal{H}^{-1}_{\mathcal{A}\cup\mathcal{B}}(\mathcal{F}(x))$ a.e. in x. Then, the envelope condition would pin down $\varepsilon(x)$ to be $\hat{\varepsilon}_{\mathcal{A}\cup\mathcal{B}}(x)$ for any equilibrium ε a.e. in x, which completes the proof. □

A.2 Proof of Essential Uniqueness in Proposition 2

Before we begin with the actual proof of essential uniqueness in proposition 2, we state and prove some subsidiary claims that will be useful.

Claim 3. *All within-contest equilibria allocations are assortative.*

Proof. Consider contest i and let us take a look at the Best Response set of an arbitrary contestant with type x given their perceived distribution of efforts in i $\hat{\Psi}_{i,x}(\tilde{e})$:

$$E(x) = \left\{ \arg\max_e \left\{ x \cdot w \left(\mathcal{H}_i^{-1} \left(1 - \int \mathbb{I}\{e \leqslant \tilde{e}\} d\hat{\Psi}_{i,x}(\tilde{e}) \right) \right) - e \right\} \right\}.$$

We would like to argue that E is non-decreasing in x and is a singleton. We can apply the steps we took in the proof of Lemma 2 if we replace $u(x,e)$ with

$$u(x,e) = x \cdot w \left(\mathcal{H}_i^{-1} \left(1 - \int \mathbb{I}\{e \leqslant \tilde{e}\} d\hat{\Psi}_{i,x}(\tilde{e}) \right) \right) - e.$$

Indeed, we have that $u(x,.)$ is quasi-supermodular and $u(x,e)$ satisfies single crossing. Thus, $E(x)$ is non-decreasing in the set sense. Applying again the results of the Envelope Theorem, we also conclude that $E(x)$ is a singleton. Best Response to any perceived distribution of efforts is a non-decreasing singleton. Thus, in equilibrium, in both contests, higher types exert higher efforts and win higher prizes *a.e.* in x. □

Due to Claim 3, in every equilibrium,

- there are two, potentially different, contest-specific single-valued non-decreasing effort functions, $\varepsilon_i(x)$, $i \in \{\mathcal{A}, \mathcal{B}\}$;
- the contest-specific consistent perceived distributions of efforts are given by

$$\hat{\Psi}_{\mathcal{A},x}(\hat{e}) = \int_0^1 \mathbb{I}\{\varepsilon_{\mathcal{A}}(z) \leqslant \hat{e}\} p_{\mathcal{A}}(z) dz \text{ and } \hat{\Psi}_{\mathcal{B},x}(\hat{e}) = \int_0^1 \mathbb{I}\{\varepsilon_{\mathcal{A}}(z) \leqslant \hat{e}\}(1 - p_{\mathcal{A}}(z)) dz; \quad (68)$$

- almost every type x expects to win, with probability 1, the contest-specific prizes $y_{\mathcal{A}}(x)$ and $y_{\mathcal{B}}(x)$ defined as

$$y_{\mathcal{A}}(x) = \mathcal{H}_{\mathcal{A}}^{-1}\left(1 - \int_x^1 p_{\mathcal{A}}(z) d\mathcal{F}(z)\right) \text{ and } y_{\mathcal{B}}(x) = \mathcal{H}_{\mathcal{B}}^{-1}\left(1 - \int_x^1 (1 - p_{\mathcal{A}}(z)) d\mathcal{F}(z)\right); \quad (69)$$

- contest-specific payoff functions $V_{\mathcal{A}}(x)$ and $V_{\mathcal{B}}(x)$ are given by

$$V_{\mathcal{A}}(x) = \int_0^x w(y_{\mathcal{A}}(z)) dz \text{ and } V_{\mathcal{B}}(x) = \int_0^x w(y_{\mathcal{B}}(z)) dz. \quad (70)$$

  $V_{\mathcal{A}}(x)$ and $V_{\mathcal{B}}(x)$ are continuous and increasing.

Moreover, we can make the following statement about the payoff functions:

Claim 4. *In all equilibria, $V_{\mathcal{A}}$ and $V_{\mathcal{B}}$ are convex functions.*

Proof. In the previous claim we have established that $x^1 \leqslant x^2$ implies $y_i(x^1) \leqslant y_i(x^2)$, $i \in \{\mathcal{A}, \mathcal{B}\}$. Let us show that $V_i(x)$ satisfies *midpoint convexity*, i.e. $\forall x_1, x_2 \in (0,1)$, $x_1 \leqslant x_2$ it holds that

$$V_i\left(\frac{x_1 + x_2}{2}\right) \leqslant \frac{1}{2} V_i(x_1) + \frac{1}{2} V_i(x_2).$$

We have that

$$V_i(x_2) - V_i\left(\frac{x_1 + x_2}{2}\right) = \int_{\frac{x_1+x_2}{2}}^{x_2} w(y_i(z)) dz \quad (71)$$

and

$$V_i\left(\frac{x_1 + x_2}{2}\right) - V_i(x_1) = \int_{x_1}^{\frac{x_1+x_2}{2}} w(y_i(z)) dz. \quad (72)$$

Since $x_2 \geqslant \frac{x_1+x_2}{2} \geqslant x_1$, $y_i()$ and $w()$ are both non-decreasing, it holds that

$$\int_{\frac{x_1+x_2}{2}}^{x_2} w(y_i(z))dz \geqslant \int_{x_1}^{\frac{x_1+x_2}{2}} w(y_i(z))dz. \tag{73}$$

Thus, $V_i(x_2) - V_i(\frac{x_1+x_2}{2}) \geqslant V_i(\frac{x_1+x_2}{2}) - V_i(x_1)$, or, $\frac{1}{2}V_i(x_1) + \frac{1}{2}V_i(x_2) \geqslant V_i\left(\frac{x_1+x_2}{2}\right)$, as desired.

V_i is continuous and, therefore, measurable. A measurable midpoint convex function is convex, a result due to Blumberg (1919) and Sierpiński (1920). □

Below is the proof of essential uniqueness.

Proof. Suppose there is a measurable subset $\tilde{X} \subset (0,1)$ with non-zero measure such that

$$y_{\mathcal{A}}(x) \neq y_{\mathcal{B}}(x), \text{ and } y_{\mathcal{A}}(x) \neq y^*(x)$$

for all $x \in \tilde{X}$. Consider an open interval $(\underaccent{\tilde}{x}, \tilde{x}) \subset \tilde{X}$. Note that if $V_{\mathcal{A}}(x) = V_{\mathcal{B}}(x)$ for all x in some open interval $\underline{x}, \overline{x}$, it follows that $\frac{dV_{\mathcal{A}}(x)}{dx} = \frac{dV_{\mathcal{B}}(x)}{dx}$, $\forall x \in (\underline{x}, \overline{x})$. Since $\frac{dV_i(x)}{dx} = y_i(x)$, $y_{\mathcal{A}}(x) \neq y_{\mathcal{B}}(x)$, $\forall x \in (\underaccent{\tilde}{x}, \tilde{x})$ implies $V_{\mathcal{A}}(x) \neq V_{\mathcal{B}}(x)$, $\forall x \in (\underaccent{\tilde}{x}, \tilde{x})$. By continuity of both V_i, we can find a non-empty interval $(\underaccent{\hat}{x}, \check{x}) \subseteq (\underaccent{\tilde}{x}, \tilde{x})$ such that either $V_{\mathcal{A}}(x) > V_{\mathcal{B}}(x)$, $\forall x \in (\underaccent{\hat}{x}, \check{x})$, or $V_{\mathcal{B}}(x) > V_{\mathcal{A}}(x)$, $\forall x \in (\underaccent{\hat}{x}, \check{x})$; and one of the following holds:

(*i*) $V_{\mathcal{A}}(\underaccent{\hat}{x}) = V_{\mathcal{B}}(\underaccent{\hat}{x})$, $V_{\mathcal{A}}(\check{x}) = V_{\mathcal{B}}(\check{x})$, $0 < \underaccent{\hat}{x}$, $\check{x} < 1$;

(*ii*) $V_{\mathcal{A}}(\check{x}) = V_{\mathcal{B}}(\check{x})$, $\underaccent{\hat}{x} = 0$, $\check{x} < 1$;

(*iii*) $V_{\mathcal{A}}(\underaccent{\hat}{x}) = V_{\mathcal{B}}(\underaccent{\hat}{x})$, $0 < \underaccent{\hat}{x}$, $\check{x} = 1$;

(*iv*) $\underaccent{\hat}{x} = 0$, $\check{x} = 1$.

Assume, without loss, that $V_{\mathcal{A}}(x) > V_{\mathcal{B}}(x)$, $\forall x \in (\underaccent{\hat}{x}, \check{x})$, and consider the above four cases in turn:

Case (*a*) $\mathbf{V}_{\mathcal{A}}(\underaccent{\hat}{\mathbf{x}}) = \mathbf{V}_{\mathcal{B}}(\underaccent{\hat}{\mathbf{x}})$, $\mathbf{V}_{\mathcal{A}}(\check{\mathbf{x}}) = \mathbf{V}_{\mathcal{B}}(\check{\mathbf{x}})$, $\mathbf{0} < \underaccent{\hat}{\mathbf{x}}$, $\check{\mathbf{x}} < \mathbf{1}$: Since $V_{\mathcal{A}}(x) > V_{\mathcal{B}}(x)$, $\forall x \in (\underaccent{\hat}{x}, \check{x})$, we have in that interval $V(x) = p_{\mathcal{A}}(x)V_{\mathcal{A}}(x) + (1 - p_{\mathcal{A}}(x))V_{\mathcal{B}}(x) = V_{\mathcal{A}}(x) \Rightarrow p_{\mathcal{A}}(x) = 0$. Hence, there is no type in $(\underaccent{\hat}{x}, \check{x})$ that joins contest $\mathcal{B}$. As a result of no one joining $\mathcal{B}$, and because equilibrium allocations are assortative, $y_{\mathcal{B}}(x) = y_{\mathcal{B}}(\underaccent{\hat}{x})$, $\forall x \in (\underaccent{\hat}{x}, \check{x}]$. Thus, the derivative of $V_{\mathcal{B}}$ is constant in $[\underaccent{\hat}{x}, \check{x}]$. Since $V_{\mathcal{B}}$ crosses $V_{\mathcal{A}}$ from below at $\check{x}$, and $V_{\mathcal{A}}$ is convex, so its slope is non-decreasing, it must be that $w(y_{\mathcal{B}}(x)) > w(y_{\mathcal{A}}(x))$, $x \in (\underaccent{\hat}{x}, \check{x})$. Thus, we have $V_{\mathcal{B}}(x) = V_{\mathcal{B}}(\underaccent{\hat}{x}) + \int_{\underaccent{\hat}{x}}^{x} w(y_{\mathcal{B}}(z))dz > V_{\mathcal{A}}(\underaccent{\hat}{x}) + \int_{\underaccent{\hat}{x}}^{x} w(y_{\mathcal{A}}(z))dz = V_{\mathcal{A}}(x)$, for $x \in (\underaccent{\hat}{x}, \check{x})$, a contradiction.

Case (*b*) $\mathbf{V}_{\mathcal{A}}(\check{\mathbf{x}}) = \mathbf{V}_{\mathcal{B}}(\check{\mathbf{x}})$, $\underaccent{\hat}{\mathbf{x}} = \mathbf{0}$, $\check{\mathbf{x}} < \mathbf{1}$: Note that $V_{\mathcal{A}}(0) = V_{\mathcal{B}}(0) = 0$. Thus, we can arrive at the contradiction following the same steps as in the previous case.

Case (*c*) $\mathbf{V}_{\mathcal{A}}(\underset{\wedge}{\mathbf{x}}) = \mathbf{V}_{\mathcal{B}}(\underset{\wedge}{\mathbf{x}})$, $0 < \underset{\wedge}{\mathbf{x}}$, $\check{\mathbf{x}} = \mathbf{1}$: This case implies that no type above $\underset{\wedge}{x}$ joins contest $\mathcal{B}$. For $x \in (\underset{\wedge}{x} - \epsilon, \underset{\wedge}{x})$, $\epsilon > 0$, either there is ϵ such that $V_{\mathcal{A}}(x) = V_{\mathcal{B}}(x)$, $\forall x \in (\underset{\wedge}{x} - \epsilon, \underset{\wedge}{x})$, or there is no such ϵ. However, having considered cases (i) and (ii), we can exclude the latter possibility. Thus, $V_{\mathcal{A}}(x)$ and $V_{\mathcal{B}}(x)$ coincide for some open interval below $\underset{\wedge}{x}$, and so do their slopes, and hence, $y_{\mathcal{A}}(x) = y_{\mathcal{B}}(x)$, $x \in (\underset{\wedge}{x} - \epsilon, \underset{\wedge}{x})$. So we have that prizes that types above $\underset{\wedge}{x}$ would get in $\mathcal{B}$ are strictly lower than the prizes they get in $\mathcal{A}$, even though the prizes coincide for types $(\underset{\wedge}{x} - \epsilon, \underset{\wedge}{x})$. Recall that both distributions of prizes have full support and the highest type joining $\mathcal{B}$ is awarded the highest prize there, valued at 1. This contradicts $\underset{\wedge}{x} < \check{x}$.

Case (*d*) $\underset{\wedge}{\mathbf{x}} = \mathbf{0}, \check{\mathbf{x}} = \mathbf{1}$: This case implies that no type joins contest $\mathcal{B}$. Then, any agent $x < \overline{x}$, who would choose to join $\mathcal{B}$ would be awarded the highest prize 1 exerting any infinitely small effort, leading to payoff close to $xw(1)$ which would be strictly greater than $V_{\mathcal{A}}(x) = xw(y_{\mathcal{A}}(x)) - \varepsilon_{\mathcal{A}}(x)$. This observation contradicts the fact that every type weakly prefers joining contest $\mathcal{A}$. □

A.3 Proof of Proposition 3, part 1, case $x^{FS} = 0$

Let us check that the profile stated in Part 1 of Proposition 3 for the case $s > \mathcal{H}_{\mathcal{A}}(\underline{y})$ is an equilibrium. To achieve that, check that there are no profitable deviations by any talent type x of any participants subgroup, constrained or mobile. Note that, fixing the behavior of others, if a type-x contestant joins $\mathcal{A}$, the mass of participants with a higher talent in $\mathcal{A}$ would be

$$\int_x^1 (1-s)f(z)dz = (1-s)(1-x). \tag{74}$$

Hence, in an assortative allocation, type x contestant would get prize $\tilde{y}$ such that $1 - \mathcal{H}_{\mathcal{A}}(\tilde{y}) = s(1-x)$, since $1 - \mathcal{H}_{\mathcal{A}}(\tilde{y})$ is a mass of prizes in $\mathcal{A}$ more valuable than $\tilde{y}$. Since we know that an assortative allocation can be supported in equilibrium by the effort profile that follows from the envelope condition, we get that $\varepsilon_{\mathcal{A}}(x)$, as expressed above, is indeed an equilibrium effort profile that supports the assortative allocation. Applying the same steps to contest $\mathcal{B}$, there, the mass of participants with a talent higher than x would be $\int_x^1 sf(z)dz = s(1-x)$. The assortative allocation would yield to x the prize $\hat{y}$ such that $1 - \mathcal{H}_{\mathcal{B}}(\hat{y}) = (1-s)(1-x)$, which is supported by the effort profile $\varepsilon_{\mathcal{B}}$.

We need to check that the mobile participants do not want to switch to $\mathcal{B}$ and the constrained participants do not want to switch to $\mathcal{A}$.

Recall our assumption that $\mathcal{H}_{\mathcal{A}} \succsim_{\text{LR}} \mathcal{H}_{\mathcal{B}}$. Our first goal is to show that the mobile contestants don't want to switch from $\mathcal{A}$ to $\mathcal{B}$ because $y_{\mathcal{A}}(x) \geqslant y_{\mathcal{B}}(x)$, $\forall x \in (0,1)$. As we proceed, we need to consider three cases: (i) $\mathcal{H}_{\mathcal{A}}^{-1}(s) > 0$ and $\mathcal{H}_{\mathcal{B}}^{-1}(1-s) > 0$, there is no unemployment under full segregation; (ii) $\mathcal{H}_{\mathcal{A}}^{-1}(s) > 0$ and $\mathcal{H}_{\mathcal{B}}^{-1}(1-s) = 0$, there is unemployment in occupation $\mathcal{B}$ under full segregation;

Start with case (i). Consider first the positions that type $x = 0$ would get in the two occupations. In occupation $\mathcal{B}$, it is going to be the position $\underline{y}_{\mathcal{B}}$ such that $\mathcal{H}_{\mathcal{B}}(\underline{y}_{\mathcal{B}}) = 1-s$, and in occupation $\mathcal{A}$ - $\underline{y}_{\mathcal{A}}$ such that $\mathcal{H}_{\mathcal{A}}(\underline{y}_{\mathcal{A}}) = s$. If the PAM allocation was achievable, type $x = 0$ there would have been getting the position $\underline{y}$ such that $\mathcal{H}_{\mathcal{A}}(\underline{y}) + \mathcal{H}_{\mathcal{B}}(\underline{y}) = 1$. Let us show that $\underline{y}_{\mathcal{A}} > \underline{y} > \underline{y}_{\mathcal{B}}$. Since we have assumed that $s > \mathcal{H}_{\mathcal{A}}(\underline{y})$, we have $\mathcal{H}_{\mathcal{A}}(\underline{y}_{\mathcal{A}}) > \mathcal{H}_{\mathcal{A}}(\underline{y})$ and, hence, $\underline{y}_{\mathcal{A}} > \underline{y}$. Also, since $\mathcal{H}_{\mathcal{A}}(\underline{y}_{\mathcal{A}}) + \mathcal{H}_{\mathcal{B}}(\underline{y}_{\mathcal{B}}) = 1$, combined with the facts that $\mathcal{H}_{\mathcal{A}}(\underline{y}) + \mathcal{H}_{\mathcal{B}}(\underline{y}) = 1$ and $\mathcal{H}_{\mathcal{A}}(\underline{y}_{\mathcal{A}}) > \mathcal{H}_{\mathcal{A}}(\underline{y})$, it must hold that $\mathcal{H}_{\mathcal{B}}(\underline{y}_{\mathcal{B}}) < \mathcal{H}_{\mathcal{B}}(\underline{y})$, from which it follows that $\underline{y}_{\mathcal{B}} < \underline{y}$.

In order to obtain the position valued as $\underline{y}_{\mathcal{A}}$ in occupation $\mathcal{B}$, the type of the contestant must be higher than the one, which obtains $\underline{y}_{\mathcal{A}}$ in $\mathcal{A}$. Let us define an auxiliary function $D(y)$ on $y \in [\underline{y}_{\mathcal{A}}, 1]$ as follows:

$$D(y) = x_{\mathcal{B}}(y) - x_{\mathcal{A}}(y) = \frac{\mathcal{H}_{\mathcal{B}}(y) - (1-s)}{s} - \frac{\mathcal{H}_{\mathcal{A}}(y) - s}{1-s}, \tag{75}$$

that is, $D(y)$ is the difference between the types that obtain prize y in contest $\mathcal{B}$ versus in contest $\mathcal{A}$. We know that $D(\underline{y}_{\mathcal{A}}) > 0$ and that $D(1) = 0$, because it is the type 1 that obtains the prize 1 in both contests. We have that

$$D'(y) = \frac{h_{\mathcal{B}}(y)}{s} - \frac{h_{\mathcal{A}}(y)}{1-s} = \frac{h_{\mathcal{A}}(y)}{s}\left(\frac{h_{\mathcal{B}}(y)}{h_{\mathcal{A}}(y)} - \frac{s}{1-s}\right). \tag{76}$$

Since $\mathcal{H}_{\mathcal{A}} \succsim_{\mathrm{LR}} \mathcal{H}_{\mathcal{B}}$, the fraction $h_{\mathcal{B}}(y)/h_{\mathcal{A}}(y)$ is decreasing. Thus, $D'(y)$ can only switch sign once, and if it does, it goes from positive to negative, as y increases; or $D'(y)$ can be negative throughout all $y \in [\underline{y}_{\mathcal{A}}, 1]$. This implies that $D(y)$ can either be first increasing and then decreasing, or it can be monotonically decreasing. Altogether, we have $D(y) \geqslant 0$ for $y \in [\underline{y}_{\mathcal{A}}, 1]$. Thus, getting the same prize in contest $\mathcal{B}$ requires a higher talent than in contest $\mathcal{A}$, which is equivalent to saying that the same talent type would get a higher prize in $\mathcal{A}$ than in $\mathcal{B}$. Hence, the mobile contestants prefer contest $\mathcal{A}$, since

$$V_{\mathcal{A}}(x) = \int_0^x w(y_{\mathcal{A}}(z))dz \geqslant \int_0^x w(y_{\mathcal{B}}(z))dz = V_{\mathcal{B}}(x). \tag{77}$$

As for the constrained workers, they cannot benefit from switching from $\mathcal{B}$ to $\mathcal{A}$ because the entry cost to $\mathcal{A}$ is too high, even for the most talented workers:

$$V_{\mathcal{A}}(x) - V_{\mathcal{B}}(x) = \int_0^x \left(w(y_{\mathcal{A}}(z)) - w(y_{\mathcal{B}}(z))\right) dz \leqslant \int_0^1 \left(w(y_{\mathcal{A}}(z)) - w(y_{\mathcal{B}}(z))\right) dz < \zeta_{\mathcal{A}}. \tag{78}$$

Continue with case (ii), $\mathcal{H}_{\mathcal{A}}^{-1}(s) > 0$ and $\mathcal{H}_{\mathcal{B}}^{-1}(1-s) = 0$. It follows that the lowest-paying position that is allocated in $\mathcal{A}$ is positive, $\underline{y}_{\mathcal{A}} > 0$, while the lowest position in $\mathcal{B}$ is zero, $\underline{y}_{\mathcal{B}} = 0$. Hence, $\underline{y}_{\mathcal{A}} > \underline{y}_{\mathcal{B}}$, as before, and the rest of the proof carries out as in case (i).

Finally, note that there cannot be unemployment in $\mathcal{A}$ under full segregation and the assumption that $s > \mathcal{H}_{\mathcal{A}}(\underline{y})$. If there is no unemployment under PAM, $\underline{y} > 0$, it implies there cannot be unemployment in $\mathcal{A}$. If there is unemployment under PAM, $\underline{y} = 0$, the assumption $s > \mathcal{H}_{\mathcal{A}}(\underline{y})$ implies that the amount of workers in $\mathcal{A}$, $1 - s$, is smaller than the amount of positive-paying positions in $\mathcal{A}$, $1 - \mathcal{H}_{\mathcal{A}}(0)$.

A.4 Proof of Proposition 3, part 1, case $x^{FS}(s) > 0$

Before starting with the proof for that case, we state and proof two auxiliary results, which will be useful with equilibrium characterization. In the claim below, we argue formally that there is a threshold type $\hat{x}$ (and a correspondent position $\hat{y}$) such that the mobile workers are indifferent between the two occupations if their types are below $\hat{x}$ and join $\mathcal{B}$ exclusively, if their types are above that threshold.

Claim 5. *Suppose $\mathcal{H}_{\mathcal{A}} \succsim_{LR} \mathcal{H}_{\mathcal{B}}$ and $s \in \left(\frac{h_{\mathcal{B}}(1)}{h_{\mathcal{A}}(1)+h_{\mathcal{B}}(1)}, \mathcal{H}_{\mathcal{A}}(\underline{y})\right)$. Then, there exists a unique $\hat{x} \in (0,1)$ such that $y_{\mathcal{A}}(\hat{x}) = \mathcal{H}_{\mathcal{A}}^{-1}(s + (1-s)\hat{x}) = y_{\mathcal{B}}(\hat{x}) = \mathcal{H}_{\mathcal{B}}^{-1}(1 - s + s\hat{x})$.*

Claim 5 above says that, under specified conditions, there is a threshold type $\hat{x}$ such that, if all the mobile workers above $\hat{x}$ were to join occupation $\mathcal{A}$ and all the constrained ones - occupation $\mathcal{B}$, $\hat{x}$ would get the same position in either of $\mathcal{A}$ and $\mathcal{B}$.

Proof. Let us argue that under $\frac{h_{\mathcal{B}}(1)}{h_{\mathcal{A}}(1)+h_{\mathcal{B}}(1)} < s < \mathcal{H}_{\mathcal{A}}(\underline{y})$, there must be exactly one intersection of $y_{\mathcal{A}}(x) = \mathcal{H}_{\mathcal{A}}^{-1}(s + (1-s)x)$ and $y_{\mathcal{B}}(x) = \mathcal{H}_{\mathcal{B}}^{-1}(1 - s + sx)$.[9] This is also equivalent to showing that function $D(y)$ admits a value of zero for some $y < 1$, where $D(y)$ is as previously defined in (75) of the proof for the case of $\overline{x}_C = 0$. Consider a talent type $x = 0$. If all mobile workers chose $\mathcal{A}$ and all constrained workers chose $\mathcal{B}$, a zero talent type would receive the position $\underline{y}_{\mathcal{A}}$ such that $\mathcal{H}_{\mathcal{A}}\left(\underline{y}_{\mathcal{A}}\right) = s$ in $\mathcal{A}$ and the position $\underline{y}_{\mathcal{B}}$ such that $\mathcal{H}_{\mathcal{B}}\left(\underline{y}_{\mathcal{B}}\right) = 1 - s$ in $\mathcal{B}$. Since $\mathcal{H}_{\mathcal{A}}\left(\underline{y}_{\mathcal{A}}\right) + \mathcal{H}_{\mathcal{B}}\left(\underline{y}_{\mathcal{B}}\right) = 1$ and $\mathcal{H}_{\mathcal{A}}(\underline{y}) + \mathcal{H}_{\mathcal{B}}(\underline{y}) = 1$, but $\mathcal{H}_{\mathcal{A}}(\underline{y}) > s$, it must hold that $\underline{y}_{\mathcal{A}} < \underline{y} < \underline{y}_{\mathcal{B}}$. Thus, a zero talent type would be allocated a higher position in $\mathcal{B}$ than in $\mathcal{A}$ under full occupational segregation. In order to get position paying value $\underline{y}_{\mathcal{B}}$ in occupation $\mathcal{A}$, a worker must be of type strictly greater than 0. Thus, $D(\underline{y}_{\mathcal{B}}) < 0$. Note also that $D(1) = 0$ (since in both occupations, only type 1 would get position 1). Check that $D'(1) = \frac{h_{\mathcal{A}}(1)}{s}\left(\frac{h_{\mathcal{B}}(1)}{h_{\mathcal{A}}(1)} - \frac{s}{1-s}\right) < 0$ under the assumption of $s > \frac{h_{\mathcal{B}}(1)}{h_{\mathcal{A}}(1)+h_{\mathcal{B}}(1)}$. Since $\frac{h_{\mathcal{B}}(y)}{h_{\mathcal{A}}(y)}$ is decreasing, $D'(y)$ can only change sign at most once, and if it does, the sign changes from positive to negative. Let $\tilde{y} \in (\underline{y}_{\mathcal{B}}, 1)$ be such that $D'(\tilde{y}) = 0$ Thus, there exists a unique y^* in $(\underline{y}_{\mathcal{B}}, \tilde{y})$ such that $D(y^*) = 0$. Then, the correspondent x^* follows as

$$x^* = \frac{\mathcal{H}_{\mathcal{B}}(y^*) - (1-s)}{s} = \frac{\mathcal{H}_{\mathcal{A}}(y^*) - s}{1-s}. \tag{79}$$

[9]Recall that $\underline{y}$ is defined as $\underline{y} = \mathcal{H}_{\mathcal{A}\cup\mathcal{B}}^{-1}(0)$.

□

Claim 6. *Suppose $\mathcal{H}_\mathcal{A} \succsim_{LR} \mathcal{H}_\mathcal{B}$ and $s \in \left(\frac{h_\mathcal{B}(1)}{h_\mathcal{A}(1)+h_\mathcal{B}(1)}, \mathcal{H}_\mathcal{A}(\underline{y})\right)$. Then, for $\hat{x}$ defined in claim 5, it holds that $\frac{h_\mathcal{B}(y^*(\hat{x}))}{h_\mathcal{A}(y^*(\hat{x}))+h_\mathcal{B}(y^*(\hat{x}))} > s$.*

What claim 6 implies, is that if mobile workers above $\hat{x}$ participated in $\mathcal{A}$ exclusively, while the constrained ones in $\mathcal{B}$ exclusively, the PAM allocation could still be achieved for workers below $\hat{x}$ by the mobile workers mixing between $\mathcal{A}$ and $\mathcal{B}$ with an appropriate probability.[10]

Proof. Consider two functions, $x_\mathcal{A}(y)$ and $x_\mathcal{B}(y)$:

$$x_\mathcal{A}(y) = \frac{\mathcal{H}_\mathcal{A}(y) - s}{1-s}, \quad x_\mathcal{B}(y) = \frac{\mathcal{H}_\mathcal{B}(y) - (1-s)}{s}. \tag{80}$$

Taking the derivatives of both functions with respect to y and evaluating them at y^*, we have

$$\left.\frac{dx_\mathcal{A}(y)}{dy}\right|_{y=y^*} = \frac{h_\mathcal{A}(y^*)}{1-s}, \quad \left.\frac{dx_\mathcal{B}(y)}{dy}\right|_{y=y^*} = \frac{h_\mathcal{B}(y^*)}{s}. \tag{81}$$

Since $D(y)$ crosses zero from below at y^* (as per proof of claim 5), $D'(y^*) = \left.\frac{dx_\mathcal{B}(y)}{dy}\right|_{y=y^*} - \left.\frac{dx_\mathcal{A}(y)}{dy}\right|_{y=y^*} > 0$, which means that $h_\mathcal{B}(y^*)/s > h_\mathcal{A}(y^*)/(1-s)$. The latter inequality can be re-arranged as

$$\frac{h_\mathcal{B}(y^*(\hat{x}))}{h_\mathcal{A}(y^*(\hat{x})) + h_\mathcal{B}(y^*(\hat{x}))} > s, \tag{82}$$

as desired. Note that another way to write the above inequality is

$$p_\mathcal{A}(\hat{x}) = \frac{h_\mathcal{A}(y^*(\hat{x})}{h_\mathcal{A}(y^*(\hat{x})) + h_\mathcal{B}(y^*(\hat{x}))} < 1-s, \tag{83}$$

□

Proof of of Proposition 3, part 1, case $x^{FS}(s) > 0$: From claim 5, we know that under the conditions, stated in Proposition 3, part 1, there is a unique $\hat{x} \in (0,1)$ such that $\mathcal{H}_\mathcal{A}^{-1}(s+(1-s)\hat{x}) = \mathcal{H}_\mathcal{B}^{-1}(1-s+s\hat{x})$. This value of $\hat{x}$ is the function $x^{FS}(s)$ stated in Proposition 3.

Suppose that all workers follow the profile, as the one described in the proposition. That is, the constrained workers always join $\mathcal{B}$ with probability one, while the mobile ones above $\hat{x}$ only join $\mathcal{A}$ with probability one and those below $\hat{x}$ are indifferent between $\mathcal{A}$ and $\mathcal{B}$ and join $\mathcal{A}$ with

[10]Namely, $p_\mathcal{A}(x; M) = \frac{h_\mathcal{A}(y^*(x))}{(h_\mathcal{A}(y^*(x))+h_\mathcal{B}(y^*(x)))(1-s)}$

probability

$$\tilde{p}_{\mathcal{A}}(x) = \frac{h_{\mathcal{A}}(y^*(x))}{(h_{\mathcal{A}}(y^*(x)) + h_{\mathcal{B}}(y^*(x)))(1-s)}. \tag{84}$$

It holds that $(1-s)\tilde{p}_{\mathcal{A}}(x) = p_{\mathcal{A}}(x)$, where $p_{\mathcal{A}}(x) = h_{\mathcal{A}}(y^*(x))/(h_{\mathcal{A}}(y^*(x))+h_{\mathcal{B}}(y^*(x)))$. Since $y^*(x)$ is increasing in x, as well as $h_{\mathcal{A}}(y)/(h_{\mathcal{A}}(y) + h_{\mathcal{B}}(y))$ is increasing in y, it holds that both $\tilde{p}_{\mathcal{A}}(x)$ and $p_{\mathcal{A}}(x)$ are both increasing. It follows from claim 6 that $p_{\mathcal{A}}(\hat{x}) < 1 - s$ and hence, $\tilde{p}_{\mathcal{A}}(\hat{x}) < 1$. Therefore, we only need to check that $p_{\mathcal{A}}(0) \geqslant 0$ to make sure that $\tilde{p}_{\mathcal{A}}(x)$ takes appropriate values for $x \in (0, \hat{x})$. But the former follows from $h_{\mathcal{A}}(y) \geqslant 0$. Overall, we have $\tilde{p}_{\mathcal{A}}(x) \in (0, 1)$, $x \in (0, \hat{x})$.

Let us now attempt to establish, which position would an arbitrary worker get in an assortative allocation if all other workers followed the above profile.

Consider first a contestant of type $x > \hat{x}$. In occupation $\mathcal{A}$, he would get $y_{\mathcal{A}}(x)$ such that $1-\mathcal{H}_{\mathcal{A}}(y_{\mathcal{A}}(x)) = (1-s)(1-x)$, while in occupation $\mathcal{B}$, he would get $y_{\mathcal{B}}(x)$ such that $1-\mathcal{H}_{\mathcal{B}}(y_{\mathcal{B}}(x)) = s(1-x)$. Note that $y_{\mathcal{A}}(1) = y_{\mathcal{B}}(1)$ and $y_{\mathcal{A}}(\hat{x}) = y_{\mathcal{B}}(\hat{x})$. Let's show that $y_{\mathcal{A}}(x) > y_{\mathcal{B}}(x)$, for $x \in (\hat{x}, 1)$. Following the steps we took in previous proofs, let $x_{\mathcal{A}}(y)$ and $x_{\mathcal{B}}(y)$ be the types of the workers, joining $\mathcal{A}$ and $\mathcal{B}$ respectively, that would get allocated the prize $y \in (y^*(\hat{x}), 1)$. Let, also, $D(y) = x_{\mathcal{B}}(y) - x_{\mathcal{A}}(y)$. Recall that $D'(y) = \frac{h_{\mathcal{A}}(y)}{s} \times (\frac{h_{\mathcal{B}}(y)}{h_{\mathcal{A}}(y)} - \frac{s}{1-s})$. Since $h_{\mathcal{B}}(y)/h_{\mathcal{A}}(y)$ is decreasing, $D'(y)$ can take a value of zero at most once in $y \in (y^*(\hat{x}), 1)$. Since $D(y^*(\hat{x})) = 0$ and $D(1) = 0$, it must be then that $D'(y^*(\hat{x})) \geqslant 0$ and $D'(1) \leqslant 0$. Thus, $D(y) \geqslant 0$ for $y \in (y^*(\hat{x}), 1)$, which means that in order to get the same-paying position in $\mathcal{B}$ as in $\mathcal{A}$, a higher type is required. The latter is equivalent to saying that the same type would get a better position in $\mathcal{A}$ than in $\mathcal{B}$, or, $y_{\mathcal{A}}(x) \geqslant y_{\mathcal{B}}(x)$, $x \in (\hat{x}, 1)$.

Let us now focus on the prizes that a type $x < \hat{x}$ would get in either contest, keeping in mind that the type $\hat{x}$ gets the same prize $y^*(\hat{x})$ in both $\mathcal{A}$ and $\mathcal{B}$. In occupation $\mathcal{B}$, for type x the correspondent prize $y_{\mathcal{B}}(x)$ follows from

$$\mathcal{H}_{\mathcal{B}}(y^*(\hat{x}) - \mathcal{H}_{\mathcal{B}}(y_{\mathcal{B}}(x)) = \int_x^{\hat{x}} (s + (1-s)\tilde{p}_{\mathcal{B}}(z))dz. \tag{85}$$

Using the fact that $s + (1-s)\tilde{p}_{\mathcal{B}}(x) = p_{\mathcal{B}}(x) = h_{\mathcal{B}}(y^*(x))/(h_{\mathcal{A}}(y^*(x)) + h_{\mathcal{B}}(y^*(x)))$ and that for $y^*(x)$, in current setting it holds that $F(x) = \mathcal{H}_{\mathcal{A}}(y^*(x)) + \mathcal{H}_{\mathcal{B}}(y^*(x)) - 1$, we can transform the integral on the left-hand side of the above expression as

$$\int_x^{\hat{x}} (s + (1-s)\tilde{p}_{\mathcal{B}}(z))dz = \int_x^{x^*} p_{\mathcal{B}}(z)dz = \int_x^{\hat{x}} \frac{h_{\mathcal{B}}(y^*(z))}{h_{\mathcal{A}}(y^*(z)) + h_{\mathcal{B}}(y^*(z))}dz = \tag{86}$$

$$= \int_{y^*(x)}^{y^*(\hat{x})} h_{\mathcal{B}}(y^*)dy^* = \mathcal{H}_{\mathcal{B}}(y^*(\hat{x}) - \mathcal{H}_{\mathcal{B}}(y^*(x)), \tag{87}$$

where we changed the variable of integration from z to y^*. From $\mathcal{H}_{\mathcal{B}}(y^*(\hat{x})) - \mathcal{H}_{\mathcal{B}}(y_{\mathcal{A}}(x)) =$

$\mathcal{H}_\mathcal{B}(y^*(\hat{x})) - \mathcal{H}_\mathcal{B}(y^*(x))$, we have that for $x < \hat{x}$, it must hold that $y_\mathcal{B}(x) = y^*(x)$. Repeating the same steps for $y_\mathcal{A}(x)$, we would get that $y_\mathcal{A}(x) = y^*(x)$ for workers $x < \hat{x}$ who join $\mathcal{A}$.

Summing up, we have that for the above profile, in an assortative allocation within each occupation, types below $\hat{x}$ would get $y^*(x)$ in each occupation, and types above $\hat{x}$ would get $y_\mathcal{A}(x)$ in $\mathcal{A}$ and $y_\mathcal{B}(x)$ in $\mathcal{B}$ such that $y_\mathcal{A}(x) \geqslant y_\mathcal{B}(x)$. The assortative allocation is supported by the education efforts which follow from the envelope condition and which are increasing due to the increasing nature of the allocation. The fact that the mobile workers are indifferent between $\mathcal{A}$ and $\mathcal{B}$ for $x < \hat{x}$ follows from

$$\int_0^x w(y_\mathcal{A}(z))dz = \int_0^x w(y_\mathcal{B}(z))dz = \int_0^x w(y^*(z))dz. \tag{88}$$

The fact that they prefer $\mathcal{A}$ for $x > \hat{x}$ follows from

$$\int_0^x w(y_\mathcal{A}(z))dz = \underbrace{\int_0^{\hat{x}} w(y_\mathcal{A}(z))dz}_{=\int_0^{\hat{x}} w(y_\mathcal{B}(z))dz} + \int_{\hat{x}}^x w(y_\mathcal{A}(z))dz \geqslant \int_0^{\hat{x}} w(y_\mathcal{B}(z))dz + \int_{\hat{x}}^x w(y_\mathcal{B}(z))dz. \tag{89}$$

The fact that all types of constrained workers prefer $\mathcal{A}$ follows from the fact that they would need to bear positive entry costs $\zeta_\mathcal{A}$ in $\mathcal{A}$ to have the same net utility as in $\mathcal{B}$ if they are below $\hat{x}$. The constrained ones above $\hat{x}$ prefer $\mathcal{B}$ because the most they can gain by switching to $\mathcal{A}$ is still less than $\zeta_\mathcal{A}$, by the condition that $\zeta_\mathcal{A} > \int_{\hat{x}}^1 \left(w(\mathcal{H}_\mathcal{A}^{-1}(s + (1-s)z)) - w(\mathcal{H}_\mathcal{B}^{-1}(1 - s + sz))\right) dz$.

A.5 Proof of Proposition 3, part 2

Let us define an auxiliary threshold:

$$x^\dagger(s) = \begin{cases} \{x : s = \frac{h_\mathcal{B}(y^*(x))}{h_\mathcal{A}(y^*(x)) + h_\mathcal{B}(y^*(x))}\} & \text{if } s < \frac{h_\mathcal{B}(y^*(0))}{h_\mathcal{A}(y^*(0)) + h_\mathcal{B}(y^*(0))}, \\ 0 & \text{if } s \geqslant \frac{h_\mathcal{B}(y^*(0))}{h_\mathcal{A}(y^*(0)) + h_\mathcal{B}(y^*(0))}. \end{cases} \tag{90}$$

$x^\dagger(s)$ is uniquely defined and weakly decreasing in s, because $\frac{h_\mathcal{B}(y^*(x))}{h_\mathcal{A}(y^*(x)) + h_\mathcal{B}(y^*(x))}$ is decreasing in x, given $\mathcal{H}_\mathcal{A} \succsim_{\text{LR}} \mathcal{H}_\mathcal{B}$, and achieves its maximum at $x = 0$. Now, consider an auxiliary profile such that above some type $\overline{x}_C$, there is PAM and below $\overline{x}_C$, there is segregation with C-workers joining $\mathcal{B}$ and M-workers joining $\mathcal{A}$. Then, assuming assortative matching within each occupation, the allocations below $\overline{x}_C$ follow from

$$\mathcal{H}_\mathcal{A}(y^*(\overline{x}_C)) - \mathcal{H}_\mathcal{A}(y_\mathcal{A}(x)) = (1-s)(\overline{x}_C - x), \tag{91}$$

$$\mathcal{H}_\mathcal{B}(y^*(\overline{x}_C)) - \mathcal{H}_\mathcal{B}(y_\mathcal{B}(x)) = s(\overline{x}_C - x), \tag{92}$$

or, equivalently

$$y_{\mathcal{A}}(x) = \mathcal{H}_{\mathcal{A}}^{-1}\left(\mathcal{H}_{\mathcal{A}}(y^*(\overline{x}_C)) - (1-s)(\overline{x}_C - x)\right), \tag{93}$$

$$y_{\mathcal{B}}(x) = \mathcal{H}_{\mathcal{B}}^{-1}\left(\mathcal{H}_{\mathcal{B}}(y^*(\overline{x}_C)) - s(\overline{x}_C - x)\right). \tag{94}$$

Claim 7. *Let $\overline{x}_C \geqslant x^\dagger(s)$. Then, either $y_{\mathcal{A}}(x) \geqslant y_{\mathcal{B}}(x)$ for all $x \in [0, \overline{x}_C)$, or, there exists a unique $\overline{x}_M \in (0, x^\dagger(s))$ such that $y_{\mathcal{A}}(x) \leqslant y_{\mathcal{B}}(x)$ for $x < \overline{x}_M$ and $y_{\mathcal{A}}(x) \geqslant y_{\mathcal{B}}(x)$ for $x \in (\overline{x}_M, \overline{x}_C)$.*

To see, why the above claim is true, consider a modified difference function, analogous to (75), but now defined for the auxiliary profile currently under attention:

$$D(y) = x_{\mathcal{B}}(y) - x_{\mathcal{A}}(y) = \frac{\mathcal{H}_{\mathcal{A}}(y^*(\overline{x}_C)) - \mathcal{H}_{\mathcal{A}}(y)}{1-s} - \frac{\mathcal{H}_{\mathcal{B}}(y^*(\overline{x}_C)) - \mathcal{H}_{\mathcal{B}}(y)}{s}. \tag{95}$$

As before, $D(y)$ measures the talent-differential between two workers obtaining identical jobs in sectors $\mathcal{A}$ and $\mathcal{B}$. Note that by construction $D(y^*(\overline{x}_C)) = 0$. Moreover,

$$D'(y)|_{y=y^*(\overline{x}_C)} = -\frac{h_{\mathcal{A}}(y^*(\overline{x}_C)}{1-s} + \frac{h_{\mathcal{B}}(y^*(\overline{x}_C))}{s} = \frac{h_{\mathcal{A}}(y^*(\overline{x}_C))}{s}\left(\frac{h_{\mathcal{B}}(y^*(\overline{x}_C))}{h_{\mathcal{A}}(y^*(\overline{x}_C))} - \frac{s}{1-s}\right) < 0, \tag{96}$$

which follows from $\overline{x}_C > x^\dagger$, the definition of $x^\dagger$, and the fact that $h_{\mathcal{B}}(y^*(x))/h_{\mathcal{A}}(y^*(x))$ is decreasing. Thus, $y_{\mathcal{A}}(x) \geqslant y_{\mathcal{B}}(x)$, for all x slightly below $\overline{x}_C$. Following the same logic as before, we thus have that either $y_{\mathcal{A}}(x) \geqslant y_{\mathcal{B}}(x)$ for all $x < \overline{x}_C$, or $y_{\mathcal{A}}(x)$ and $y_{\mathcal{B}}(x)$ cross exactly once, so that $\overline{x}_M > 0$ is given implicitly as the unique solution to the equation

$$y_{\mathcal{A}}(\overline{x}_M) - y_{\mathcal{B}}(\overline{x}_M) = 0. \tag{97}$$

In the latter case, $\overline{x}_M$ is decreasing as a function of $\overline{x}_C$ and for $\overline{x}_C = x^\dagger$, the two thresholds become equal, i.e. $\overline{x}_M = \overline{x}_C$. This follows from the fact that (97) is increasing not only in $\overline{x}_M$ but also in $\overline{x}_C$ for all $\overline{x}_C > x^\dagger$ because

$$(y_{\mathcal{A}}(x))'_{\overline{x}_C} = \frac{1}{h_{\mathcal{A}}(y_{\mathcal{A}}(x))}\left(\frac{h_{\mathcal{A}}(y^*(\overline{x}_C))}{h_{\mathcal{A}}(y^*(\overline{x}_C)) + h_{\mathcal{B}}(y^*(\overline{x}_C))} - (1-s)\right) \geqslant 0 \tag{98}$$

$$(y_{\mathcal{B}}(x))'_{\overline{x}_C} = \frac{1}{h_{\mathcal{B}}(y_{\mathcal{B}}(x))}\left(\frac{h_{\mathcal{B}}(y^*(\overline{x}_C))}{h_{\mathcal{A}}(y^*(\overline{x}_C)) + h_{\mathcal{B}}(y^*(\overline{x}_C))} - s\right) \leqslant 0, \tag{99}$$

where we have used that $(y^*(\overline{x}_C))'_{\overline{x}_C} = [h_{\mathcal{A}}(y^*(\overline{x}_C)) + h_{\mathcal{B}}(y^*(\overline{x}_C))]^{-1}$. It is also useful to notice that $\overline{x}_M = x^{FS}(s)$ when $\overline{x}_C = 1$.

Claim 8. *For any $\zeta_{\mathcal{A}} < \overline{\zeta}_{\mathcal{A}}(s)$ there exists a unique pair of thresholds $(\overline{x}_M, \overline{x}_C)$ such that*

$$\zeta_{\mathcal{A}} = \int_{\overline{x}_M}^{\overline{x}_C} \left(w(y_{\mathcal{A}}(x)) - w(y_{\mathcal{B}}(x))\right) dx, \tag{100}$$

with $y_{\mathcal{A}}(x)$ and $y_{\mathcal{B}}(x)$ as defined in (93)-(94).

To see why this claim is true, note that our analysis above implies that an increase in $\overline{x}_C$ strictly increases the integral in (100) because it raises the positive-valued integrand while reducing, as shown above, the lower limit of integration. The range of that integral is given by $(0, \overline{\zeta}_{\mathcal{A}}(s))$. This means that there exists a unique value $\overline{x}_C \in (x^{\dagger}, 1)$ for which the integral equals $\zeta_{\mathcal{A}}$, which together with the previous claim proves the existence of a unique pair of thresholds $(\overline{x}_M, \overline{x}_C)$ satisfying (100).

We now consider a second, modified, auxiliary profile, featuring PAM below $\overline{x}_M$ and PAM above $\overline{x}_C$, while between the two critical values of talent, the allocations are as in (93)-(94). This profile can be supported, as an equilibrium, by the mobile workers mixing appropriately below $\overline{x}_M$, and the constrained workers mixing appropriately above $\overline{x}_C$. The incentives of workers are readily satisfied, because workers with talents below $\overline{x}_M$ or above $\overline{x}_C$ are indifferent between sectors $\mathcal{A}$ and $\mathcal{B}$ by construction of the auxiliary allocations and given equality (100), whereas workers with talents between $\overline{x}_M$ and $\overline{x}_C$ strictly prefer sector $\mathcal{A}$ if they are mobile and $\mathcal{B}$ when they are constrained. It only remains to check that the entry-probabilities necessary to induce PAM below $\overline{x}_M$ or above $\overline{x}_C$ lie in $[0, 1]$.

When $\overline{x}_M > 0$, then for all $x < \overline{x}_M$, PAM requires mobile workers' probability of entering $\mathcal{A}$ to satisfy $p^*_{\mathcal{A}}(x) = (1 - s)p_{\mathcal{A}}(x; M)$ or

$$p_{\mathcal{A}}(x; M) = \frac{h_{\mathcal{A}}(y^*(x))}{h_{\mathcal{A}}(y^*(x)) + h_{\mathcal{B}}(y^*(x))} \frac{1}{1 - s}. \tag{101}$$

Since the likelihood ratio is increasing, and $x < \overline{x}_M \leqslant x^{\dagger}$, it follows, from the definition of $x^{\dagger}$ that

$$\frac{h_{\mathcal{A}}(y^*(x))}{h_{\mathcal{A}}(y^*(x)) + h_{\mathcal{B}}(y^*(x))} < \frac{h_{\mathcal{A}}(y^*(x^{\dagger}(s)))}{h_{\mathcal{A}}(y^*(x^{\dagger}(s)) + h_{\mathcal{B}}(y^*(x^{\dagger}(s))} = 1 - s, \tag{102}$$

and hence $p_{\mathcal{A}}(x, M) \in (0, 1)$. Similarly, for all $x > \overline{x}_C$, PAM requires the constrained workers' probability of entering $\mathcal{A}$ to satisfy $p^*_{\mathcal{A}}(x) = (1 - s) + sp_{\mathcal{A}}(x; C)$ or

$$p_{\mathcal{A}}(x; C) = 1 - \frac{h_{\mathcal{B}}(y^*(x))}{h_{\mathcal{A}}(y^*(x)) + h_{\mathcal{B}}(y^*(x))} \frac{1}{s}. \tag{103}$$

Since the likelihood ratio is decreasing and $x^\dagger \leq \overline{x}_C < x$ it hold that

$$\frac{h_\mathcal{B}(y^*(x))}{h_\mathcal{A}(y^*(x)) + h_\mathcal{B}(y^*(x))} < \frac{h_\mathcal{B}(y^*(x^\dagger(s)))}{h_\mathcal{A}(y^*(x^\dagger(s))) + h_\mathcal{B}(y^*(x^\dagger(s)))} \leqslant s, \tag{104}$$

which implies $p_\mathcal{A}(x, C) \in (0, 1)$.

The final remarks are regarding the condition for $\overline{x}_M$ being equal to zero. Note that $\overline{x}_M \leqslant x^\dagger(s)$. Thus $x^\dagger(s) = 0$ is a sufficient condition for $\overline{x}_M = 0$ (the former happens when $s \geqslant \frac{h_\mathcal{B}(y^*(0))}{h_\mathcal{A}(y^*(0))+h_\mathcal{B}(y^*(0))}$). However, $\overline{x_M} = 0$ can occur for lower s when $\zeta_\mathcal{A}$ is low enough, relative to s. Let us find the value of $\overline{x}_C$ and s such that $y_\mathcal{A}(x)$ and $y_\mathcal{B}(x)$ coincide exactly at zero. Let us denote such critical values by $s^\bigstar$ and $\overline{x}_C^\bigstar$. They follow from

$$\mathcal{H}_\mathcal{A}^{-1}\left(\mathcal{H}_\mathcal{A}(y^*(\overline{x}_C^\bigstar)) - (1-s)\overline{x}_C^\bigstar\right) = \mathcal{H}_\mathcal{B}^{-1}\left(\mathcal{H}_\mathcal{B}(y^*(\overline{x}_C^\bigstar)) - s\overline{x}_C^\bigstar\right). \tag{105}$$

When s is greater than $s^\bigstar$, given $\overline{x}_C^\bigstar$, $y_\mathcal{A}(0)$ is greater than $y_\mathcal{B}(0)$ and vice versa. On the other hand, the given $\overline{x}_C^\bigstar$, the critical value of $\zeta_\mathcal{A}$ (denoted, similarly, by $\zeta_\mathcal{A}^\bigstar$) follows from

$$\int_0^{\overline{x}_C^\bigstar} \left(w(y_\mathcal{A}(z)) - w(y_\mathcal{B}(z))\right) dz = \zeta_\mathcal{A}^\bigstar. \tag{106}$$

Taken together, equations (105)-(106) form a system with two equations and three unknowns, $(s^\bigstar, \overline{x}_C^\bigstar, \zeta^\bigstar)$. We can treat $s^\bigstar$ as an exogenous variable, that determines the remaining two variables $(\overline{x}_C^\bigstar, \zeta^\bigstar)$ as endogenous ones. Thus, we have a function $\zeta_\mathcal{A}^\bigstar(s^\bigstar)$. When, for a given $s^\bigstar$, a given $\zeta_\mathcal{A}$ is smaller than $\zeta_\mathcal{A}^\bigstar(s^\bigstar)$, $\overline{x}_C$ needs to become less than $\overline{x}_C^\bigstar(s^\bigstar)$, thus making $\overline{x}_M$ be greater than zero. Thus, $\zeta^\bigstar(s^\bigstar)$ gives us the critical value for the entry cost, above which, $\overline{x}_M$ is equal to zero. This applies when $s < \frac{h_\mathcal{B}(y^*(0))}{h_\mathcal{A}(y^*(0))+h_\mathcal{B}(y^*(0))}$. When $s \geqslant \frac{h_\mathcal{B}(y^*(0))}{h_\mathcal{A}(y^*(0))+h_\mathcal{B}(y^*(0))}$, $\overline{x}_M$ remains zero for any entry cost $\zeta_\mathcal{A}$.

To see that $\zeta_\mathcal{A}^\bigstar$ is decreasing with an increase of $s^\bigstar$, consider first the behavior of $\overline{x}_C^\bigstar$. Taking the full derivative of (105) with respect to $\overline{x}_C^\bigstar$ and $s^\bigstar$, and after transforming the terms, we get

$$\frac{d\overline{x}_C^\bigstar}{ds^\bigstar}\left(\frac{\frac{h_\mathcal{A}(y^*(\overline{x}_C^\bigstar))}{h_\mathcal{A}(y^*(\overline{x}_C^\bigstar))+h_\mathcal{B}(y^*(\overline{x}_C^\bigstar))} - (1-s)}{h_\mathcal{A}(\underline{y})} - \frac{\frac{h_\mathcal{B}(y^*(\overline{x}_C^\bigstar))}{h_\mathcal{A}(y^*(\overline{x}_C^\bigstar))+h_\mathcal{B}(y^*(\overline{x}_C^\bigstar))} - s}{h_\mathcal{B}(\underline{y})}\right) = -\left(\frac{\overline{x}_C^\bigstar}{h_\mathcal{A}(\underline{y})} + \frac{\overline{x}_C^\bigstar}{h_\mathcal{B}(\underline{y})}\right), \tag{107}$$

where the term on the left-hand side is positive, due to the previously employed properties of $\overline{x}_C$. Thus, $\overline{x}_C^\bigstar$ is decreasing with respect to $s^\bigstar$. Since this means that the integral on the left-hand side of (106) is also decreasing, we can conclude that $\zeta_\mathcal{A}^\bigstar$ is decreasing with respect to $s^\bigstar$.

A.6 Proof of Corrollary 2

Another interesting observation concerns the efforts of the constrained workers. It turns out that in equilibrium, sufficiently high-talented constrained workers always exert higher efforts than under PAM. To see this, consider the difference in effort for the highest constrained worker with $x = 1$:

$$\Delta\varepsilon(1, C) \equiv \varepsilon_{\mathcal{B}}(1; C) - \varepsilon^*(1) = w(y_{\mathcal{B}}(1)) - \int_0^1 w(y_{\mathcal{B}}(z))dz - w(y^*(1)) + \int_0^1 w(y^*(z))dz \tag{108}$$

$$= \int_0^1 \left[w(y^*(z)) - w(y_{\mathcal{B}}(z))\right] dz > 0. \tag{109}$$

Furthermore, consider the lowest type among the constrained workers, who would get the same job in equilibrium, as under PAM, $\overline{x}_C$. For the constrained workers close to $\overline{x}_C$ but below, we have

$$\Delta\varepsilon'_x(x, C) \simeq \overline{x}_C w'(y_{\mathcal{B}}(\overline{x}_C))(y_{\mathcal{B}}(\overline{x}_C))' - \overline{x}_C w'(y^*(\overline{x}_C))(y^*(\overline{x}_C))' = \tag{110}$$

$$= \overline{x}_C w'(y^*(\overline{x}_C)) \left(\frac{s}{h_{\mathcal{B}}(y^*(\overline{x}_C))} - \frac{1}{h_{\mathcal{A}}(y^*(\overline{x}_C)) + h_{\mathcal{B}}(y^*(\overline{x}_C))} \right) > 0, \tag{111}$$

while for the constrained workers in $[\overline{x}_C, 1]$ it holds that $\Delta\varepsilon(x, C)'_x = 0$. Thus, in equilibrium, the difference in effort that the constrained workers have to exert is the highest for types in $[\overline{x}_C, 1]$. Moreover, there is a unique threshold type $\hat{x}_C \in [0, \overline{x}_C)$ such that constrained workers above that type exert higher effort in equilibrium than under PAM. In order to demonstrate this, consider $\Delta\varepsilon'_x(x, C)$ for an arbitrary talent type x,

$$\Delta\varepsilon'_x(x, C) = x \left(w'(y_{\mathcal{B}}(x)) \frac{s}{h_{\mathcal{B}}(y_{\mathcal{B}}(x))} - w'(y^*(x)) \frac{1}{h_{\mathcal{A}}(y^*(x)) + h_{\mathcal{B}}(y^*(x))} \right), \tag{112}$$

and notice that due to $w()$ being concave, $w'(y_{\mathcal{B}}(x))$ is larger than $w'(y^*(x))$. Combined with the assumption of $\mathcal{H}_{\mathcal{A}} \gtrsim_{LR} \mathcal{H}_{\mathcal{B}}$, this ensures that $\varepsilon'_x(x, C)$ is monotonically increasing in x and thus, that $\varepsilon(x, C) > 0$ for $x > \hat{x}_C$.

It is also worth noting that if $w(.)$ is concave enough, $\Delta\varepsilon(x, C)$ can be positive for all $x \in [0, 1]$. For this, consider an example. Let $w(y) = \sqrt{y}$, $\mathcal{H}_{\mathcal{A}}(y) = y^2$, and $\mathcal{H}_{\mathcal{B}}(y) = y$, $s = 0.43$, and $\zeta_{\mathcal{A}}$ high enough so that none of the constrained workers are willing participate in $\mathcal{A}$. Under such setting, PAM job allocation is $y^*(x) = \frac{1}{2}\left(\sqrt{5 + 4x} - 1\right)$ and equilibrium allocation to constrained workers is $y_{\mathcal{B}}(x) = 1 - s + sx$. We have that $\Delta\varepsilon(x, C)$ is strictly increasing because $\Delta\varepsilon'_c(x, C) > 0$ for x

arbitrary close to zero:

$$\frac{\Delta\varepsilon'_x(x,C)}{x} \simeq \left(\frac{s}{2\sqrt{1-s+sx}} - \frac{1}{2\sqrt{\frac{1}{2}(\sqrt{5+4x}-1)}} \times \frac{1}{\sqrt{5+4x}}\right)\Bigg|_{x=0,s=0.43} \simeq 3.4\times 10^{-4} > 0. \tag{113}$$

This shows that $\Delta\varepsilon(x,C)$ can be positive for all types of constrained workers. Note that for this to occur, it is necessary that there is no "Diversification" by the low-ability mobile workers, since there must be a high enough gap between jobs awarded to type 0 in equilibrium and under PAM.

A.7 Missing Expressions

The expressions for the allocations in the illustrative example, assimilation case, are as follows:

$$y(x;C) = \begin{cases} \frac{1}{2}(\sqrt{5+4x}-1), & \text{if } x \geqslant \overline{x}_C \\ \frac{1}{2}(\sqrt{5+4\overline{x}_C}-1) - s(\overline{x}_C - x), & \text{if } x \in [\overline{x}_M, \overline{x}_C) \\ \frac{1}{2}(\sqrt{5+4x}-1), & \text{if } x < \overline{x}_M \end{cases} \tag{114}$$

$$y(x;M) = \begin{cases} \frac{1}{2}(\sqrt{5+4x}-1), & \text{if } x \geqslant \overline{x}_C \\ \sqrt{\frac{3}{2} + s\overline{x}_C + (1-s)x - \frac{1}{2}\sqrt{5+4\overline{x}_C}}, & \text{if } x \in [\overline{x}_M, \overline{x}_C) \\ \frac{1}{2}(\sqrt{5+4x}-1), & \text{if } x < \overline{x}_M. \end{cases} \tag{115}$$

The value of $\overline{x}_M$ is pinned down by $y(\overline{x}_M;C) = y(\overline{x}_M;M)$ and is given by $\overline{x}_M(\overline{x}_C) = 1/s^2 + \overline{x}_C - \sqrt{5+4\overline{x}_C}/s$. The value of $\overline{x}_C$ is pinned down by $\zeta_{\mathcal{A}} = \int_{\overline{x}_M(\overline{x}_C)}^{\overline{x}_C} w(y(z,M)) - w(y(z,C))dz$. This way, $\overline{x}_C$ is the lowest of the constrained contestants who becomes indifferent between the two contests. This region corresponds to part 2 (assimilation) of proposition 3, with $\overline{x}_M = 0$. Other than $\overline{x}_M = 0$ instead of $\overline{x}_M > 0$, this region coincides with the previous one. This time, the value of $\overline{x}_C$ is pinned down by $\zeta_{\mathcal{A}} = \int_0^{\overline{x}_C} w(y(z,M)) - w(y(z,C))dz$.

References

Akerlof, George A and Rachel E Kranton, "Identity and the Economics of Organizations," *Journal of Economic perspectives*, 2005, *19* (1), 9–32.

Azmat, Ghazala and Marc Möller, "The distribution of talent across contests," *The economic journal*, 2018, *128* (609), 471–509.

Becker, Gary S., "A Theory of Marriage. Part I," *Journal of Political Economy*, 1973, *81* (4), 813–846.

Blumberg, Henry, "On convex functions," *Transactions of the American Mathematical Society*, 1919, *20* (1), 40–44.

Charles, Maria and David B Grusky, *Occupational ghettos: The worldwide segregation of women and men*, Vol. 200, Stanford University Press Stanford, CA, 2005.

Delfino, Alexia, "Breaking Gender Barriers: Experimental Evidence on Men in Pink-Collar Jobs," *American Economic Review*, 2024, *114* (6), 1816–1853.

Hopkins, Ed, "Job Market Signalling of Relative Position, or Becker Married to Spence," *Journal of the European Economic Association*, 2012, *10* (2), 290–322.

Hoppe, Heidrun, Benny Moldovanu, and Aner Sela, "The Theory of Assortative Matching Based on Costly Signals," *The Review of Economic Studies*, 2009, *76*, 253–281.

Kuhn, Andreas and Stefan C Wolter, "Things versus people: Gender differences in vocational interests and in occupational preferences," *Journal of Economic Behavior & Organization*, 2022, *203*, 210–234.

Lazear, Edward P. and Sherwin Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 1981, *89* (5).

Milgrom, Paul and Chris Shannon, "Monotone comparative statics," *Econometrica: Journal of the Econometric Society*, 1994, pp. 157–180.

_ **and Ilya Segal**, "Envelope theorems for arbitrary choice sets," *Econometrica*, 2002, *70* (2), 583–601.

Morgan, John, Dana Sisak, and Felix Várdy, "The ponds dilemma," *The Economic Journal*, 2018, *128* (611), 1634–1682.

Olszewski, Wojciech and Ron Siegel, "Large contests," *Econometrica*, 2016, *84* (2), 835–854.

Shaked, Moshe and J. George Shanthikumar, *Stochastic orders*, Springer Science & Business Media, 2007.

Sierpiński, Wacław, "Sur les fonctions convexes mesurables," *Fundamenta Mathematicae*, 1920, *1* (1), 125–128.