I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14855

Shining with the Stars: Competition, Screening, and Concern for Coworkers' Quality

Francesca Barigozzi
Helmuth Cremer

NOVEMBER 2021

DISCUSSION PAPER SERIES

IZA DP No. 14855

Shining with the Stars: Competition, Screening, and Concern for Coworkers' Quality

Francesca Barigozzi
University of Bologna

Helmuth Cremer
Toulouse School of Economics and IZA

NOVEMBER 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Shining with the Stars: Competition, Screening, and Concern for Coworkers' Quality*

We study how workers' concern for coworkers' ability (CfCA) affects competition in the labor market. We consider two firms offering nonlinear contracts to a unit mass of prospective workers. Firms may differ in their marginal productivity, while workers are heterogeneous in their ability (high or low), and in their taste for being employed by any of the two firms. Workers receive a utility premium when employed by the firm hiring the workforce with larger average ability and they suffer a utility loss in the opposite case. These premiums/losses are endogenously determined. When workers' ability is observable and the diffrrence in firms' marginal productivities is strictly positive, we show that CfCA increases surplus but it also increases firms' competition for high-ability workers. As a result, CfCA benefits high-ability workers but is detrimental to firms. In addition, CfCA exacerbates the existing distortion in sorting of high-ability workers to firms: too many workers are hired by the least efficient firm. When ability is not observable, the additional surplus appropriated by high-ability workers is eroded by overincentivization (countervailing incentives) and the more so when CfCA is high. Conversely, high-types' sorting improves when CfCA is low and remains the same when it is high.

JEL Classification: D82, L13, M54

Keywords: concern for coworkers' quality, competition, screening, sorting

Corresponding author:
Helmuth Cremer
Toulouse School of Economics
University of Toulouse Capitole
1 Esplanade de l'Université
31080 Toulouse Cedex 06
France
E-mail: helmuth.cremer@tse-fr.eu

* This paper has been presented at the 20th Journées Louis-André Gérard-Varet, at the EEA-ESEM Virtual conference 2021 and at the 48th EARIE Annual Conference. We are grateful to seminar participants for their constructive comments. A special thanks to Giacomo Calzolari for his suggestions. Financial support from the Chaire "Marché des risques et creation de valeur" of the FdR/SCOR is gratefully acknowledged, together with the funding received by TSE from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program).

1 Introduction

Organizations are increasingly concerned with the quality of their workers as a source of competitive advantage and are particularly eager to attract top performers. Studies of financial-services professionals and of lawyers have shown that the skills and experience of top professionals constitute "general human capital" that contributes to the prestige of the organization; see Groysberg and Lee (2008). The same holds true for professional business services firms and research institutions. As a result, many organizations compete aggressively to attract top performers. And, in some industries, the mobility of top workers is further enhanced by their direct and exclusive relationship to their clients, who are loyal to the professionals providing the service rather than to the firm that employs them. In the same way, highly productive researchers bring their ability and their network to the organization, inside and outside Academia, contributing to its success and attractiveness.

On the supply side of the market, job seekers choose which organization to work for. In the existing economic literature, this decision generally depends on the applicants' preferences for the organizations offering a vacancy and on the monetary compensation associated with the posted job offers. We innovate by assuming that workers' choice also depends on the quality of their coworkers. Specifically, we assume that the attractiveness of an organization increases with the quality of its workforce. Think about young lawyers who just graduated from a prestigious Law School. To which law firm should they apply for a position? Cravath or Skadden? The choice naturally depends on the offered salaries and on the two firms' amenities, but it is likely to be also affected by the (endogenous) average skills of the lawyers employed by each of the two firms.

Why is workers' utility increasing in their coworkers' quality? First, workers' utility may be increasing in the share of high-ability coworkers because being employed in an organization that hires a qualified workforce increases the workers' future career prospects outside the firm, for example because of a signaling mechanism. Second, working with top professionals may give preferential access to resources, opportunities, and general perks/benefits inside and outside the organization.[1] Third, top workers bring social status to the firm and the latter may be a source of utility *per se* for its employees. Note that we disregard complementarities and possible spillovers in term of productivity, which have been considered before, mainly in the management literature (see Groysberg and Lee, 2008, Ertug et al., 2018, Tan and Netessine, 2019).

We take a first step towards analyzing the role played by the concern for coworkers' ability (CfCA) in the hiring process. We interpret such concern as a utility premium accruing to

[1] Think about the increased opportunity to have access to research funds and external contracts in renown research institutions that have a strong reputation and visibility thanks to their top researchers and international experts.

workers employed in the organization hiring the majority of high-ability workers. In a labor market where organizations compete to attract the best workers by offering them nonlinear contracts, we investigate how CfCA affects workers' selection. By doing so we want to address the following questions. How does CfCA affect competition to attract the best talents? How does it shape nonlinear contracts and workers' sorting between competing firms? How does workers' private information on ability (and the subsequent screening designed by employers) affect workers' sorting when CfCA matters?

To study these questions we consider two firms and a unit mass of prospective workers. Firms may differ in their marginal productivity while workers are heterogeneous with respect to their ability, high or low, and with respect to their taste for being employed by any of the two firms. In addition, high-ability workers care for the ability of their colleagues.[2] Specifically, their utility increases if they are employed by the firm hiring the larger share of high-ability workers, and it decreases if they are employed by the firm hiring the lower share of high-types. Firms compete to attract workers by offering nonlinear contracts. Optimal contracts are contingent on workers' ability and are designed in the *utility space* so that they are characterized by the (gross) indirect utility (rent) offered to the worker, and by the worker's labor supply which corresponds to an observable and contractible level of effort. Workers' sorting depends on the relative magnitude of indirect utilities offered by each firm to workers of different ability.

We first derive the labor market equilibrium when workers' ability is observable, but their taste for firms is not. This case is relevant for the senior job-market where candidates' previous outcomes are observable (e.g. successful lawsuits for a lawyer and the publications list for a researcher). We find that, when firms are identical, CfCA does not affect surplus because firms equally share the workforce of both types and neither premiums nor utility losses emerge. When instead firms are heterogeneous, CfCA matters because workers' sorting to firms is asymmetric. The more productive firm hires a larger share of high-types and, to a lower extent, also a larger share of low-types. As a result, the more efficient firm always hires the workforce characterized by the higher average ability. Here CfCA increases total surplus and high-ability workers' utility but it reduces both firms' profits. Intuitively, CfCA increases competition for high-ability workers by reducing their mismatch disutility, and is thus detrimental to firms. If CfCA is sufficiently large, a corner solution emerges, where the more efficient firm hires all high-types.

We derive the allocation that maximizes an utilitarian social welfare function and compare it to the market equilibrium. Workers' sorting is always inefficient when the two firms are heterogeneous. Three different distortions of marginal workers sum up in the market equilibrium, each of them results in having too many workers employed by the least efficient firm. The first

[2]In an extension of the model we also study the case in which both high-and low-ability workers care about the ability of their coworkers.

distortion is caused by profit maximization: firms disregard mismatch disutility of all the workers except the marginal ones. The second one depends on strategic interaction: the least efficient firm competes too aggressively while the most efficient one accommodates too much. The third distortion is the one generated by CfCA (and, again, strategic interaction); the latter implies a positive externality for workers employed by the more efficient firm and a negative one for workers hired by the least efficient firm that are only partially internalized in equilibrium.

We then derive screening contracts and workers' sorting when neither taste for firms nor workers' ability are observable by the firms. Private information on ability is relevant for instance in the case of junior job-market applicants who had no opportunities yet to prove their talent in practice. We show that, if the two firms are identical and/or CfCA is sufficiently low, then the market allocation is incentive compatible. In case firms have different productivities the market allocation continues to be incentive compatible when CfCA is sufficiently low and firms' heterogeneity is sufficiently lower than workers' heterogeneity. Otherwise, the market allocation is not incentive compatible and, depending on which incentive constraints are binding, one of three different regimes emerges. Regime 1 realizes for low values of CfCA and low workers' heterogeneity, Regime 2 emerges for intermediate values of CfCA, while Regime 3 emerges for sufficiently larger levels of CfCA. In the three regimes, obviously screening contracts entail some upward/downward distortions of effort levels. Consequently, inefficient effort levels obtain on top of the distortions in workers' sorting.

In Regime 1 and 2, workers' sorting differs from the one obtained under full information on ability because screening contracts alter the difference between indirect utilities that firms offer to the workers. Specifically, in Regime 1 and 2, the share of high-types hired by the more efficient firm increases and, as a result, distortions in the sorting of high-ability workers decrease with respect to the full information market equilibrium. Conversely, the share of low-types hired by the more efficient firm falls so that distortions in sorting of low-ability workers increase. Sorting obtained under Regime 2 is overall less distorted than the one obtained under Regime 1. In Regime 3 sorting remains the same as under full information. Countervailing incentives emerge in all three regimes. In regimes 2 and 3, high-ability workers are worse off than under full information because of upward distorted effort levels and lower indirect utilities. Low-ability workers on the other hand are better off because their utility increases. Results are ambiguous to this respect in Regime 1.

We conclude that, when firms have different marginal productivities, CfCA increases surplus but it also increases firms' competition for high-ability workers. As a result, CfCA benefits high-ability workers but is detrimental to firms. CfCA increases the existing distortion in sorting of high-ability workers to firms: too many workers are hired by the least efficient firm. When ability is not observable, screening contracts are such that this distortion decreases when CfCA is low and remains unchanged when CfCA is high. In addition, countervailing incentives (partially) erode the additional surplus appropriated by high-ability workers in the full information equilibrium, and the more so when CfCA is high.

Considering that the case where ability is observable can be interpreted as the study of selection of senior job market candidates, while the case of private information may correspond to selection of junior candidates, we observe the following. CfCA empowers all *senior* talented job market applicants, including the ones employed by the least efficient firm, but *junior* talented applicants entering the labor market for the first time are not able to appropriate the same surplus. The latter is substantially eroded by screening contracts which imply lower rent for and overincentivization of talented workers.

As mentioned before, in the main text we study a specification of the model where only high-ability workers care for their coworkers' ability. This specification is tractable and intuitions are easy to grasp. In Appendix A.9, we study a richer version of the model, where both workers' types are concerned with their coworkers' ability. We show that our simplified model is able to capture all main results on the market equilibrium and on workers' sorting obtained with the richer specification.

1.1 Related literature

From an analytical point of view, our paper draws from the literature on multi-principals initiated by the seminal contributions of Martimort (1992) and Stole (1992). Within this literature, the paper that is most closely related to ours is Rochet and Stole (2002) which extends the analysis carried out in Stole (1995) and studies duopolists competing in nonlinear prices in the presence of both vertical and horizontal preference uncertainty.[3] We depart from Rochet and Stole (2002) in that they only consider symmetric firms and thus find that incentive compatibility constraints are always slack for all firms, so that efficient quality allocations with cost-plus-fixed-fee pricing emerge in equilibrium.[4]

In the literature on workers' selection, the papers closest to ours are Bénabou and Tirole (2016) and Barigozzi and Burani (2019). Bénabou and Tirole (2016) embed multitasking and screening in a Hotelling framework. Workers engage in two activities, one in which individual contributions are not measurable and are driven by motivation, and the other which is contractible and dependent upon a worker's ability. When motivation is observable, while ability is

[3] Two other papers analyzing optimal contracts by multiple principals that are related are Biglaiser and Mezzetti (2000) and Lehmann *et al.* (2014). The former studies an incentive auction in which multiple principals bid for the exclusive services of an agent, who has private information about ability. The latter considers optimal nonlinear income taxes levied by two competing governments.

[4] Precisely the same result can be found in Armstrong and Vickers (2001) who model firms as directly supplying utility to consumers.

private information, equilibria range from the case of monopsonistic underincentivization of low-skilled work to the other extreme case of perfectly competitive overincentivization of high-skilled work. With respect to that paper we innovate in several directions. First, we introduce CfCA in the workplace. Second, we consider heterogeneous firms. Third, in our setup, workers' taste for firms is not observable, it influences the sorting of workers into firms and interacts with skills in determining incentive pay in equilibrium. In terms of results, we share with Bénabou and Tirole (2016) the fact that competition for the most talented workers generates countervailing incentives for high-ability types. We find screening contracts similar to the ones of Bénabou and Tirole (2016) as a special case (see our Regime 3). Specifically, when CfCA is sufficiently large, we show that both firms distort the effort of high-ability workers upward. However, the interaction between firms' heterogeneity and CfCA generates new results: for low relevance of CfCA, we find equilibria where the least efficient firm always distorts effort of high-ability workers upward and the more efficient firm may or may not distort effort of low-skilled workers downward.

Barigozzi and Burani (2019) study a setting with a for-profit and a non-profit firm competing to attract workers who are intrinsically motivated to contribute to the mission of the non-profit firm.[5] The setting of the two papers presents some similarities because workers differ in ability and in a second characteristic which corresponds to intrinsic motivation in Barigozzi and Burani (2019) and in "taste for firms" in this papers. In the two papers both characteristics are the workers' private information and intrinsic motivation in Barigozzi and Burani (2019) is uniformly distributed among the applicants, like the taste for firms in the present setting. However, our setting is different because CfCA generates a peer effect in workers' preferences which translates into an additional interdependence in labor demands of the two firms. This is why the equilibrium set of optimal screening contracts is richer in our paper than in Barigozzi and Burani (2019).

2 The model

We study a Hotelling-like competitive screening model, where workers care about the ability of their coworkers. Two firms compete to hire workers: firm A is located at zero whereas firm B is located at 1. Each worker (she) can work exclusively for one firm and supplies effort, which represents the only input necessary to produce. Firms and workers are risk neutral.

Firms

[5] Barigozzi and Burani (2019) is in turn related to Barigozzi and Burani (2016). The latter considers output-oriented motivation, so that a worker's intrinsic satisfaction depends on her personal contribution to the output produced. In Barigozzi and Burani (2019) instead, workers' motivation does not depend on effort (or output) provision so that the single-crossing condition holds and, like in the current paper, firms only screen workers for their ability.

Let x denote the *observable and measurable* effort level that workers are asked to provide. Firms' production functions display constant returns to effort and the amount of output produced is $q_i(x) = k_i x$ for firm $i = A, B$, where the marginal product of labor k_i is firm-specific. Without generality loss we assume that firm A has a weak competitive advantage so that $k_A \geq k_B$.

Payoffs *per-worker*, conditional on the worker being hired, are given by

$$\pi_i(x) = q_i(x) - w_i(x) = k_i x - w_i(x), \tag{1}$$

where $w_i(x)$ is the wage or salary paid by firm i to the worker exerting effort x and where the unit price of output is exogenous and set to 1.

Workers

There is a unit mass of workers who are uniformly distributed on the Hotelling line. They differ in two characteristics: *ability* and the *taste for firms*. Ability is inversely related to the cost of providing effort and is denoted as θ_j, with $\theta_j \in \{\theta_1, \theta_2\}$, where $\theta_2 > \theta_1$ and $\Delta\theta \equiv \theta_2 - \theta_1$. A fraction λ_1 of workers has a low cost of effort (i.e., high ability) θ_1 and a fraction $\lambda_2 = 1 - \lambda_1$ has a high effort cost (i.e., low ability) θ_2. Workers' average ability is denoted by $\mathrm{E}(\theta) = \lambda_1\theta_1 + \lambda_2\theta_2$. The mismatch disutility depends on the worker's location on the Hotelling line γ, which is uniformly distributed on the interval $[0, 1]$, and by the cost per unit distance σ.

Let us define $\widehat{\gamma}_j$, where $0 \leq \widehat{\gamma}_j \leq 1$, $j = 1, 2$, the type-specific marginal worker who is indifferent between being hired by firm A and by firm B. Given that firm A is located in 0 and firm B is located in 1, $\lambda_1\widehat{\gamma}_1 + \lambda_2\widehat{\gamma}_2$ is the workforce employed by firm A while $\lambda_1(1 - \widehat{\gamma}_1) + \lambda_2(1 - \widehat{\gamma}_2)$ is the workforce employed by firm B.

We innovate with respect to the existing literature by assuming that workers care about their coworkers' ability. Specifically, high-ability workers receive a utility premium if their employer hires a majority of them and suffer a disutility otherwise. To be precise, take firm A and the share of its high-type employees, namely $\widehat{\gamma}_1$. When $\widehat{\gamma}_1 > 1/2$ high-ability workers hired in firm A receive a premium, whereas high-ability workers hired by the competitor suffer a loss of the same amount. When $\widehat{\gamma}_1 < 1/2$ premium and loss are reversed. Because high-ability workers receive a benefit that is increasing in the share of colleagues of the same type, we can say that the workplace displays homophyly among high-skill workers.[6] In the conclusion we discuss possible job-market mechanisms resulting in a workers' utility function which increases with the quality of their coworkers.

[6] Using the terminology of the Management Literature, our high-ability workers can be interpreted as "status stars" who bring social status to their peers (as opposed to "performance stars" who increase the overall performance of the organization); see Kehoe et al. (2016) and the references within.

In the Economic Literature, a recent paper by Bolte et al. (2020) studies the consequences of homophily in the worplace. In their setting, referrals and homophily lead to social immobility. Specifically, a demographic group's low current employment rate leads that group to have relatively low future employment as well.

The workers' utility function when hired by firm A and by firm B, respectively, are given by:

$$u_A(x_A, w_A; \theta_j, \gamma) = w_A(x_A) - \frac{1}{2}\theta_j x_A^2 \qquad - \gamma\sigma \qquad + \alpha_j\left(\hat{\gamma}_1 - \tfrac{1}{2}\right), \quad (2)$$

$$u_B(x_B, w_B; \theta_j, \gamma) = \underbrace{w_B(x_B) - \frac{1}{2}\theta_j x_B^2}_{\text{net compensation } U_i(\theta_j)} \qquad \underbrace{-(1-\gamma)\sigma}_{\text{mismatch disutility}} \qquad \underbrace{-\alpha_j\left(\hat{\gamma}_1 - \tfrac{1}{2}\right)}_{\text{concern for coworkers' ability}}, \quad (3)$$

where the relevance of CfCA is represented by the parameter $\alpha_1 \geq 0$, while $\alpha_2 = 0$. In words, workers utilities depend on *(i)* their net compensation, $U_i(\theta_j)$,i.e., the salary less the cost of effort provision *(ii)* their mismatch disutility and, for high-ability workers, *(iii)* the utility premium (loss) when their coworkers include the majority (the minority) of the high-skill workforce. Hence, CfCA translates into a premium for high-types if their employer is able to hire a larger share of high-ability workers than its competitor and in a utility loss suffered by high-types employed by the firm hiring the lower share of them.[7] Note that, when employed by firm B, a high-type worker's premium for coworkers' ability is $+\alpha_1((1-\hat{\gamma}_1) - 1/2) = -\alpha_1(\hat{\gamma}_1 - 1/2)$. In Appendix A.9 we present and discuss a richer specification of the utility function with $\alpha_2 > 0$ so that premiums or losses from coworkers' ability accrue to both types of workers.[8]

The average ability of workers employed by firm $i = A, B$, $\mathrm{E}_i(\theta)$, writes:

$$\mathrm{E}_A(\theta) = \frac{\lambda_1\theta_1\widehat{\gamma}_1 + \lambda_2\theta_2\widehat{\gamma}_2}{\lambda_1\widehat{\gamma}_1 + \lambda_2\widehat{\gamma}_2},$$

$$\mathrm{E}_B(\theta) = \frac{\lambda_1\theta_1(1-\widehat{\gamma}_1) + \lambda_2\theta_2(1-\widehat{\gamma}_2)}{\lambda_1(1-\widehat{\gamma}_1) + \lambda_2(1-\widehat{\gamma}_2)}.$$

Note that *a more efficient workforce is characterized by a lower* $\mathrm{E}_i(\theta)$, $i = A, B$, because $\theta_2 > \theta_1$. The following three possible workers' sorting patterns exist.

Lemma 1 ***Workers' sorting.***
(i) when $\widehat{\gamma}_1 = \widehat{\gamma}_2$, each firm hires the same share of high- and low-ability workers and the average ability of the workforce is the same for the two firms: $E_A(\theta) = E_B(\theta) = E(\theta)$.
(ii) when $\widehat{\gamma}_1 > \widehat{\gamma}_2$, firm A hires a larger share of high- than of low-ability workers so that it employes the workforce with the higher average ability: $E_A(\theta) < E_B(\theta)$.
(iii) when $\widehat{\gamma}_1 < \widehat{\gamma}_2$ firm A hires a lower share of high- than of low-ability workers so that it employes the workforce with the lower average ability: $E_A(\theta) > E_B(\theta)$.

Expressions (2) and (3) imply that neither the mismatch disutility nor CfCA are related to effort exertion and they do not affect directly the firm's output. This implies that a worker's

[7] The fact that $\hat{\gamma}_1 > 1/2$ does not necessarily imply that firm A's workforce has a larger average ability than firm B. Indeed this requires that $\hat{\gamma}_1 > \hat{\gamma}_2$ as we show below.

[8] Referring again to the Management literature, in this case "status stars" bring social status not only to their peers but to all colleagues.

indifference curves have positive slope in the (x, w) plane and that the single-crossing property holds, no matter the hiring firm.

Contracts and screening mechanism

Anticipating the workers' decisions, firms $i = A, B$ offer incentive-compatible non-linear wage schedules $w_i(x_i)$ that are conditional on the effort target. Recall that workers of any type θ_j have preferences over effort-salary pairs which are independent of γ and of $\widehat{\gamma}_1$, (conditional on being hired by a given firm). To determine the wage schedules we study the direct revelation mechanism such that each firm offers two incentive-compatible contracts, one for each ability type θ_j, consisting in an effort target and a wage rate, i.e. $\{x_i(\theta_j), w_i(\theta_j)\}_{i=A,B;\ j=1,2}$. The contracts offered by the two firms, determine the indirect (gross) utilities of a worker who truthfully reports her ability type θ_j. We then use these to tackle the worker's self-selection problem across firms, which depends on mismatch disutility γ and on the concern for the coworkers quality $(\widehat{\gamma}_1 - 1/2)$. We thus treat the firms' contract design problem as independent of the workers' choice about which firm to work for. The latter is considered as an indirect mechanism, because no report on γ is required. Finally, it is convenient to focus on workers' indirect utility $U_i(\theta_j)$, gross of the mismatch disutility and of the premium for coworkers quality. Consequently, we derive contracts of the form $\{x_i(\theta_j), U_i(\theta_j)\}_{i=A,B;\ j=1,2}$.

2.1 Marginal workers

Given the non-linear wage schedule $w_i(x_i)$ offered by firms $i = A, B$, a worker of type θ_j employed by firm i, solves

$$\max_{x_i} \quad w_i(x_i) - \frac{1}{2}\theta_j x_i^2.$$

Denoting by $x_i(\theta_j)$ the solution to this, one can write

$$U_i(\theta_j) = w_i(x_i(\theta_j)) - \frac{1}{2}\theta_j x_i^2(\theta_j), \tag{4}$$

where $U_i(\theta_j)$ is the *indirect utility* of an agent of type θ_j who is hired by firm i, *absent* the mismatch disutility and the premium/loss from coworkers' ability. Hence, a worker of type (θ_j, γ) gets *total* indirect utility

$$\mathcal{U}_A(\theta_j) = U_A(\theta_j) - \gamma\sigma + \alpha_j\left(\hat{\gamma}_1 - \tfrac{1}{2}\right) \tag{5}$$

if employed by firm A and *total* indirect utility

$$\mathcal{U}_B(\theta_j) = U_B(\theta_j) - (1-\gamma)\sigma - \alpha_j\left(\hat{\gamma}_1 - \tfrac{1}{2}\right) \tag{6}$$

if employed by firm B.

The participation constraints require that

$$\mathcal{U}_A(\theta_j) \geq 0 \qquad \text{and} \qquad \mathcal{U}_B(\theta_j) \geq 0 \qquad \text{for all } \theta_j \in \{\theta_1, \theta_2\}. \tag{PC}$$

When the market is fully covered, given firm i's offer, the outside option of each type of worker is represented by the contract offered by the rival firm $-i$.[9]

We are now in the position to determine the share of workers of each type employed by the two firms.

The worker who is indifferent between working for firm A and for firm B is $\widehat{\gamma}_j$ such that $\mathcal{U}_A(\theta_j) = \mathcal{U}_B(\theta_j)$, $j = 1, 2$. Using (5) and (6) yields

$$\widehat{\gamma}_1 = \frac{1}{2} + \frac{U_A(\theta_1) - U_B(\theta_1)}{2(\sigma - \alpha_1)}, \tag{7}$$

$$\widehat{\gamma}_2 = \frac{1}{2} + \frac{U_A(\theta_2) - U_B(\theta_2)}{2\sigma}. \tag{8}$$

Note that, when $\alpha_1 = 0$ we return to the standard Hotelling labor demands: $\widehat{\gamma}_j = 1/2 + (U_A(\theta_j) - U_B(\theta_j))/2\sigma$, $j = 1, 2$. When instead $\alpha_1 = \sigma$, the marginal worker of type θ_1 is indeterminate. If $\alpha_1 < \sigma$, high-ability workers' CfCA is not so strong to reverse the standard Hotelling "forces" in (7) and an interior solution for $\widehat{\gamma}_1$ is possible. Formally, when it comes to the determination of the marginal worker, α_1 is equivalent to a reduction in the mismatch disutility.

Interestingly, when $\alpha_1 > \sigma$, CfCA leads to a corner solution with all high-types employed by one firm. To see this, use (5), to write the utility of the high-type marginal worker when hired by firm A as:

$$\mathcal{U}_A(\theta_1, \widehat{\gamma}_1) = U_A(\theta_1) - \widehat{\gamma}_1\sigma + \alpha_1\left(\widehat{\gamma}_1 - \tfrac{1}{2}\right) = U_A(\theta_1) - \widehat{\gamma}_1(\sigma - \alpha_1) - \tfrac{1}{2}\alpha_1. \tag{9}$$

From (9), when $\alpha_1 > \sigma$ we have that $\mathcal{U}_A(\theta_1)$ is monotonically increasing in $\widehat{\gamma}_1$. In addition, $\mathcal{U}_A(\theta_1, \gamma) \geq \mathcal{U}_A(\theta_1, \widehat{\gamma}_1)$ $\forall \gamma \leq \widehat{\gamma}_1$ because the mismatch disutility is lower for workers located to the left of the marginal worker $\widehat{\gamma}_1$ but the premium/loss for coworkers' quality is the same as for worker $(\theta_1, \widehat{\gamma}_1)$. In other words, when $\alpha_1 > \sigma$, the utility of type $(\theta_1, \widehat{\gamma}_1)$ hired by A, and that of all types with $\gamma < \widehat{\gamma}_1$, increases monotonically with $\widehat{\gamma}_1$. Similarly, the utility of type $(\theta_1, \widehat{\gamma}_1)$ hired by B, and that of all types located on the right of $\widehat{\gamma}_1$, decreases monotonically with $\widehat{\gamma}_1$.

[9] The option that workers prefer to remain unemployed is excluded by assuming that the market is fully covered or that

$$U_A(\theta_j) + U_B(\theta_j) > \sigma, \forall j = 1, 2;$$

see Rochet and Stole (2002, page 290). This is equivalent to say that the *total* utilities of the marginal workers are non-negative: $\mathcal{U}_A^*(\theta_j, \widehat{\gamma}_j^*) = \mathcal{U}_B^*(\theta_j, \widehat{\gamma}_j^*) \geq 0$, $j = 1, 2$, where $\widehat{\gamma}_j^*$ is the marginal worker of type j in equilibrium. In our setting, these inequalities hold if k_i, $i = A, B$ is sufficiently larger than σ, which we assume. See also our comments before Proposition 1.

Thus, if $U_A(\theta_1) > U_B(\theta_1)$ there is corner solution with $\widehat{\gamma}_1 = 1$. If $U_A(\theta_1) < U_B(\theta_1)$ the corner solution entails $\widehat{\gamma}_1 = 0$.

To better understand the impact of α_1 on workers' sorting when $\sigma > \alpha_1$, or when an interior solution for $\widehat{\gamma}_1$ is possible, let us consider (7) and (8) and observe that

$$\widehat{\gamma}_1 - \widehat{\gamma}_2 \geq 0 \quad \text{if and only if} \quad \frac{(U_A(\theta_1) - U_B(\theta_1)) - (U_A(\theta_2) - U_B(\theta_2))}{2(\sigma - \alpha_1)} \geq 0.$$

Hence we can state the following Lemma.

Lemma 2 *If $\alpha_1 < \sigma$, then:*

$$E_A(\theta) \leq E_B(\theta) \Longleftrightarrow U_A(\theta_1) - U_A(\theta_2) \geq U_B(\theta_1) - U_B(\theta_2) \tag{10}$$

and $E_A(\theta) = E_B(\theta)$ requires that $U_A(\theta_1) - U_B(\theta_1) = U_A(\theta_2) - U_B(\theta_2)$. A sufficient (but not necessary) condition is $U_A(\theta_1) = U_B(\theta_1)$ and $U_A(\theta_2) = U_B(\theta_2)$.

In other words, if $\alpha_1 < \sigma$, firm A is able to hire a better workforce if and only if it offers high-types a larger return to ability than its competitor $(U_A(\theta_1) - U_A(\theta_2) \geq U_B(\theta_1) - U_B(\theta_2))$.

3 Equilibrium contracts when taste for firms is not observable

Suppose that $\alpha_1 < \sigma$ and that workers' ability is observable, while mismatch disutility γ is the workers' private information. We derive optimal contracts $\{x_i(\theta_j), U_i(\theta_j)\}_{i=A,B;\ j=1,2}$ under full information on ability.

Let us write the firms' profits as a function of the workers' utility. Solving (4) for the wage rate:

$$w_i(\theta_j) = U_i(\theta_j) + \frac{1}{2}\theta_j x_i^2(\theta_j). \tag{11}$$

Plugging the previous expression into the firms' payoffs (1), we can rewrite per-worker profits relative to each type θ_j as

$$\pi_i(\theta_j) = k_i x_i(\theta_j) - \frac{1}{2}\theta_j x_i^2(\theta_j) - U_i(\theta_j). \tag{12}$$

Each firm maximizes profits obtained by multiplying (12) with their workforce determined by expressions (7) and (8). Hence, firm A and B respectively solves the following program:

$$\begin{aligned} \max_{\{x_A(\theta_j),U_A(\theta_j)\}_{j=1,2}} \pi_A = \; & \lambda_1 \left(\tfrac{1}{2} + \tfrac{U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)}\right)\left(k_A x_A(\theta_1) - \tfrac{1}{2}\theta_1 x_A^2(\theta_1) - U_A(\theta_1)\right) \\ & + \lambda_2 \left(\tfrac{1}{2} + \tfrac{U_A(\theta_2)-U_B(\theta_2)}{2\sigma}\right)\left(k_A x_A(\theta_2) - \tfrac{1}{2}\theta_2 x_A^2(\theta_2) - U_A(\theta_2)\right) \end{aligned}$$

$$\begin{aligned} \max_{\{x_B(\theta_j),U_B(\theta_j)\}_{j=1,2}} \pi_B = \; & \lambda_1 \left(\tfrac{1}{2} - \tfrac{U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)}\right)\left(k_B x_B(\theta_1) - \tfrac{1}{2}\theta_1 x_B^2(\theta_1) - U_B(\theta_1)\right) \\ & + \lambda_2 \left(\tfrac{1}{2} - \tfrac{U_A(\theta_2)-U_B(\theta_2)}{2\sigma}\right)\left(k_B x_B(\theta_2) - \tfrac{1}{2}\theta_2 x_B^2(\theta_2) - U_B(\theta_2)\right) \end{aligned} \tag{P_i}$$

Note that $U_{-i}(\theta_j)$, which enters the expression of the marginal worker $\widehat{\gamma}_j$, $j = 1, 2$, is taken as given by the two firms. Because the worker's type θ_j is observable and, $\widehat{\gamma}_j$ only depends on $U_i(\theta_j)$ and $U_{-i}(\theta_j)$ (and not on $U_i(\theta_{-j})$ and $U_{-i}(\theta_{-j})$), firms indeed maximize profits per-worker's for each *type* and Program P_i can be decomposed into two programs.[10] Firms compete to attract high-types and respectively solve:

$$\begin{aligned} \max_{\{x_A(\theta_1),U_A(\theta_1)\}} \pi_A(\theta_1) = & \left(\tfrac{1}{2} + \tfrac{U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)}\right)\left(k_A x_A(\theta_1) - \tfrac{1}{2}\theta_1 x_A^2(\theta_1) - U_A(\theta_1)\right) \\ \max_{\{x_B(\theta_1),U_B(\theta_1)\}} \pi_B(\theta_1) = & \left(\tfrac{1}{2} - \tfrac{U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)}\right)\left(k_B x_B(\theta_1) - \tfrac{1}{2}\theta_1 x_B^2(\theta_1) - U_B(\theta_1)\right) \end{aligned} \tag{$P1_i$}$$

Simultaneously, firms compete to attract low-types and respectively solve:

$$\begin{aligned} \max_{\{x_A(\theta_2),U_A(\theta_2)\}} \pi_A(\theta_2) = & \left(\tfrac{1}{2} + \tfrac{U_A(\theta_1)-U_B(\theta_1)}{2\sigma}\right)\left(k_A x_A(\theta_2) - \tfrac{1}{2}\theta_2 x_A^2(\theta_1) - U_A(\theta_2)\right) \\ \max_{\{x_B(\theta_2),U_B(\theta_2)\}} \pi_B(\theta_2) = & \left(\tfrac{1}{2} - \tfrac{U_A(\theta_1)-U_B(\theta_1)}{2\sigma}\right)\left(k_B x_B(\theta_2) - \tfrac{1}{2}\theta_2 x_B^2(\theta_2) - U_B(\theta_2)\right) \end{aligned} \tag{$P2_i$}$$

One can easily check that the second order conditions with respect to $U_i(\theta_1)$, $i = A, B$, require:

$$\alpha_1 < \sigma, \tag{Equ. SOC}$$

which is the same condition to possibly have an interior solution for high-types' marginal worker.

The workers' types being observable, firms are able to require the efficient effort level from each worker:

$$x_i^*(\theta_j) = x_i^{fb}(\theta_j) = \frac{k_i}{\theta_j}. \tag{13}$$

These effort levels ensure that the surplus per-worker, $S_i(\theta_j) \equiv k_i x_i(\theta_j) - \frac{1}{2}\theta_j x_i^2(\theta_j)$, is maximized. Intuitively, the best that each firm can do is to maximize the surplus per-worker and then use a fraction of the surplus to attract the workers.

Let us substitute efforts in firms' Programs $P1_i$ and $P2_i$ by their first best levels (13) and then derive firm i's profits with respect to $U_i(\theta_j)$, $j = 1, 2$, by taking $U_{-i}(\theta_j)$ as given. One obtains two reaction functions for each firm in which indirect utility $U_i(\theta_j)$ offered by firm i is a function of $U_{-i}(\theta_j)$ offered by the rival firm. For high-types, reaction functions are

$$U_A(\theta_1) = \frac{k_A^2 - 2\theta_1(\sigma - \alpha_1)}{4\theta_1} + \frac{1}{2}U_B(\theta_1), \tag{14}$$

$$U_B(\theta_1) = \frac{k_B^2 - 2\theta_1(\sigma - \alpha_1)}{4\theta_1} + \frac{1}{2}U_A(\theta_1), \tag{15}$$

[10] See Appendix A.9 for a richer model where marginal types $\widehat{\gamma}_j$, $j = 1, 2$, depend on $U_i(\theta_j)$, $U_{-i}(\theta_j)$, $U_i(\theta_{-j})$ and $U_{-i}(\theta_{-j})$), i.e. on indirect utilities of both types. Despite full information on ability, here firms maximize *expected* profits instead of profits per-workers's type.

and for type θ_2

$$U_A(\theta_2) = \frac{k_A^2 - 2\theta_2\sigma}{4\theta_2} + \frac{1}{2}U_B(\theta_2), \tag{16}$$

$$U_B(\theta_2) = \frac{k_B^2 - 2\theta_2\sigma}{4\theta_2} + \frac{1}{2}U_A(\theta_2). \tag{17}$$

The two pairs of expressions (14)-(15) and (16)-(17) show that indirect utilities $U_i(\theta_j)$, $i = A, B$, $j = 1, 2$, are strategic complements.

Then we solve the two systems of two reaction functions in two unknowns and obtain the four indirect utilities in the equilibrium with full information on ability:

$$U_A^*(\theta_1) = \frac{2k_A^2 + k_B^2}{6\theta_1} - \sigma + \alpha_1, \tag{18}$$

$$U_B^*(\theta_1) = \frac{k_A^2 + 2k_B^2}{6\theta_1} - \sigma + \alpha_1, \tag{19}$$

$$U_A^*(\theta_2) = \frac{2k_A^2 + k_B^2}{6\theta_2} - \sigma, \tag{20}$$

$$U_B^*(\theta_2) = \frac{k_A^2 + 2k_B^2}{6\theta_2} - \sigma. \tag{21}$$

Hence, $k_A > k_B$ implies that $U_A^*(\theta_1) > U_B^*(\theta_1)$ and $U_A^*(\theta_2) > U_B^*(\theta_2)$; if the firms are identical, indirect utilities are the same and the equilibrium is symmetric. From (18)-(19) one can also check that CfCA benefits high-type workers who receive a larger $U_i^*(\theta_1)$ when $\alpha_1 > 0$ than when $\alpha_1 = 0$. Specifically, the indirect utility of all high-ability workers increases by the amount α_1.

Substituting $U_i^*(\theta_i)$, $i = A, B$, $j = 1, 2$, in (7) and (8) one obtains the expressions for marginal workers:

$$\widehat{\gamma}_1^* = \frac{1}{2} + \frac{k_A^2 - k_B^2}{12\theta_1(\sigma - \alpha_1)} \geq \frac{1}{2}, \tag{22}$$

$$\widehat{\gamma}_2^* = \frac{1}{2} + \frac{k_A^2 - k_B^2}{12\theta_2\sigma} \geq \frac{1}{2}, \tag{23}$$

where $\widehat{\gamma}_1^*$ is increasing in α_1 and the firm with a competitive advantage hires a larger share of both high and low-type workers. Interior solutions require

$$\widehat{\gamma}_1^* < 1 \Leftrightarrow k_A^2 - k_B^2 < 6\theta_1(\sigma - \alpha_1), \tag{24}$$

$$\widehat{\gamma}_2^* < 1 \Leftrightarrow k_A^2 - k_B^2 < 6\theta_2\sigma, \tag{25}$$

where inequality (24) implies (25) because $6\theta_1(\sigma - \alpha_1) < 6\theta_2\sigma$. Intuitively, firm B remains active in the market only if firm A's competitive advantage is not too high relatively to workers' mismatch disutility. The condition for an interior $\widehat{\gamma}_1$ can also be written as

$$\widehat{\gamma}_1^* < 1 \quad \Leftrightarrow \quad \alpha_1 < \alpha_1' \equiv \sigma - \tfrac{k_A^2 - k_B^2}{6\theta_1}. \tag{26}$$

This implies that, when $k_A > k_B$ but firms' heterogeneity is not too high, starting from a value of α_1 close to zero and letting α_1 grow larger, an interior solution where $\widehat{\gamma}_1^* < 1$ first exists. Then $\widehat{\gamma}_1^*$ increases with α_1 and hits the corner solution $\widehat{\gamma}_1^* = 1$ for $\alpha_1 \geq \alpha_1'$.

Finally, we have

$$\widehat{\gamma}_1^* - \widehat{\gamma}_2^* = \frac{\left(k_A^2 - k_B^2\right)\left(2\alpha_1\theta_1 + \sigma(\theta_2 - \theta_1)\right)}{12\theta_1\theta_2\sigma(\sigma - \alpha_1)} \geq 0, \tag{27}$$

so that firm A, holding a competitive advantage, not only hires a larger share of both workers' types, but also employes a workforce characterized by a larger average ability.

Total utilities $U_i(\theta_j)$ of marginal workers in equilibrium are given by:

$$\mathcal{U}_A^*(\theta_1, \widehat{\gamma}_1^*) = \mathcal{U}_B^*(\theta_1, \widehat{\gamma}_1^*) = \frac{k_A^2 + k_B^2 - 6\theta_1\sigma}{4\theta_1} + \alpha_1, \tag{28}$$

$$\mathcal{U}_A^*(\theta_2, \widehat{\gamma}_2^*) = \mathcal{U}_B^*(\theta_2, \widehat{\gamma}_2^*) = \frac{k_A^2 + k_B^2 - 6\theta_2\sigma}{4\theta_2}. \tag{29}$$

Total utilities are increasing moving from the marginal workers to the workers located at the two extremes of the Hotelling line. This is because taste for firms, γ, is not observable so that all the workers different from the marginal ones obtain an additional rent. Hence, once the participation constraints of the two marginal workers are met, all the other workers necessarily receive a strictly positive payoff. Inspection of (28) and (29) confirms that high-ability workers' payoff is increasing in the concern for coworkers' quality.

Note that having $U_A^*\left(\theta_j, \widehat{\gamma}_j^*\right) = U_B^*\left(\theta_j, \widehat{\gamma}_j^*\right) \geq 0$, $j = 1, 2$, not only ensures that all workers receive a positive payoff so that their participation constraint is satisfied, but also that the market is fully covered. Using (28) and (29) we conclude that full market coverage requires:

$$\sigma < \min\{\frac{k_A^2 + k_B^2 + 4\alpha_1\theta_1}{6\theta_1}, \frac{k_A^2 + k_B^2}{6\theta_2}\}.$$

Thus, the condition for a fully covered market requires a σ sufficiently lower than $k_A^2 + k_B^2$, while, from (24) and (25), the condition for an interior solution requires a σ sufficiently larger than $k_A^2 - k_B^2$.[11]

Let us now consider profits in equilibrium. By plugging expressions for effort levels $x_i^*(\theta_j)$ and indirect utilities $U_i^*(\theta_j)$, $i = A, B$, $j = 1, 2$, into $(P1)$ one can check that firm B earns

[11] Putting all conditions together, market is fully covered and the solution is interior for workers of type θ_2 ($\widehat{\gamma}_2^* < 1$) if:

$$\frac{k_A^2 - k_B^2}{6\theta_2} < \sigma < \min\{\frac{k_A^2 + k_B^2 + 4\alpha_1\theta_1}{6\theta_1}, \frac{k_A^2 + k_B^2}{6\theta_2}\}.$$

Market is fully covered and the solution is interior for workers of both type θ_1 and θ_2 $\left(\widehat{\gamma}_j^* < 1 \text{ for } j = 1, 2\right)$ if:

$$\frac{k_A^2 - k_B^2 + 6\alpha_1\theta_1}{6\theta_1} < \sigma < \min\{\frac{k_A^2 + k_B^2 + 4\alpha_1\theta_1}{6\theta_1}, \frac{k_A^2 + k_B^2}{6\theta_2}\}.$$

positive profits and that firm A earns higher profits than $B : \pi_A^* > \pi_B^* > 0$. Interestingly, the derivative of the two firms' profits with respect to α_1 writes:

$$\frac{\partial \pi_A^*}{\partial \alpha_1} = \frac{\partial \pi_B^*}{\partial \alpha_1} = \frac{\lambda_1}{72}\left(\frac{\left(k_A^2 - k_B^2\right)^2}{\theta_1^2\left(\sigma - \alpha_1\right)^2} - 36\right),$$

which is negative under condition (24). In words: when an interior solution exists and both firms hire a positive share of high-ability workers, CfCA decreases firms' profits.

Results so far are summarized in the following proposition.

Proposition 1 ***Full information on ability.*** *(i) When ability is observable while mismatch disutility is the workers' private information, equilibrium contracts are the Nash equilibrium contracts* $\{x_i^*(\theta_j), U_i^*(\theta_j)\}_{j=1,2;\ i=A,B}$ *of the game in which firms compete in the utility space and are defined by* efficient *efforts (13) and by indirect utilities (18)-(21).*
(ii) When firms are identical $(k_A = k_B)$ *they equally share the workforce of both types:* $\widehat{\gamma}_1^* = \widehat{\gamma}_2^* = \frac{1}{2}$ *and* $E_A^*(\theta) = E_B^*(\theta)$.
(iii) When $k_A > k_B$, *then firm A hires a larger share of both types and the better workforce:* $\widehat{\gamma}_1^* > \widehat{\gamma}_2^* > \frac{1}{2}$ *and* $E_A^*(\theta) < E_B^*(\theta)$.
(iv) The share of high-types $\widehat{\gamma}_1^*$ *hired by firm A increases with* α_1 *and* $\widehat{\gamma}_1^* < 1$ *holds for* $k_A^2 - k_B^2 < 6\theta_1(\sigma - \alpha_1)$ *and* $\alpha_1 < \alpha_1'$, *where* α_1' *is expressed in (26). If one of the previous two conditions does not hold, then* $\widehat{\gamma}_1^* = 1$. *The share of low-types* $\widehat{\gamma}_2^*$ *hired by firm A is independent of* α_1 *and* $\widehat{\gamma}_2^* < 1$ *holds for* $k_A^2 - k_B^2 < 6\theta_2\sigma$. *If the opposite inequality holds, then* $\widehat{\gamma}_2^* = 1$.
(v) The concern for coworkers' quality benefits high-ability *workers (including the ones hired by firm B) but is detrimental to firms.*

Let us consider point *(v)* of the above proposition. From expressions (18) and (19) we observe that high-type's indirect utility is increasing in α_1. Intuitively, high-type workers hired by firm B must be compensated for the utility loss suffered because they belong to the workforce with the relatively lower average ability. But, given that indirect utilities are strategic complements, workers employed by firm A also have to be compensated accordingly. By contrast, CfCA is detrimental to firms. Intuitively, α_1 decreases the mismatch disutility of high-type workers and thus increases competition. As a result, the intercepts of the two reaction functions move in opposite directions and reaction functions cross each other farther away from the origin (see equations (14) and (15)): firms offer higher indirect utilities to the workers when $\alpha_1 > 0$ than when $\alpha_1 = 0$. Note that low-type workers are not affected by CfCA.

In Appendix A.9, we solve again for market equilibrium using the richer specification for CfCA. By doing so we show that our reduced model, together with being tractable, is able to capture the main results on market equilibrium and on workers' sorting obtained with the richer specification.

In the following section we study the optimal allocation which maximizes a social welfare function and compare it to the equilibrium.

4 Welfare analysis

To assess how CfCA affects surplus and whether the market equilibrium under full information on ability yields efficient workers' sorting, one has to compare the equilibrium allocation with the one that maximizes total surplus.

4.1 The efficient allocation

We assume an utilitarian social welfare defined as the sum of the firms' profits and workers' utility which includes the concern for coworkers' quality

$$\max_{\{x_A(\theta_j),U_A(\theta_j),\hat{\gamma}_j\}_{j=1,2}} SW = \begin{array}{l} \sum_{j=1}^{2}\int_0^{\widehat{\gamma}_j}\left[\pi_A\left(\theta_j\right)+\mathcal{U}_A\left(\theta_j\right)\right]dF\left(\gamma\right) \\ +\sum_{j=1}^{2}\int_{\widehat{\gamma}_j}^{1}\left[\pi_B\left(\theta_j\right)+\mathcal{U}_B\left(\theta_j\right)\right]dF\left(\gamma\right) \end{array}, \qquad (P_S W)$$

where profits $\pi_i\left(\theta_j\right)$ are defined in (12) and workers' total utilities $\mathcal{U}_A\left(\theta_j\right)$ and $\mathcal{U}_B\left(\theta_j\right)$ are expressed in (5) and (6). Effort levels $\{x_A\left(\theta_j\right)\}_{j=1,2}$ are the efficient ones (see 13).

In Appendix A.1 we show that program P_{SW} can be rewritten as:

$$\max_{\{\hat{\gamma}_j\}_{j=1,2}} SW = \frac{1}{2\theta_1\theta_2}\left[\left(k_A^2-k_B^2\right)\left(\widehat{\gamma}_1\lambda_1\theta_2+\widehat{\gamma}_2\lambda_2\theta_1\right)+k_B^2\left(\lambda_1\theta_2+\lambda_2\theta_1\right)\right]+ \qquad (30)$$

$$-\frac{1}{2}\sigma\left[\lambda_1\hat{\gamma}_1^2+\lambda_2\hat{\gamma}_2^2+\lambda_1\left(1-\hat{\gamma}_1\right)^2+\lambda_2\left(1-\hat{\gamma}_2\right)^2\right] \qquad (31)$$

$$+2\lambda_1\left(\hat{\gamma}_1-\tfrac{1}{2}\right)^2\alpha_1. \qquad (32)$$

In words, welfare depends on the surplus produced by the specific matching of firms and workers, on the mismatch disutility paid by workers and on the (net) premium/loss received by high-type workers because of their CfCA.

Specifically, the first line (30) of SW shows how surplus is affected by the two firms' marginal productivity. When firm A has a competitive advantage, it hires a relatively larger workforce which increases social welfare because this increases the benefit from the matching between firms and workers. The second line (31) shows the total mismatch disutility. Finally, line (32) indicates the total premium from coworkers' quality accruing high-types employed in A (given by $\lambda_1\hat{\gamma}_1\alpha_1\left(\hat{\gamma}_1-\frac{1}{2}\right)$) net of the disutility experienced by workers hired by firm B (given by $-\lambda_1(1-\hat{\gamma}_1)\alpha_1\left(\hat{\gamma}_1-\frac{1}{2}\right)$). Note that the last term is the unique one which depends on α_1 and that SW is monotonically increasing in α_1 provided that $\hat{\gamma}_1\neq\frac{1}{2}$: CfCA increases social surplus. One can easily check that the derivative of $\max_{\{\hat{\gamma}_j\}_{j=1,2}} SW$ with respect to α_1 is positive, confirming that, when concern for coworkers matters, the efficient sorting increases surplus.

In a symmetric allocation with $\widehat{\gamma}_1 = 1/2$, the surplus generated by the premium for coworkers' ability vanishes while the total mismatch disutility is minimized at $-\sigma/4$. An asymmetric allocation is optimal when $k_A > k_B$ because it creates both an additional surplus from the matching of firm A and high-types and a net premium from coworkers' quality which together are larger than the additional mismatch disutility.

The second order condition with respect to $\widehat{\gamma}_1$ requires that

$$\alpha_1 < \frac{1}{2}\sigma, \qquad \text{(First-best SOC)}$$

which is more stringent than the SOC of the firms' program requiring $\alpha_1 < \sigma$. To be able to compare the market allocation with the efficient one we assume from now on that First-best SOC holds.

Efficient sorting entails:

$$\widehat{\gamma}_1^{fb} = \frac{1}{2} + \frac{k_A^2 - k_B^2}{4\theta_1(\sigma - 2\alpha_1)} \geq \frac{1}{2} \tag{33}$$

$$\widehat{\gamma}_2^{fb} = \frac{1}{2} + \frac{k_A^2 - k_B^2}{4\theta_2\sigma} \geq \frac{1}{2} \tag{34}$$

$$\widehat{\gamma}_1^{fb} - \widehat{\gamma}_2^{fb} = \frac{\left(k_A^2 - k_B^2\right)\left(2\alpha_1\theta_1 + \sigma(\theta_2 - \theta_1)\right)}{4\theta_1\theta_2\sigma(\sigma - 2\alpha_1)} \geq 0. \tag{35}$$

Confirming that, when $k_A > k_B$, it is efficient that firm A hires a larger share of workers of each type. We also observe that $\widehat{\gamma}_1^{fb}$ is monotonically increasing in α_1.

An efficient interior allocation requires:

$$\begin{aligned} \widehat{\gamma}_1^{fb} < 1 &\iff k_A^2 - k_B^2 < 2\theta_1(\sigma - 2\alpha_1) \\ \widehat{\gamma}_2^{fb} < 1 &\iff k_A^2 - k_B^2 < 2\theta_2\sigma. \end{aligned} \tag{36}$$

showing that firms' heterogeneity must be sufficiently low. Moreover, if we have an interior solution for $\widehat{\gamma}_1$, we also have one for $\widehat{\gamma}_2$. The interior condition for $\widehat{\gamma}_1^{fb}$ can also be written as

$$\widehat{\gamma}_1^{fb} < 1 \iff \alpha_1 < \alpha_1'' \equiv \tfrac{1}{2}\sigma - \tfrac{k_A^2 - k_B^2}{4\theta_1}. \tag{37}$$

Like in the market equilibrium (see expressions (24) and (25)), when $k_A > k_B$ but firms' heterogeneity is not too high, starting from a value of α_1 close to zero and letting α_1 increase, an interior solution where $\widehat{\gamma}_1^{fb} < 1$ emerges first. Then $\widehat{\gamma}_1^{fb}$ increases with α_1 and hits the corner solution $\widehat{\gamma}_1^{fb} = 1$ for $\alpha_1 \geq \alpha_1''$. Thresholds levels are however different than in the market equilibrium, as we explain below.

The following proposition summarizes results on the efficient matching of workers and firms.

Proposition 2 ***Efficient sorting.*** *(i) When firms are identical ($k_A = k_B$) they equally share the workforce of both types: $\widehat{\gamma}_1^{fb} = \widehat{\gamma}_2^{fb} = 1/2$ and $E_A^{fb}(\theta) = E_B^{fb}(\theta)$.*

(ii) The concern for coworkers ability increases total surplus.
(iii) When $k_A > k_B$ then firm A hires a larger share of both types and a workforce characterized by larger average ability: $\widehat{\gamma}_1^{fb} > \widehat{\gamma}_2^{fb} > 1/2$ and $E_A^{fb}(\theta) < E_B^{fb}(\theta)$.
(iv) The optimal share of high-types $\widehat{\gamma}_1^{fb}$ is increasing in α_1 and is interior ($\widehat{\gamma}_1^{fb} < 1$) if $k_A^2 - k_B^2 < 2\theta_1(\sigma - 2\alpha_1)$ and $\alpha_1 < \alpha_1''$, where α_1'' is expressed in (37). If one of the previous two conditions does not hold, then $\widehat{\gamma}_1^{fb} = 1$. The solution for low-types is interior ($\widehat{\gamma}_2^{fb} < 1$) if $k_A^2 - k_B^2 < 2\theta_2\sigma$. If the opposite inequality holds, then $\widehat{\gamma}_2^{fb} = 1$.

To understand the economic forces generating workers' sorting when $k_A > k_B$, let us start with low-ability workers. The additional mismatch disutility arising when $\widehat{\gamma}_2$ moves on the right of 1/2 is traded off with having a larger share of workers employed by the relatively more productive firm A. A similar reasoning applies for high-type workers who, being relatively more productive, benefit even more from the good matching with the more efficient firm so that $\widehat{\gamma}_1^{fb}|_{\alpha_1=0} > \widehat{\gamma}_2^{fb}$ (compare expression (33) when $\alpha_1 = 0$ with (34)). But now CfCA becomes also relevant. Specifically, a second benefit from moving $\widehat{\gamma}_1$ on the right arises from the larger share of high-type workers employed by firm A who enjoy the premium from coworkers' ability. A third one arises because, as a result, there are fewer high-type workers employed by firm B who suffer the disutility from coworkers' lower-than-average ability.

The following chain of inequalities holds: $\widehat{\gamma}_1^{fb} > \widehat{\gamma}_1^{fb}|_{\alpha_1=0} > \widehat{\gamma}_2^{fb}$ and the higher α_1, the higher the benefit from moving $\widehat{\gamma}_1$ to the right of 1/2. As a consequence the difference between marginal types $\left(\widehat{\gamma}_1^{fb} - \widehat{\gamma}_2^{fb}\right)$ and the average ability of the workforce in firm A both increase with α_1. Since, in the market allocation, efforts are set at the efficient level, the unique possible distortion is in workers' sorting to firms.

4.2 Is market equilibrium efficient?

Recall that, in the market equilibrium, each firm determines the indirect utilities to be offered to its workers by maximizing its profits while taking the indirect utility offered by the rival firm as given. Marginal workers are then determined indirectly by substituting the equilibrium indirect utilities (18)-(21) into (7) and (8). In the first best, instead, marginal workers are such that the sum of firms' profits and workers' utilities is maximized. Recall that efforts are set at their efficient levels in equilibrium.

Let us compare (22)-(23) and (33)-(34). From Propositions 1 and 2 it follows:

Corollary 1 *(i) When firms are identical ($k_A = k_B$) the concern for coworkers' ability does not affect surplus and the market allocation is fully efficient.*
(ii) When firms are heterogeneous ($k_A > k_B$), the concern for coworkers' ability increases total surplus but reduces firms' profits. Specifically:

(iia) When condition (24) holds so that an interior solution emerges for both marginal workers, market sorting is inefficient because the share of high- and low-ability workers employed by firm A is too low. In addition, the average ability characterizing the workforce hired by firm A is too low and the one of firm B is too high ($E^*_A(\theta) > E^{fb}_A(\theta)$ and $E^*_B(\theta) < E^{fb}_B(\theta)$).
(iib) In the market equilibrium, an interior solution emerges too often.

Interestingly, only high-ability workers appropriate the surplus generated by CfCA when firms are heterogenous. When $\alpha_1 > 0$, firms get a lower share of a larger surplus and are worse off. Strategic interaction prevents even the more efficient firm A from appropriating a part of the increased return from the matching between high-ability workers and the more efficient firm. We further elaborate on that in Proposition 3.

As expressed in part *(iia)* of the corollary, too few high-ability and too few low-ability workers are employed by firm A in equilibrium ($\widehat{\gamma}^*_j < \widehat{\gamma}^{fb}_j$, $j = 1, 2$). This result in itself is not sufficient to compare the average ability of workers hired by the two firms. However, one can easily check that $\widehat{\gamma}^*_1 - \widehat{\gamma}^*_2 < \widehat{\gamma}^{fb}_1 - \widehat{\gamma}^{fb}_2$ so that that average ability in firm A is inefficiently low; see expressions (27) and (35). Finally, part *(iib)* of the corollary is explained by the fact that threshold values for interior solutions are such that $\alpha'_1 > \alpha''_1$; see (26) and (37). Hence, as α_1 increases, a corner solution is reached faster in the efficient allocation than in the market equilibrium. In other words, the region of the parameters such that $\widehat{\gamma}^*_1 < 1$ is too large.

What is the source of the inefficient sorting observed at the market equilibrium? Is it a consequence of strategic interaction between the two firms, a result of profit maximization, or both? To address these questions we study the multi-firm monopsonist's solution in Appendix A.2. We show that sorting obtained by a monopsonist is inefficient, but to a lesser extent than sorting in the market allocation:

$$\widehat{\gamma}^*_1 < \widehat{\gamma}^M_1 < \widehat{\gamma}^{fb}_1, \tag{38}$$

$$\widehat{\gamma}^*_2 < \widehat{\gamma}^M_2 < \widehat{\gamma}^{fb}_2. \tag{39}$$

Consequently, we conclude that strategic interaction and profit maximization jointly contribute to the distortion in workers' sorting. The following proposition provides further details:

Proposition 3 *Market sorting is inefficient for three reasons that sum up and all contribute to the downward distortion of $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$. (i)* ***Profit maximization*** *implies that* average *mismatch disutility is disregarded and only the* highest *mismatch disutility, i.e. the one of marginal workers, is considered. (ii)* ***Strategic interaction*** *acts through two channels. (iia)* Indirect utilities *are set in such a way that the more efficient firm A accommodates too much while the less efficient firm B competes too aggressively. (iib) In the Nash equilibrium, the* externality *generated by the concern for coworkers' ability is only partially internalized by the two firms.*

This specific channel due to strategic interaction pushes towards a lower $\widehat{\gamma}_1$ *but does not affect* $\widehat{\gamma}_2$.

The effect of profit maximization, accounted for in part (i) of the proposition is relevant both for the monopsonist and for the two competing firms. Basically, when maximizing profits, firms focus on the two marginal workers and on their specific mismatch disutility while disregarding the average mismatch disutility of the whole workforce (the latter corresponds to the term (31) in the expression of the social welfare function). By so doing the monopsonist and the competing firms weight the mismatch disutility of the marginal workers too much and, as a result, marginal workers are too close to 1/2. Profit maximization explains why the inequalities $\widehat{\gamma}_1^M < \widehat{\gamma}_1^{fb}$ and $\widehat{\gamma}_2^M < \widehat{\gamma}_2^{fb}$ in (38) and (39) hold.

Let us move to part (ii) of the proposition and, for the sake of exposition, consider the case $\alpha_1 = 0$. First of all recall that, under full information on ability, efforts are set at the efficient levels and thus here competition *does not* increase allocative efficiency. In other words, competition only generates a distortion in sorting due to strategic interaction (that sum up to the inefficiency due to profit maximization) and explains the inequalities $\widehat{\gamma}_1^*|_{\alpha_1=0} < \widehat{\gamma}_1^M|_{\alpha_1=0}$ and $\widehat{\gamma}_2^* < \widehat{\gamma}_2^M$, the latter appearing in (39). By taking the indirect utilities offered by the competing firm as given, firm A ends up being too accommodating (firm A does not pay workers enough) while firm B is too aggressive (firm B pays workers too much) so that too many workers are employed in the less efficient firm B.

Finally, part (iib) of the proposition indicates that, when CfCA matters ($\alpha_1 > 0$), we observe an additional source of distortion in sorting of high-ability workers due to strategic interaction and which operates thought the externality introduced by workers' peer effects. Specifically, firm A disregards the utility loss suffered by high-type workers employed by firm B while firm B disregards the premium accruing high-types hired by firm A. This further reduces $\widehat{\gamma}_1$.

5 Equilibrium contracts when neither taste for firms nor ability are observable

We now assume that workers' abilities are no longer observable. The objective of firms A, B continues to be represented by Program P_i. However, unlike in the previous sections, each firm has now to consider total profits, and not just profits by type $P1_i$ and $P2_i$. Most importantly, firms now take into account the workers' incentive compatibility constraints. Provided that both firms are able to hire workers with both ability levels, there are two incentive compatibility constraints for each firm: the *downward incentive constraint* (henceforth DIC) requiring that high-ability types are not attracted by the contract offered to low-ability types and the *upward incentive constraint* (henceforth UIC) requiring that low-ability types do not gain by mimicking high-ability workers. For each firm $i = A, B$, these constraints (written in terms of effort levels and utilities) are given by[12]

$$U_i(\theta_1) \geq U_i(\theta_2) + \frac{1}{2}(\theta_2 - \theta_1) x_i^2(\theta_2), \qquad (DIC_i)$$

and

$$U_i(\theta_2) \geq U_i(\theta_1) - \frac{1}{2}(\theta_2 - \theta_1) x_i^2(\theta_1). \qquad (UIC_i)$$

These constraints depend neither on mismatch disutility γ nor on the marginal worker $\widehat{\gamma}_j$, $j = 1, 2$. Combining DIC_i and UIC_i yields

$$\frac{1}{2}(\theta_2 - \theta_1) x_i^2(\theta_2) \leq U_i(\theta_1) - U_i(\theta_2) \leq \frac{1}{2}(\theta_2 - \theta_1) x_i^2(\theta_1), \qquad (40)$$

which shows that incentive compatible contracts must satisfy: (i) the monotonicity condition $x_i(\theta_1) \geq x_i(\theta_2)$, requiring that high-ability workers exert more effort than low-ability types at each firm $i = A, B$; and (ii) condition $U_i(\theta_1) \geq U_i(\theta_2)$, requiring that high-ability workers get an indirect utility not lower than the one of low-ability types, for each employer $i = A, B$.

In Lemma 4 (see Appendix A.5), among other results, we show that the two constraints cannot be binding simultaneously when $\lambda_2 \leq \lambda_1$. This suggests that, as it is generally the case in this type of models (see also Bénabou and Tirole 2016), only one or the other incentive constraint will typically bind at a given point, which we now assume.

In addition, as under full information, the participation constraints PC must be met.

To sum up, firms simultaneously design menus of contracts of the form $\{x_i(\theta_j), U_i(\theta_j)\}_{i=A,B;\, j=1,2}$ by maximizing Program P_i with respect to the effort level and the indirect utility associated to each type θ worker, taking as given the indirect utility $U_{-i}(\theta)$ that the rival firm leaves to the worker, and subject to the two incentive compatibility constraints DIC_i and UIC_i and to the participation constraints PC. Once optimal screening contracts $\{x_i(\theta_j), U_i(\theta_j)\}_{i=A,B;\, j=1,2}$ are derived, workers compute the corresponding non-linear transfer schedule $w_i(x_i)$ for $i = A, B$, select the preferred one and thus choose which firm to work for.

In what follows we assume that screening continues to entail $\widehat{\gamma}_1 \geq \widehat{\gamma}_2$ (as under full information on ability). Hence, the firm with a competitive advantage hires a share of high-types larger than the share of low-types and employes a workforce characterized by a larger average ability.[13] In other words, we assume that distortions introduced by asymmetric information do not affect indirect utilities so much to change the nature of workers' sorting.

We first study under which conditions, if any, the full information equilibrium is incentive compatible so that it remains the solution when types are not observable. Then we turn to the

[12] DIC_i is obtained by considering $U_i(\theta_1) \geq w_i(\theta_2) - \frac{1}{2}\theta_1 x_i^2(\theta_2)$, where the r.h.s. of the previous inequality is equal to $w_i(\theta_2) - \frac{1}{2}\theta_2 x_i^2(\theta_2) + \frac{1}{2}\theta_2 x_i^2(\theta_2) - \frac{1}{2}\theta_1 x_i^2(\theta_2) = U_i(\theta_2) + \frac{1}{2}(\theta_2 - \theta_1) x_i^2(\theta_2)$.

[13] Specifically, this assumption will be relevant for the proof of Lemma 4; see Appendix A.5. An equivalent assumption is used in Barigozzi and Burani (2019).

case where at least one firm has a binding incentive constraint and study the different regimes that can occur.

5.1 Neither DIC_i nor UIC_i are binding

We first check conditions, if they exist, such that the market equilibrium obtained when ability is observable is incentive compatible. Consider contracts $\{x_i^*(\theta_j), U_i^*(\theta_j)\}$, $i = A, B$, $j = 1, 2$, where $x_i^*(\theta_j) = x_i^{fb}(\theta_j) = k_i/\theta_j$ and $U_i^*(\theta_j)$ are described in (18)-(21) and substitute them into DIC_i and UIC_i, $i = A, B$. One immediately observes that UIC_B is always met (see Appendix A.3 for more details). Rearranging the other incentive constraints one finds that they are met if the following conditions hold:

$$UIC_B \text{ holds if} : \alpha_1 \leq \frac{\theta_2 - \theta_1}{6\theta_2\theta_1^2}\left[3k_B^2(\theta_2 - \theta_1) - \theta_1(k_A^2 - k_B^2)\right] \equiv \alpha_1^a, \tag{41}$$

$$DIC_A \text{ holds if} : \alpha_1 \geq \frac{\theta_2 - \theta_1}{6\theta_1\theta_2^2}\left[\theta_2(k_A^2 - k_B^2) - 3k_A^2(\theta_2 - \theta_1)\right] \equiv \alpha_1^b, \tag{42}$$

$$UIC_A \text{ holds if} : \alpha_1 \leq \frac{\theta_2 - \theta_1}{6\theta_2\theta_1^2}\left[\theta_1(k_A^2 - k_B^2) + 3k_A^2(\theta_2 - \theta_1)\right] \equiv \alpha_1^c. \tag{43}$$

In Appendix A.3 we prove the following result.

Lemma 3 *(i) When $k_A = k_B = k$ the market allocation described in Proposition 1 is incentive compatible if the following condition holds*

$$0 \leq \alpha_1 \leq \frac{(\theta_2 - \theta_1)^2}{2\theta_2\theta_1^2}k^2 = \alpha_1^a = \alpha_1^c. \tag{44}$$

(ii) When $k_A > k_B$ the market allocation described in Proposition 1 is incentive compatible if the following two conditions hold

$$0 \leq \alpha_1 \leq \alpha_1^a; \tag{45}$$

$$\frac{\theta_2 - \theta_1}{\theta_2} \geq \frac{k_A^2 - k_B^2}{3k_B^2}. \tag{46}$$

When $k_A = k_B = k$, DIC_i, $i = A, B$, are always satisfied and, UIC_i, $i = A, B$, are met if Condition (44) is satisfied. The condition shows that, if CfCA is sufficiently small and/or heterogeneity in workers' ability $(\theta_2 - \theta_1)$ sufficiently large, then contracts offered under full information on ability are incentive compatible. Interestingly, if workers do not care for their coworkers' ability (i.e., if $\alpha_1 = 0$), the market allocation is *always* incentive compatible when the two firms are identical.[14] Hence, CfCA reduces the likelihood that the allocation characterized in Proposition 1 is incentive compatible.

[14] In the case where $k_A = k_B = k$ and $\alpha_1 = 0$, the result is reminiscent of Rochet and Stole (2002) who consider identical firms and find that incentive constraints are always slack for all firms, so that efficient quality allocations with cost-plus-fixed-fee pricing emerge at equilibrium. See also Armstrong and Vickers (2001) and Barigozzi and Burani (2019).

When $k_A > k_B$ conditions are more stringent. First, condition (45) continues to require that α_1 must be small. In addition, heterogeneity in workers' ability $(\theta_2 - \theta_1)$ must be relatively larger than heterogeneity in firms' productivity $(k_A - k_B)$; see condition (46).

Conditions (44)–(46) together show that incentive compatibility is more likely to be achieved when workers' heterogeneity is sufficiently large. Indeed, when workers' types are sufficiently different from each other, mimicking is too costly to be attractive. Specifically, Condition (46) indicates that, to have incentive compatible full information contracts, workers' heterogeneity must be sufficiently higher than firms' heterogeneity.

5.2 Screening contracts

We now proceed with the characterization of optimal screening contracts, when conditions (44)–(46) are not met so that full information contracts are no longer incentive compatible. Firms will then design contracts that *are constrained* by incentive compatibility. Which constraints are relevant depends on the parameters' value and different regimes have to be considered. The following analysis holds when each firm is able to hire both high- and low-ability workers, that is when the second chain of inequalities in Footnote 11 is met.

In Remark 1 (see Appendix A.4), we study the ranking of the three threshold values α_1^a, α_1^b and α_1^c defined in (41)-(43) and we identify conditions on α_1, Δk and $\Delta\theta$ such that each incentive constraint starts to be binding when (44)-(46) are not met. Figure 1 below reports the two possible rankings of the threshold values, α_1^a, α_1^b and α_1^c, depending on whether heterogeneity in workers' ability is larger or lower than firms' heterogeneity.

Lemma 4 (see Appendix A.5) complements Remark 1. It studies the two firms' programs P_i, $i = A, B$, when DIC_i and UIC_i are taken into account and only one constraint may bind for each firm. We show that DIC cannot be binding for firm B, whereas for firm A, we show that UIC_A can be binding only if α_1 is sufficiently large.

Combining results from Remark 1 and from Lemma 4 established the following proposition (see also Figure 1).

Proposition 4 *Under competition and screening, when conditions (44)–(46) do not hold, UIC_B is always binding whereas DIC_B is always slack.*
Letting α_1 grow larger and considering the threshold values appearing in (41)-(43), the following three regimes become relevant in turn:

Regime 1 *Both UIC_B and DIC_A are binding for*

$$0 < \alpha_1 \leq \alpha_1^b \quad \textit{and} \quad \frac{\theta_2 - \theta_1}{\theta_1} \leq \frac{k_A^2 - k_B^2}{3k_A^2}.$$

and H/peer effects/R1YCK400.wmf

Figure 1: The different regimes according to the relevance of the concern for coworkers' ability.

Regime 2 *Only UIC_B is binding either for*

$$\frac{\theta_2-\theta_1}{\theta_1} \leq \frac{k_A^2-k_B^2}{3k_A^2} \quad \textit{and} \quad \alpha_1^b < \alpha_1 \leq \alpha_1^c;$$

or for

$$\frac{\theta_2-\theta_1}{\theta_2} \geq \frac{k_A^2-k_B^2}{3k_B^2}, k_A \neq k_B \quad \textit{and} \quad \alpha_1^a < \alpha_1 \leq \alpha_1^c.$$

Regime 3 *Both UIC_A and UIC_B are binding for*

$$\alpha_1 > \alpha_1^c \quad \textit{and} \quad \frac{\sigma}{\sigma-\alpha_1} \geq \frac{\pi_A(\theta_2)}{\pi_A(\theta_1)} > 1.$$

Regime 1 occurs for low values of α_1, because the condition $0 < \alpha_1 \leq \alpha_1^b$ ensures that DIC_A is binding (see Remark 1). Moreover when, $(\theta_2-\theta_1)/\theta_1 \leq (k_A^2-k_B^2)/3k_A^2$, UIC_B is necessarily binding because $\alpha_1^a < 0$. Hence, this regime holds for firms' heterogeneity relatively larger than heterogeneity in workers' ability. This means that Regime 1 never occurs when firms are identical.

When α_1 grows larger Regime 2 becomes relevant. Condition $(\theta_2-\theta_1)/\theta_1 \leq (k_A^2-k_B^2)/3k_A^2$ again implies that UIC_B is binding because $\alpha_1^a < 0$, while $\alpha_1^b < \alpha_1 \leq \alpha_1^c$ ensures that DIC_A and UIC_A are both slack (see Remark 1). When instead $(\theta_2-\theta_1)/\theta_1 > (k_A^2-k_B^2)/3k_A^2$ then DIC_A is always slack because $\alpha_1^b \leq 0$, whereas $\alpha_1^a > 0$ holds so that $\alpha_1 > \alpha_1^a > 0$ implies that UIC_B is binding. The condition $\alpha_1^a < \alpha_1 < \alpha_1^c$ ensures that UIC_B is binding but UIC_A is slack. Note that, when $k_A = k_B$, then $\alpha_1^a \equiv \alpha_1^c$ and this regime disappears.

Finally, when $\alpha_1 > \alpha_1^c$, UIC_B is binding and, provided that condition $\sigma/(\sigma-\alpha_1) \geq \pi_A(\theta_2)/\pi_A(\theta_1) > 1$ is also met, UIC_A is binding as well. The chain of two inequalities is necessary for having UIC binding and DIC slack for firm A (see Remark 4). Absent CfCA ($\alpha_1 = 0$), the chain of two inequalities would not hold, UIC_A could not be binding and this regime would not exist. Interestingly, Regime 3 is the only one that is compatible with the case of identical firms.

The following proposition, established in Appendix A.7, A.6 and A.8, summarizes the main properties of the equilibria achieved in the different regimes. Recall that superscript $*$ denotes the equilibrium when ability is observable (characterized in Section 3); now superscript $**$ indicates the equilibrium under screening.

Proposition 5 ***Equilibrium contracts under screening.***
Optimal contracts $\{x_i^{**}(\theta_j), U_i^{**}(\theta_j)\}_{i=A,B;\ j=1,2}$ are such that:

Regime 1 (*i*) *Firm A sets the efficient effort level for high-ability workers,* $x_A^{**}(\theta_1) = x_A^{fb}(\theta_1)$, *whereas it distorts downward the effort of low-ability workers,* $x_A^{**}(\theta_2) < x_A^{fb}(\theta_2)$. *Firm B sets the efficient effort level for low-ability workers,* $x_B^{**}(\theta_2) = x_B^{fb}(\theta_2)$, *whereas it distorts upward the effort of high-ability workers,* $x_B^{**}(\theta_1) > x_B^{fb}(\theta_1)$; (*ii*) *In firm A,* $U_A^{**}(\theta_1) > U_A^{*}(\theta_1)$ *and* $U_A^{**}(\theta_2) < U_A^{*}(\theta_2)$ *whereas, in firm B,* $U_B^{**}(\theta_1) < U_B^{*}(\theta_1)$ *and* $U_B^{**}(\theta_2) > U_B^{*}(\theta_2)$. *Strategic complementarity between indirect utilities offered to the same workers' type mitigate overall departures from the values of* $U_i^{*}(\theta_j)$, $i = A, B;\ j = 1, 2$, *obtained under full information.*

Regime 2 (*i*) *Firm A sets the efficient effort level for both high and low-ability workers,* $x_A^{**}(\theta_1) = x_A^{fb}(\theta_1)$ *and* $x_A^{**}(\theta_2) = x_A^{fb}(\theta_2)$; *firm B sets the efficient effort level for low-ability workers,* $x_B^{**}(\theta_2) = x_B^{fb}(\theta_2)$, *whereas it distorts upward the effort of high-ability workers,* $x_B^{**}(\theta_1) > x_B^{fb}(\theta_1)$; (*ii*) *High-types' indirect utilities are lower than at the full information equilibrium* ($U_i^{**}(\theta_1) < U_i^{*}(\theta_1)$) *whereas low-types' ones are higher* ($U_i^{**}(\theta_2) > U_i^{*}(\theta_2)$), $i = A, B$.

Regime 3 (*i*) *Firms A and B set the efficient effort level for low-ability workers,* $x_i^{**}(\theta_2) = x_i^{fb}(\theta_2)$, $i = A, B$, *whereas they both distort upward the effort level of high-ability workers,* $x_i^{**}(\theta_1) > x_i^{fb}(\theta_1)$, $i = A, B$. (*ii*) *High-types' marginal utilities are lower than at the full information equilibrium* ($U_i^{**}(\theta_1) < U_i^{*}(\theta_1)$) *whereas low-types' ones are higher* ($U_i^{**}(\theta_2) > U_i^{*}(\theta_2)$), $i = A, B$.

First of all, recall that reaction functions (14)-(17) imply that indirect utilities offered to the same workers' type under full information on ability are strategic complement. Let us start from Regime 2 where only UIC_B is binding. Here, firm B needs to increase $U_B(\theta_2)$ and to decrease $U_B(\theta_1)$ in order to discourage mimicking by low-types. And, given strategic complementarity, firm A changes its indirect utilities accordingly and in the same direction, but the change is lower than the one implemented by firm B. Overall, this will make $\widehat{\gamma}_1$ increase and $\widehat{\gamma}_2$ decrease. Let us now move to Regime 1 where both UIC_B and DIC_A are binding. In Regime 1, firm B still needs to increase $U_B(\theta_2)$ and decrease $U_B(\theta_1)$ as before, but now firm A needs to decrease $U_A(\theta_2)$ and to increase $U_A(\theta_1)$ in order to discourage mimicking by high-types. All changes in $U_i(\theta_1)$ and $U_i(\theta_2)$, $i = A, B$, induce a reactions by the competitor, via strategic complementarity, and $U_{-i}(\theta_1)$ and $U_{-i}(\theta_2)$ will change accordingly. But now the two firms change indirect utilities in opposite directions and those changes partially offset each others. As a consequence $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ will move in the same direction as under Regime 2, but the overall change in $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ will be smaller. This in turn implies that, under Regime 1, strategic complementarity between indirect utilities mitigates overall departures from the values of $U_i^{*}(\theta_j)$ obtained under full information.

Distortions in effort levels are larger in Regime 1 than in Regime 2 because, under the latter, only the effort level of high-types employed by firm B is distorted, while all the other effort levels are efficient. Conversely, changes in the location of marginal workers are larger in Regime 2 as stated in the following corollary:

Corollary 2 ***Sorting under screening***. *Workers' sorting is such that:*

Regime 1 *(i) The share of high-types employed in firm A increases ($\widehat{\gamma}_1^{**} > \widehat{\gamma}_1^*$) whereas the share of low-types employed in firm A ($\widehat{\gamma}_2^{**} < \widehat{\gamma}_2^*$) decreases. (ii) Screening contracts improve average quality of the workforce employed in firm A and impair average quality of the workforce employed in firm B ($\widehat{\gamma}_1^{**} - \widehat{\gamma}_2^{**} > \widehat{\gamma}_1^* - \widehat{\gamma}_2^*$).*

Regime 2 *Sorting is like under Regime 1. Points (i) and (ii) above continue to hold but the changes in marginal types and in the workforce's average quality are larger.*

Regime 3 *(i) Screening contracts do not affect the average quality of the workforce because the share of high and low-types employed by the two firms remains constant: $\widehat{\gamma}_1^{**} = \widehat{\gamma}_1^*$ and $\widehat{\gamma}_2^{**} = \widehat{\gamma}_2^*$. (ii) If $k_A > k_B$ then $\widehat{\gamma}_1^{**} = \widehat{\gamma}_1^* > \widehat{\gamma}_2^{**} = \widehat{\gamma}_2^*$; if $k_A = k_B$ then $\widehat{\gamma}_1^{**} = \widehat{\gamma}_1^* = \widehat{\gamma}_2^{**} = \widehat{\gamma}_2^* = 1/2$.*

Recall that, when $k_A > k_B$, workers' sorting is inefficient in the market allocation with full information on ability because both marginal workers are located too close to 1/2 and because the average ability of workers hired by firm A is too low. By increasing the share of high-types hired by firm A, Regime 1 and 2 decrease distortions in the sorting of high-ability workers. At the same time, by decreasing the share of low-types hired by firm A, they also increase distortions in the sorting of low-ability workers. Given that the two effects together imply $\widehat{\gamma}_1^{**} - \widehat{\gamma}_2^{**} > \widehat{\gamma}_1^* - \widehat{\gamma}_2^*$, the distortion in average ability of the workforce employed by firm A decreases. For the reasons explained below Proposition 5, those effects are stronger in Regime 2 than in Regime 1. Hence we conclude that Regime 2 decreases distortions in workers' sorting more than Regime 1. Given that Regime 2 is also characterized by a lower distortions in effort levels, we can conclude that the allocation obtained under Regime 2 is overall more efficient than the one obtained under Regime 1.

Let us now move to Regime 3. Under this regime, distortions in workers' sorting remain the same as in the market equilibrium under full information on ability. However, indirect utilities and the effort levels of high-ability workers change in such a way that low-types are better off, while both the firms and high-types are worse off. This regime may occur both with identical and heterogeneous firms (i.e. for $k_A \geq k_B$). Looking at the two cases separately, when firms are identical ($k_A = k_B$), workers' sorting is not distorted ($\widehat{\gamma}_j^{**} = \widehat{\gamma}_j^* = \widehat{\gamma}_j^{fb} = 1/2$, $j = 1, 2$) but effort levels of high-types are upward distorted. When instead firms differ ($k_A > k_B$), both workers' sorting and high-types' effort levels are distorted.

To sum up, when ability is not observable and either Regimes 2 or Regime 3 prevails, the distortion in sorting of high-ability workers decreases while the one of low-ability workers increases with respect to market equilibrium under full information. When Lemma 3 holds and under Regime 3, instead, workers' sorting remains the same. Note that CfCA substantially enriches the set of possible solutions under screening. Indeed, when $\alpha_1 = 0$, only two cases may occur: either full information contracts are incentive compatible or Regime 1 emerges (see Figure 1). CfCA makes Regime 2 and 3 possible.

From the point of view of the workforce, private information on ability impairs high-ability workers and benefits low-types both under Regime 2 and under Regime 3. Specifically, the effort exerted by high-types is upward distorted, at least in firm B, so that we observe overincentivization of high-skilled work like in Bénabou and Tirole (2016). In addition, indirect utility of high-types is reduced with respect to the case of private information, hence talented workers are worse off. Conversely, low-types still exert the efficient level of effort and receive a larger indirect utility than under full information, thus they are better off. This welfare comparison is ambiguous in Regime 1 because a different incentive constraint is binding for each firm. A general result in our setting is that, when (45) and (46) do not hold so that the full information solution is not incentive compatible, no matter the prevailing regime, private information on ability leads to an upward distortion of the effort exerted by high-types employed by the least efficient firm B and to a fall of their indirect utility. Hence, we can conclude that CfCA benefits all high-types under full information but that their additional surplus is at least partially eroded when ability is not observable. Returning to our example, this implies that CfCA empowers all senior talented job market applicants, also the ones employed by the least efficient firm, but junior applicants entering the job market for the first time are disadvantaged by their private information and are not able to appropriate the same surplus.

6 Concluding remarks

Consider a Ph.D. candidate receiving an offer from the Department of Economics of both University-X and College-Y. Which offer should the young economist accept? The choice is also likely to depend on the overall quality of the recruitment accomplished by each Department. Indeed the candidates' academic network, his/her future publishing prospects and research funds opportunities all tend to increase with the quality of the faculty and the prestige of the Department.

We consider a model where workers' utility is increasing in the share of high-ability coworkers. Specifically, high-ability workers' utility increases if they are employed by the firm hiring the larger share of high-ability workers, while it decreases in the opposite case. By taking a first step towards analyzing the role played by the concern for coworkers' quality in the hiring process, we contribute to the theory of organizations and to personnel economics. In addition, by studying screening contracts we contribute to the literature on competition and screening when workers's ability is not observable to firms.

We consider two (possibly) heterogeneous firms, located at the two extremes of the Hotelling line, competing to attract workers whose ability can be either high or low and who are uniformly distributed. The location on the Hotelling line represents workers' taste for firm (mismatch disutility) and is always the workers' private information.

Under full information on ability, we show that CfCA expands total surplus, but is detrimental to firms because it increases competition for high-ability workers who appropriate all the additional surplus. Except when firms are identical and hire half of the workforce of each type, workers' sorting to firms is distorted. The distortion in sorting is the results of three different forces, all pushing toward an excess of workers of both types employed by the least efficient firm: profit maximization, strategic interaction and the externality generated by CfCA which is only partially internalized by firms in equilibrium.

When ability is not observable, full information contracts are incentive compatible if CfCA and firms' heterogeneity is sufficiently low and/or workers' heterogeneity is large enough. When full information contracts are not incentive compatible then, depending on which incentive constraints are binding, one of three possible regimes emerges where high-types face countervailing incentives in at least one firm. Consequently, private information on ability erodes at least part of the surplus that high-ability workers obtain via CfCA and the more so the higher the relevance of CfCA. As for sorting, the opposite pattern occurs since when CfCA is low, sorting is less distorted under asymmetric information than in full information; while a high CfCA implies that the distortion in sorting does not change with information structure.

Our paper represents a first step in the study of peer effects in the workplace when they are not related to (positive or negative) spillovers on workers' productivity. We focus on those organizations where top workers bring value to the firm and their employees as research institutions and firms providing professional services. In the model we treated CfCA as a black box and assumed that workers' utility is increasing in the share of top workers employed by the firm. While this is a shortcut that allows us to keep the model tractable, it can be explained by different economics mechanisms. We present two examples.

First, in the case of junior job market candidates whose ability is not observable by firms, coworkers' quality may increase the worker's career prospects outside the firm via a signaling mechanism. Future prospective employers will perceive a junior job market candidate previously employed by the firm hiring the majority of top workers as a worker above average. As a consequence, in equilibrium, the discounted utility from profitable future matching will accrue all junior workers hired by the more prestigious firm.

Second, let us introduce the product side of the market and consider that the two firms also compete to attract consumers characterized by heterogeneous willingness to pay for product's quality. This generates a setting with both competition for talented workers in the labor market and competition with vertical differentiation à la Shaked and Sutton (1982) in the product market. In the case of firms selling professional services, product's quality is likely to increase with the share of high-type workers that one firm is able to hire. In turn, by hiring the larger share of high-type workers, a firm is able to offer a higher quality which translates into higher profits. Hence, in case profits are partially shared with employees, workers' utility increases with the quality of their coworkers because a more qualified workforce produces a better output, which implies higher profits for the firm and a larger payoff for its employees.

The model could be extended in many ways. The more natural one is considering performance and productivity spillovers in the workplace; see for example Groysberg and Lee (2008), Ertug et al. (2018), Tan and Netessine (2019). A positive externality exerted by top workers on the productivity of their colleagues is likely to further increase the ability of the more efficient firm to attract the best talents. Conversely, a negative externality exerted by top workers on their coworkers' productivity will tend to mitigate both the boost in utility of high-ability workers and the attractiveness of the more efficient firm. Career concerns could also be taken into account. A larger share of high-ability colleagues may imply a lower probability of promotions which could partially or totally offset the premium from coworkers' quality.

References

[1] Akerlof, G.A., and Kranton R.E. (2008), "Identity, Supervision, and Work Groups." *American Economic Review*, 98 (2): 212-17.

[2] Armstrong M. and J. Vickers (2001), "Competitive Price Discrimination", *The RAND Journal of Economics,* 32(4), 1-27.

[3] Barigozzi F. and N. Burani (2016), "Competition and Screening with Motivated Health Professionals", *Journal of Health Economics*, 50, 358-371.

[4] Barigozzi F. and N. Burani (2019), "Competition for Talent when Firms' Mission Matters", *GAMES and Economic Behavior* 116, 128-151.

[5] Bénabou R. and J. Tirole (2016), "Bonus Culture: Competitive Pay, Screening, and Multitasking", *Journal of Political Economy,* 124(2), 305-370.

[6] Biglaiser G. and C. Mezzetti (2000), "Incentive Auctions and Information Revelation", *The RAND Journal of Economics* 31(1), 145-164.

[7] Bolte, L., N. Immorlica, and M.O. Jackson (2020), "The Role of Referrals in Immobility, Inequality, and Inefficiency in Labor Markets". Available at SSRN: https://ssrn.com/abstract=3512293

[8] Burke, M.A. and T.R. Sass (2013), "Classroom Peer Effects and Student Achievement", *Journal of labor Economics*, 31(1), 51- 82.

[9] Donze J., T. Gunnes (2018), "Becoming "We" instead of "I", identity management and incentives in the workplace," *Journal of Economic Behavior & Organization*, 148, 105-120.

[10] Ertug, G., Gargiulo, M., Galunic, C.; and Zou T. (2018), "Homophily and Individual Performance." *Organization Science*, 29(5), 912-930.

[11] Groysberg, B. and Lee, L.-E. (2008), The effect of colleague quality on top performance: the case of security analysts. J. Organiz. Behav., 29: 1123-1144. https://doi.org/10.1002/job.558

[12] Kehoe, R.R., D.P. Lepak, F.S. Bentley (2016), "Let's Call a Star a Star: Task Performance, External Status, and Exceptional Contributors in Organizations", *Journal of Management* 44(5), 1848-1872.

[13] Hwang H., W.R. Reed and C. Hubbard (1992), "Compensating Wage Differentials and Unobserved Productivity", *Journal of Political Economy* 100(4), 835-858.

[14] Lehmann E., L. Simula and A. Trannoy (2014), "Tax Me if You Can! Optimal Nonlinear Income Tax Between Competing Governments", *Quarterly Journal of Economics*, 129(4), 1995-2030.

[15] Lewis T.R. and D. Sappington (1989), "Countervailing incentives in agency problems", *Journal of Economic Theory*, 49(2), 294-313.

[16] Martimort D. (1992), "Multi-principaux avec anti-selection", *Annales d'Economie et de Statistique*, 28, 1-37.

[17] Rochet J.-C. and L.A. Stole (2002), "Nonlinear Pricing with Random Participation" *Review of Economic Studies*, 69, 277-311.

[18] Rosen, S. (1986), "The Theory of Equalizing Differences", in: Ashenfelter, O., Layard, R. (Eds.), *Handbook of Labor Economics*, vol. 1, North Holland, 641-692.

[19] Shaked, A. and J. Sutton (1982), "Relaxing Price Competition Through Product Differentiation", *The Review of Economic Studies*, 49(1), 3-13.

[20] Stole L.A. (1992), "Mechanism Design Under Common Agency", mimeo, University of Chicago.

[21] Stole L.A. (1995), "Nonlinear pricing and oligopoly", *Journal of Economics & Management Strategy* 4, 529-562.

[22] Tan TF, Netessine S (2019), "When you work with a Superman, will you also fly? An empirical study of the effect of coworkers on worker performance." *Management Science*, 65(8), 3495-3517.

Appendix

A.1 Maximizing the social welfare function

Indirect utilities in SW, $U_i(\theta_j)$, cancel out and social welfare SW in the main text writes:

$$\begin{aligned}\max_{\{x_A(\theta_j),\hat{\gamma}_j\}_{j=1,2}} SW = \; & \lambda_1 \int_0^{\widehat{\gamma}_1} \left[k_A x_A(\theta_1) - \tfrac{1}{2}\theta_1 x_A^2(\theta_1) - \gamma\sigma + \alpha_1\left(\hat{\gamma}_1 - \tfrac{1}{2}\right)\right] dF(\gamma) \\ & + \lambda_2 \int_0^{\widehat{\gamma}_2} \left[k_A x_A(\theta_2) - \tfrac{1}{2}\theta_2 x_A^2(\theta_2) - \gamma\sigma\right] dF(\gamma) \\ & + \lambda_1 \int_{\widehat{\gamma}_1}^1 \left[k_B x_B(\theta_1) - \tfrac{1}{2}\theta_1 x_B^2(\theta_1) - (1-\gamma)\sigma - \alpha_1\left(\hat{\gamma}_1 - \tfrac{1}{2}\right)\right] dF(\gamma) \\ & + \lambda_2 \int_{\widehat{\gamma}_2}^1 \left[k_B x_B(\theta_2) - \tfrac{1}{2}\theta_2 x_B^2(\theta_2) - (1-\gamma)\sigma\right] dF(\gamma).\end{aligned}$$

Plugging the efficient effort levels (13) in the social welfare function the problem simplifies to

$$\begin{aligned}\max_{\{\hat{\gamma}_j\}_{j=1,2}} SW = \; & \lambda_1 \int_0^{\widehat{\gamma}_1} \left[\frac{k_A^2}{2\theta_1} - \gamma\sigma + \alpha_1\left(\hat{\gamma}_1 - \tfrac{1}{2}\right)\right] dF(\gamma) \\ & + \lambda_2 \int_0^{\widehat{\gamma}_2} \left[\frac{k_A^2}{2\theta_2} - \gamma\sigma\right] dF(\gamma) \\ & + \lambda_1 \int_{\widehat{\gamma}_1}^1 \left[\frac{k_B^2}{2\theta_1} - (1-\gamma)\sigma - \alpha_1\left(\hat{\gamma}_1 - \tfrac{1}{2}\right)\right] dF(\gamma) \\ & + \lambda_2 \int_{\widehat{\gamma}_2}^1 \left[\frac{k_B^2}{2\theta_2} - (1-\gamma)\sigma\right] dF(\gamma)\end{aligned}$$

Solving the integral and rearranging yields the following expression

$$\begin{aligned}\max_{\{\hat{\gamma}_j\}_{j=1,2}} \pi_i(\theta_j) = \; & \lambda_1 \widehat{\gamma}_1 \left[\frac{k_A^2}{2\theta_1} - \tfrac{1}{2}\sigma\hat{\gamma}_1 + \alpha_1\left(\hat{\gamma}_1 - \tfrac{1}{2}\right)\right] \\ & + \lambda_2 \widehat{\gamma}_2 \left[\frac{k_A^2}{2\theta_2} - \tfrac{1}{2}\sigma\hat{\gamma}_2\right] \\ & + \lambda_1 (1 - \widehat{\gamma}_1) \left[\left(\frac{k_B^2}{2\theta_1} - \tfrac{1}{2}(1-\widehat{\gamma}_1)\sigma - \alpha_1\left(\hat{\gamma}_1 - \tfrac{1}{2}\right)\right)\right] \\ & + \lambda_2 (1 - \widehat{\gamma}_2) \left[\left(\frac{k_B^2}{2\theta_2} - \tfrac{1}{2}(1-\widehat{\gamma}_2)\sigma\right)\right].\end{aligned}$$

Rearranging the previous formulation of the SW and isolating its three components, one derives the three expressions (30), (31) and (32) in the main text.

A.2 The multi-firm monopsonist

To understand why workers' sorting at the market equilibrium is inefficient let us derive the allocation generated by a multi-firm monopsonist maximizing the joint profits of firm A and firm B. This allows us to disentangle the profit maximization and the strategic interaction effects.

The monopsonist solves:

$$\max_{\{x_i(\theta_j),U_i(\theta_j)\}_{i=A,B;j=1,2}} E\left(\pi^M\right) = \begin{array}{l} \lambda_1\widehat{\gamma}_1\pi_A\left(\theta_1\right)+\lambda_2\widehat{\gamma}_2\pi_A\left(\theta_2\right) \\ +\lambda_1\left(1-\widehat{\gamma}_1\right)\pi_B\left(\theta_1\right)+\lambda_2\left(1-\widehat{\gamma}_2\right)\pi_B\left(\theta_2\right) \end{array} \qquad (P^M)$$

where $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ are given by (7) and (8) respectively, while $\pi_i\left(\theta_j\right)$ is defined by (12).

The firm optimally sets the utilities of marginal workers to zero: $\mathcal{U}_A\left(\theta_1,\widehat{\gamma}_1\right)=\mathcal{U}_B\left(\theta_1,\widehat{\gamma}_1\right)=0$ and $\mathcal{U}_A\left(\theta_2,\widehat{\gamma}_2\right)=\mathcal{U}_B\left(\theta_2,\widehat{\gamma}_2\right)=0$. This implies:

$$U_A\left(\theta_1\right)=\widehat{\gamma}_1\sigma-\alpha_1\left(\hat{\gamma}_1-\tfrac{1}{2}\right) \qquad (A.1)$$

$$U_B\left(\theta_1\right)=\left(1-\widehat{\gamma}_1\right)\sigma+\alpha_1\left(\hat{\gamma}_1-\tfrac{1}{2}\right) \qquad (A.2)$$

$$U_A\left(\theta_2\right)=\widehat{\gamma}_2\sigma \qquad (A.3)$$

$$U_B\left(\theta_2\right)=\left(1-\widehat{\gamma}_2\right)\sigma \qquad (A.4)$$

Substituting the first-best effort levels, $x_i^{fb}\left(\theta_j\right)$, in $\pi_i\left(\theta_j\right)$ and plugging the indirect utilities (A.1)-(A.4) into the expressions for $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ and $\pi_i\left(\theta_j\right)$, we obtain a simplified version of Program P^M which only depends on $\widehat{\gamma}_j$, $j=1,2$. Hence the monopsonist solves: $\max_{\{\widehat{\gamma}_j\}_{j=1,2}} E\left(\pi^M\right)$ and optimal marginal workers are:

$$\widehat{\gamma}_1^M = \frac{1}{2}+\frac{k_A^2-k_B^2}{8\theta_1\left(\sigma-\alpha_1\right)} \geq \frac{1}{2}$$

$$\widehat{\gamma}_2^M = \frac{1}{2}+\frac{k_A^2-k_B^2}{8\theta_2\sigma} \geq \frac{1}{2}$$

$$\widehat{\gamma}_1^M-\widehat{\gamma}_2^M = \frac{\left(k_A^2-k_B^2\right)\left(2\alpha_1\theta_1+\sigma(\theta_2-\theta_1)\right)}{8\theta_1\theta_2\sigma\left(\sigma-\alpha_1\right)} \geq 0$$

Comparing the previous inequalities with (33)-(34) and with (22)-(23) shows that the ranking of marginal types is like in (38) and in (39) in the main text. Hence, sorting designed by the monopsonist is not efficient but the distortion is lower than the one at the market equilibrium. One can easily check that, like $\widehat{\gamma}_1^*$ and $\widehat{\gamma}_1^{fb}$, also $\widehat{\gamma}_1^M$ is increasing in α_1. Finally, $\partial(\widehat{\gamma}_1^{fb}-\widehat{\gamma}_1^M)/\partial\alpha_1>0$ and $\partial(\widehat{\gamma}_1^M-\widehat{\gamma}_1^*)/\partial\alpha_1>0$ hold.

A.3 Proof of Remark 3

Substituting equilibrium contracts into the incentive compatibility constraints DIC_i and UIC_i, $i = A, B$, one can check that they are incentive compatible if the following conditions are met:

$$DIC_A : 3k_A^2(\theta_2 - \theta_1)^2 + 6\alpha_1\theta_1\theta_2^2 \geq \theta_2(\theta_2 - \theta_1)(k_A^2 - k_B^2) \tag{A.5}$$

$$UIC_A : 3k_A^2(\theta_2 - \theta_1)^2 + \theta_1(\theta_2 - \theta_1)(k_A^2 - k_B^2) \geq 6\alpha_1\theta_2\theta_1^2 \tag{A.6}$$

$$DIC_B : 3k_B^2(\theta_2 - \theta_1)^2 + \theta_2(\theta_2 - \theta_1)(k_A^2 - k_B^2) + 6\alpha_1\theta_2^2\theta_1 \geq 0 \tag{A.7}$$

$$UIC_B : 3k_B^2(\theta_2 - \theta_1)^2 \geq \theta_1(\theta_2 - \theta_1)(k_A^2 - k_B^2) + 6\alpha_1\theta_2\theta_1^2 \tag{A.8}$$

Hence, DIC_B always hold.

First consider $k_A = k_B = k$. One can see that DIC_A is always met in this case and that UIC_A and UIC_B become identical and they are satisfied if condition (44) holds. This proves part (i) of Remark 3.

Solving (A.5), (A.6) and (A.8) for α_1 (we omit A.7 because DIC_B is always slack) one finds conditions (41)-(43) in the main text and the three relevant threshold values for α_1. Recall that, if (41)-(43) are met, then all UIC and DIC are *slack* and the market equilibrium is incentive compatible. The best case scenario is when $\alpha_1^b \leq 0$ so that DIC_A is always met, together with $\alpha_1^a > 0$ so that UIC_B can be met for $\alpha_1 < \alpha_1^a$. Note that $\alpha_1^b \leq 0$ holds when $3k_A^2(\theta_2 - \theta_1) \geq \theta_2(k_A^2 - k_B^2))$ while $\alpha_1^a > 0$ holds if $3k_B^2(\theta_2 - \theta_1) > \theta_1(k_A^2 - k_B^2)$. Both the previous inequalities are thus met if $3k_B^2(\theta_2 - \theta_1) > \theta_2(k_A^2 - k_B^2))$, which proves part ($ii$) of Remark 3.

A.4 A first step to derive the relevant incentive constraints

Let us consider again incentive constraints in (41)–(43) and check which constraints becomes binding starting from $\alpha_1 = 0$ and letting α_1 grow larger. To do so one has to rank the threshold values α_1^a, α_1^b and α_1^c.

Remark 1 *Let us consider the market equilibrium under full information on ability and assume that conditions (44)-(46) do not hold; depending on the value of* α_1, *incentive constraints become relevant as follows.*
(i) For $\alpha_1 > \alpha_1^c$ *the binding constraints are* UIC_A *and* UIC_B.
(ii) When $\frac{\theta_2-\theta_1}{\theta_1} \leq \frac{k_A^2-k_B^2}{3k_B^2}$ *and* $\frac{\theta_2-\theta_1}{\theta_2} < \frac{k_A^2-k_B^2}{3k_A^2}$:
- for $0 < \alpha_1 \leq \alpha_1^b$ *the binding constraints are* UIC_B *and* DIC_A;
- for $\alpha_1^b < \alpha_1 \leq \alpha_1^c$ *the binding constraint is* UIC_B.
(iii) When $\frac{\theta_2-\theta_1}{\theta_1} > \frac{k_A^2-k_B^2}{3k_B^2}$ *and* $\frac{\theta_2-\theta_1}{\theta_2} \geq \frac{k_A^2-k_B^2}{3k_A^2}$:
- for $0 < \alpha_1 \leq \alpha_1^a$ *equilibrium contracts are incentive compatible*
- for $\alpha_1^a < \alpha_1 < \alpha_1^c$ *the binding constraint is* UIC_B *(if* $k_A = k_B$, *then* $\alpha_1^a = \alpha_1^c$ *holds and this case disappears).*

Proof. *(i)* The threshold α_1^c is always the highest among the three. Note that UIC_A binds for $\alpha_1 > \alpha_1^c$. If $\alpha_1 > \alpha_1^c$, then DIC_A is slack but UIC_B is binding. This explains point *(i)* of Remark 1.

The ranking of α_1^a and α_1^b depends on the relative magnitude of $\Delta\theta$ and Δk as follows.

(ii) When $\frac{\theta_2-\theta_1}{\theta_1} < \frac{k_A^2-k_B^2}{3k_B^2}$, $\alpha_1^a < 0$ holds and UIC_B necessarily binds. When $\frac{\theta_2-\theta_1}{\theta_2} < \frac{k_A^2-k_B^2}{3k_A^2}$, $\alpha_1^b > 0$ holds and DIC_A binds for $0 < \alpha_1 < \alpha_1^b$. Hence, when $\frac{\theta_2-\theta_1}{\theta_1} \leq \frac{k_A^2-k_B^2}{3k_B^2}$ and $\frac{\theta_2-\theta_1}{\theta_2} < \frac{k_A^2-k_B^2}{3k_A^2}$ the ranking of the three thresholds is $\alpha_1^a < 0 < \alpha_1^b < \alpha_1^c$. The binding constraints are thus as depicted in part *(ii)* of Remark 1.

(iii) When $\frac{\theta_2-\theta_1}{\theta_1} > \frac{k_A^2-k_B^2}{3k_B^2}$, $\alpha_1^a > 0$ and UIC_B binds only for $\alpha_1 > \alpha_1^a$. When $\frac{\theta_2-\theta_1}{\theta_2} \geq \frac{k_A^2-k_B^2}{3k_A^2}$, $\alpha_1^b < 0$ holds and DIC_A is always slack. Hence, when $\frac{\theta_2-\theta_1}{\theta_1} > \frac{k_A^2-k_B^2}{3k_B^2}$ and $\frac{\theta_2-\theta_1}{\theta_2} \geq \frac{k_A^2-k_B^2}{3k_A^2}$ the ranking of the three thresholds is $\alpha_1^b < 0 < \alpha_1^a < \alpha_1^c$. Hence, in this case the binding constraints are as depicted in part *(iii)* of Remark 1. ■

A.5 A second step to derive the relevant incentive constraints

The following results help us to fully characterize the regimes that are relevant for the firms.

Lemma 4 *(i) Two programs are relevant for firm A : the one where UIC_A is slack while DIC_A is binding and the one where DIC_A is slack while UIC_A is binding. The latter requires that $\frac{\sigma}{\sigma-\alpha_1} > \frac{\pi_A(\theta_2)}{\pi_A(\theta_1)} > 1$. (ii) Only one program is relevant for firm B, namely the one where DIC_B is slack while UIC_B is binding.*

In order to prove Lemma 4, let us first consider a preliminary step. Let us express incentive constraints in terms of firm's payoffs relative to each ability type, whereby DIC_i becomes

$$\pi_i(\theta_1) - \pi_i(\theta_2) \leq S_i(\theta_1) - S_i(\theta_2) - \frac{1}{2}(\theta_2 - \theta_1)\, x_i^2(\theta_2)$$

and UIC_i takes the form

$$\pi_i(\theta_1) - \pi_i(\theta_2) \geq S_i(\theta_1) - S_i(\theta_2) - \frac{1}{2}(\theta_2 - \theta_1)\, x_i^2(\theta_1),$$

where

$$S_i(\theta_j) \equiv k_i x_i(\theta_j) - \frac{1}{2}\theta_j x_i^2(\theta_j)$$

is the surplus realized by a worker of type θ_j providing effort $x_i(\theta_j)$ for firm i (again, absent the mismatch disutility and the benefit accruing from the premium for coworkers' ability and the mismatch disutility, when $j = 1$).

Remark 2 (i) *If DIC_i is binding for firm $i = A, B$, then per-worker payoffs are such that $\pi_i(\theta_1) > \pi_i(\theta_2)$.* ($ii$) *If UIC_i is binding for firm $i = A, B$, then per-worker payoffs are such that $\pi_i(\theta_2) > \pi_i(\theta_1)$.*

Proof. The proof of this result follows an argument similar to the one developed by Rochet and Stole (2002). When DIC_i is binding for firm $i = A, B$, effort levels are such that $x_i(\theta_2) \leq x_i^{FB}(\theta_2)$ and $x_i(\theta_1) = x_i^{FB}(\theta_1)$; namely, the high-ability type gets the first-best allocation while the effort of the low-ability type is downward distorted. Moreover, when DIC_i is binding, one has

$$\pi_i(\theta_1) - \pi_i(\theta_2) = S_i(\theta_1) - S_i(\theta_2) - \frac{1}{2}(\theta_2 - \theta_1) x_i^2(\theta_2).$$

The right-hand-side of the above equality is minimized when $x_i(\theta_2)$ is the highest possible, that is when it equals the first-best effort level and surplus $S_i(\theta_2)$ is maximized. Substituting for such effort level yields

$$\begin{array}{c} \pi_i(\theta_1) - \pi_i(\theta_2) = S_i(\theta_1) - S_i(\theta_2) - \frac{1}{2}(\theta_2 - \theta_1) x_i^2(\theta_2) \\ \geq \frac{k_i^2(\theta_2-\theta_1)}{2\theta_1\theta_2} - \frac{k_i^2(\theta_2-\theta_1)}{2\theta_2^2} = \frac{k_i^2(\theta_2-\theta_1)^2}{2\theta_1\theta_2^2} > 0 \end{array}.$$

Similarly, when UIC_i is binding for firm $i = A, B$, effort levels are such that $x_i(\theta_2) = x_i^{FB}(\theta_2)$ and $x_i(\theta_1) \geq x_i^{FB}(\theta_1)$; namely, the low-ability type gets the first-best while the effort of the high-ability type is distorted upwards. Moreover, when UIC_i is binding, one has

$$\pi_i(\theta_1) - \pi_i(\theta_2) = S_i(\theta_1) - S_i(\theta_2) - \frac{1}{2}(\theta_2 - \theta_1) x_i^2(\theta_1).$$

The right-hand-side of the above equality is maximized when $x_i(\theta_1)$ is the lowest possible, that is when it equals the first-best effort level and surplus $S_i(\theta_1)$ is maximized. Substituting for such effort level yields

$$\pi_i(\theta_1) - \pi_i(\theta_2) = S_i(\theta_1) - S_i(\theta_2) - \tfrac{1}{2}(\theta_2 - \theta_1) x_i^2(\theta_1) \quad \leq -\frac{k_i^2(\theta_2-\theta_1)^2}{2\theta_1^2\theta_2} < 0 .$$

When neither DIC_i nor UIC_i is binding, then each firm sets all effort levels at the first-best and the difference in per-worker payoffs $\pi_i(\theta_1) - \pi_i(\theta_2)$ can be either positive or negative. ■

Let us now move to the actual proof of Lemma 4. As mentioned in the main text we assume that, under asymmetric information on ability, it is still true that $\hat{\gamma}_1 \geq \hat{\gamma}_2$. What follows builds on the proof of Propositions 4 and 5 in Barigozzi and Burani (2019).

- Firm A solves:

$$\begin{array}{lll} \max_{\{x_A(\theta_j),U_A(\theta_j)\}_{j=1,2}} E(\pi_A) = & \lambda_1 \frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)}\left(k_A x_A(\theta_1) - \frac{1}{2}\theta_1 x_A^2(\theta_1) - U_A(\theta_1)\right) & \\ & +\lambda_2 \frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{2\sigma}\left(k_A x_A(\theta_2) - \frac{1}{2}\theta_2 x_A^2(\theta_2) - U_A(\theta_2)\right) & \\ & & \\ s.t. & U_A(\theta_1) - U_A(\theta_2) - \frac{1}{2}(\theta_2-\theta_1) x_A^2(\theta_2) \geq 0 & (\mu_{D_A}) \\ & U_A(\theta_2) - U_A(\theta_1) + \frac{1}{2}(\theta_2-\theta_1) x_A^2(\theta_1) \geq 0 & (\mu_{U_A}) \end{array}$$

$$(P_A)$$

where $\frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)} = \hat{\gamma}_1$ and $\frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{2\sigma} = \hat{\gamma}_2$. In addition, $\mu_{D_A} \geq 0$ and $\mu_{U_A} \geq 0$ are the Lagrangian multiplier of the DIC_A and UIC_A incentive constraint, respectively.

FOCs w.r.t. $x_A(\theta_j)$, $j = 1,2$, respectively are:

$$\lambda_1 \frac{\sigma - \alpha_1 + U_A(\theta_1) - U_B(\theta_1)}{2(\sigma - \alpha_1)} (k_A - \theta_1 x_A(\theta_1)) + \mu_{U_A} (\theta_2 - \theta_1) x_A(\theta_1) = 0 \tag{A.9}$$

$$\lambda_2 \frac{\sigma + U_A(\theta_2) - U_B(\theta_2)}{2\sigma} (k_A - \theta_2 x_A(\theta_2)) - \mu_{D_A} (\theta_2 - \theta_1) x_A(\theta_2) = 0 \tag{A.10}$$

FOCs w.r.t. $U_A(\theta_j)$, $j = 1,2$, respectively are:

$$\frac{\lambda_1}{2(\sigma - \alpha_1)} \left(k_A x_A(\theta_1) - \frac{1}{2}\theta_1 x_A^2(\theta_1) - U_A(\theta_1) \right) - \lambda_1 \frac{\sigma - \alpha_1 + U_A(\theta_1) - U_B(\theta_1)}{2(\sigma - \alpha_1)} + \mu_{D_A} - \mu_{U_A} = 0 \tag{A.11}$$

$$\frac{\lambda_2}{2\sigma} \left(k_A x_A(\theta_2) - \frac{1}{2}\theta_2 x_A^2(\theta_2) - U_A(\theta_2) \right) - \lambda_2 \frac{\sigma + U_A(\theta_2) - U_B(\theta_2)}{2\sigma} - \mu_{D_A} + \mu_{U_A} = 0 \tag{A.12}$$

- When DIC and UIC are both binding, (40) writes:

$$\frac{1}{2}(\theta_2 - \theta_1) x_A^2(\theta_2) = U_A(\theta_1) - U_A(\theta_2) = \frac{1}{2}(\theta_2 - \theta_1) x_A^2(\theta_1)$$

and $x_A(\theta_1) = x_A(\theta_2) = x_A$ must hold. In addition μ_{D_A} and μ_{U_A} must be strictly positive and (A.9) and (A.10) imply:

$$\lambda_1 \hat{\gamma}_1 (k_A - \theta_1 x_A) + \mu_{U_A} (\theta_2 - \theta_1) x_A = 0 \tag{A.13}$$

$$-\lambda_2 \hat{\gamma}_2 (k_A - \theta_2 x_A) + \mu_{D_A} (\theta_2 - \theta_1) x_A = 0 \tag{A.14}$$

Summing up (A.13) and (A.14) gives:

$$\lambda_1 \hat{\gamma}_1 (k_A - \theta_1 x_A) - \lambda_2 \hat{\gamma}_2 (k_A - \theta_2 x_A) + (\mu_{U_A} + \mu_{D_A}) (\theta_2 - \theta_1) x_A = 0.$$

Hence:

$$\mu_{U_A} + \mu_{D_A} = \frac{\lambda_2 \hat{\gamma}_2 (k_A - \theta_2 x_A) - \lambda_1 \hat{\gamma}_1 (k_A - \theta_1 x_A)}{(\theta_2 - \theta_1) x_A} > 0,$$

which implies $\lambda_2 > \lambda_1 \frac{(k_A - \theta_1 x_A)}{(k_A - \theta_2 x_A)} \frac{\hat{\gamma}_1}{\hat{\gamma}_2}$ or, given that both ratios appearing in the right hand side of the previous inequality are larger than one, $\lambda_2 \gg \lambda_1$.

When instead λ_2 is not larger enough than λ_1, DIC and UIC cannot be both binding because FOCs (A.9) and (A.10) become mutually incompatible. This suggests that only one or the other incentive constraint will typically bind at a given point.

- When DIC_A is slack while UIC_A is binding, then $\mu_{D_A} = 0$ and $\mu_{U_A} > 0$. Hence (A.11) and (A.12) become:

$$\frac{\lambda_1}{2(\sigma - \alpha_1)} \left(k_A x_A(\theta_1) - \frac{1}{2}\theta_1 x_A^2(\theta_1) - U_A(\theta_1) \right) - \lambda_1 \frac{\sigma - \alpha_1 + U_A(\theta_1) - U_B(\theta_1)}{2(\sigma - \alpha_1)} - \mu_{U_A} = 0$$

$$\frac{\lambda_2}{2\sigma} \left(k_A x_A(\theta_2) - \frac{1}{2}\theta_2 x_A^2(\theta_2) - U_A(\theta_2) \right) - \lambda_2 \frac{\sigma + U_A(\theta_2) - U_B(\theta_2)}{2\sigma} + \mu_{U_A} = 0$$

Hence, dropping μ_{U_A}

$$\frac{\lambda_1}{2(\sigma-\alpha_1)}\left(k_A x_A(\theta_1)-\frac{1}{2}\theta_1 x_A^2(\theta_1)-U_A(\theta_1)\right)-\lambda_1\frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)}>0 \tag{A.15}$$

$$\frac{\lambda_2}{2\sigma}\left(k_A x_A(\theta_2)-\frac{1}{2}\theta_2 x_A^2(\theta_2)-U_A(\theta_2)\right)-\lambda_2\frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{2\sigma}<0 \tag{A.16}$$

where $\pi_A(\theta_1)\equiv k_A x_A(\theta_1)-\frac{1}{2}\theta_1 x_A^2(\theta_1)-U_A(\theta_1)$ and $\pi_A(\theta_2)\equiv k_A x_A(\theta_2)-\frac{1}{2}\theta_2 x_A^2(\theta_2)-U_A(\theta_2)$. Substituting per-worker profits in (A.15) and (A.25) and simplifying:

$$\frac{\pi_A(\theta_1)}{2(\sigma-\alpha_1)}>\frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)}$$
$$\frac{\pi_A(\theta_2)}{2\sigma}<\frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{2\sigma}$$

Recall that it must be $\hat{\gamma}_1\geq\hat{\gamma}_2$ or $\frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)}\geq\frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{2\sigma}$. Hence, the previous two inequalities imply:

$$\frac{\pi_A(\theta_1)}{\sigma-\alpha_1}>\frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{\sigma-\alpha_1}\geq\frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{\sigma}>\frac{\pi_A(\theta_2)}{\sigma}.$$

Per-worker profits must thus satisfy: $\frac{\pi_A(\theta_1)}{\sigma-\alpha_1}>\frac{\pi_A(\theta_2)}{\sigma}$. In addition, from Remark 2, UIC_A binding implies that $\pi_A(\theta_1)<\pi_A(\theta_2)$. Hence, when UIC_A is binding, $\frac{\sigma}{\sigma-\alpha_1}>\frac{\pi_A(\theta_2)}{\pi_A(\theta_1)}>1$ must hold.

- When UIC_A is slack while DIC_A is binding, then $\mu_{D_A}>0$ and $\mu_{U_A}=0$. Hence (A.11) and (A.12) become:

$$\frac{\lambda_1}{2(\sigma-\alpha_1)}\left(k_A x_A(\theta_1)-\frac{1}{2}\theta_1 x_A^2(\theta_1)-U_A(\theta_1)\right)-\lambda_1\frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)}+\mu_{D_A}=0 \tag{A.17}$$

$$\frac{\lambda_2}{2\sigma}\left(k_A x_A(\theta_2)-\frac{1}{2}\theta_2 x_A^2(\theta_2)-U_A(\theta_2)\right)-\lambda_2\frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{2\sigma}-\mu_{D_A}=0 \tag{A.18}$$

Substituting for per-worker profits, dropping μ_{D_A} and rearranging:

$$\frac{\pi_A(\theta_1)}{2(\sigma-\alpha_1)}<\frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)}$$
$$\frac{\pi_A(\theta_2)}{2\sigma}>\frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{2\sigma}$$

Recall that it must be $\hat{\gamma}_1\geq\hat{\gamma}_2$, which implies $\frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)}\geq\frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{2\sigma}$. Hence the previous two inequalities are compatible with both the following chains of inequalities:

$$\frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{\sigma-\alpha_1}>\frac{\pi_A(\theta_1)}{\sigma-\alpha_1}>\frac{\pi_A(\theta_2)}{\sigma}>\frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{\sigma} \tag{A.19}$$

and

$$\frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{\sigma-\alpha_1}>\frac{\pi_A(\theta_2)}{\sigma}>\frac{\pi_A(\theta_1)}{\sigma-\alpha_1}>\frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{\sigma}.$$

From Remark 2, DIC_A binding implies that $\pi_A(\theta_1)>\pi_A(\theta_2)$. Note that, if $\pi_A(\theta_1)>\pi_A(\theta_2)$, then *a fortiori* $\frac{\pi_A(\theta_1)}{\sigma-\alpha_1}>\frac{\pi_A(\theta_2)}{\sigma}$ and thus the chain of inequalities in (A.19) must hold. To conclude, the program where UIC_A is slack while DIC_A is binding is possible without additional constraints on $\sigma-\alpha_1$.

- Firm B solves:

$$\begin{aligned}\max_{\{x_B(\theta_j),U_B(\theta_j)\}_{j=1,2}} E(\pi_B)= \;& \lambda_1\tfrac{\sigma-\alpha_1-U_A(\theta_1)+U_B(\theta_1)}{2(\sigma-\alpha_1)}\left(k_Bx_B(\theta_1)-\tfrac{1}{2}\theta_1x_B^2(\theta_1)-U_B(\theta_1)\right)\\ &+\lambda_2\tfrac{\sigma-U_A(\theta_2)+U_B(\theta_2)}{2\sigma}\left(k_Bx_B(\theta_2)-\tfrac{1}{2}\theta_2x_B^2(\theta_2)-U_B(\theta_2)\right)\end{aligned}$$

$$\begin{aligned} s.t.\quad & U_B(\theta_1)-U_B(\theta_2)-\tfrac{1}{2}(\theta_2-\theta_1)x_B^2(\theta_2)\geq 0 \qquad (\mu_{D_B})\\ & U_B(\theta_2)-U_B(\theta_1)+\tfrac{1}{2}(\theta_2-\theta_1)x_B^2(\theta_1)\geq 0 \qquad (\mu_{U_B})\end{aligned} \tag{P_B}$$

Where $\mu_{D_B}\geq 0$ and $\mu_{U_B}\geq 0$ are the Lagrangian multiplier of the DIC_B and UIC_B incentive constraint, respectively. Using the same reasoning as before, only one or the other incentive constraint will typically bind at a given point.

FOCs w.r.t. $x_B(\theta_j)$, $j=1,2$, respectively are:

$$\lambda_1\left(\frac{\sigma-\alpha_1-U_A(\theta_1)+U_B(\theta_1)}{2(\sigma-\alpha_1)}\right)(k_B-\theta_1x_B(\theta_1))+\mu_{U_B}(\theta_2-\theta_1)x_B(\theta_1)=0 \tag{A.20}$$

$$\lambda_2\left(\frac{\sigma-U_A(\theta_2)+U_B(\theta_2)}{2\sigma}\right)(k_B-\theta_2x_B(\theta_2))-\mu_{D_B}(\theta_2-\theta_1)x_B(\theta_2)=0 \tag{A.21}$$

FOCs w.r.t. $U_B(\theta_j)$, $j=1,2$, respectively are:

$$\frac{\lambda_1}{2(\sigma-\alpha_1)}\left(k_Bx_B(\theta_1)-\frac{1}{2}\theta_1x_B^2(\theta_1)-U_B(\theta_1)\right)-\lambda_1\frac{\sigma-\alpha_1-U_A(\theta_1)+U_B(\theta_1)}{2(\sigma-\alpha_1)}+\mu_{D_B}-\mu_{U_B}=0 \tag{A.22}$$

$$\frac{\lambda_2}{2\sigma}\left(k_Bx_B(\theta_2)-\frac{1}{2}\theta_2x_B^2(\theta_2)-U_B(\theta_2)\right)-\lambda_2\frac{\sigma-U_A(\theta_2)+U_B(\theta_2)}{2\sigma}-\mu_{D_A}+\mu_{U_A}=0 \tag{A.23}$$

- Let us consider the instance where $\mu_{D_B}=0$ while $\mu_{U_B}>0$ so that DIC_B is slack while UIC_B is binding. From (A.22) and (A.23):

$$\frac{\lambda_1}{2(\sigma-\alpha_1)}\left(k_Bx_B(\theta_1)-\frac{1}{2}\theta_1x_B^2(\theta_1)-U_B(\theta_1)\right)-\lambda_1\frac{\sigma-\alpha_1-U_A(\theta_1)+U_B(\theta_1)}{2(\sigma-\alpha_1)}-\mu_{U_B}=0 \tag{A.24}$$

$$\frac{\lambda_2}{2\sigma}\left(k_Bx_B(\theta_2)-\frac{1}{2}\theta_2x_B^2(\theta_2)-U_B(\theta_2)\right)-\lambda_2\frac{\sigma-U_A(\theta_2)+U_B(\theta_2)}{2\sigma}+\mu_{U_A}=0 \tag{A.25}$$

Substituting for per-workers profits $\pi_B(\theta_1)$ in (A.24) and (A.25) and dropping μ_{U_B} :

$$\frac{\pi_B(\theta_1)}{2(\sigma-\alpha_1)} > \frac{\sigma-\alpha_1-U_A(\theta_1)+U_B(\theta_1)}{2(\sigma-\alpha_1)} \tag{A.26}$$

$$\frac{\pi_B(\theta_2)}{2\sigma} < \frac{\sigma-U_A(\theta_2)+U_B(\theta_2)}{2\sigma} \tag{A.27}$$

Recall that it must be $1-\hat{\gamma}_1 \leq 1-\hat{\gamma}_2$, which implies $\frac{\sigma-\alpha_1-U_A(\theta_1)+U_B(\theta_1)}{\sigma-\alpha_1} \leq \frac{\sigma-U_A(\theta_2)+U_B(\theta_2)}{\sigma}$. In addition, from Remark 2, UIC_B binding implies that $\pi_B(\theta_1) < \pi_B(\theta_2)$. Thus, inequalities (A.26) and (A.27) are fully compatible and we may have either $\frac{\pi_B(\theta_2)}{\sigma} > \frac{\pi_B(\theta_1)}{\sigma-\alpha_1}$ if α_1 is sufficiently small or the opposite. So the case where DIC_B is slack while UIC_B is binding is possible and no additional constraints are required.

- Let us consider the case where $\mu_{D_B} > 0$ while $\mu_{U_B} = 0$ so that DIC_B is binding while UIC_B is slack. From (A.22) and (A.23):

$$\frac{\lambda_1}{2(\sigma-\alpha_1)}\left(k_B x_B(\theta_1) - \frac{1}{2}\theta_1 x_B^2(\theta_1) - U_B(\theta_1)\right) - \lambda_1\frac{\sigma-\alpha_1-U_A(\theta_1)+U_B(\theta_1)}{2(\sigma-\alpha_1)} + \mu_{D_B} = 0 \tag{A.28}$$

$$\frac{\lambda_2}{2\sigma}\left(k_B x_B(\theta_2) - \frac{1}{2}\theta_2 x_B^2(\theta_2) - U_B(\theta_2)\right) - \lambda_2\frac{\sigma-U_A(\theta_2)+U_B(\theta_2)}{2\sigma} - \mu_{D_A} = 0 \tag{A.29}$$

Substituting for profits in (A.28) and (A.29) and dropping μ_{D_B}:

$$\frac{\pi_B(\theta_1)}{\sigma-\alpha_1} < \frac{\sigma-\alpha_1-U_A(\theta_1)+U_B(\theta_1)}{\sigma-\alpha_1} \tag{A.30}$$

$$\frac{\pi_B(\theta_2)}{\sigma} > \frac{\sigma-U_A(\theta_2)+U_B(\theta_2)}{\sigma} \tag{A.31}$$

Considering that it must be $\frac{\sigma-U_A(\theta_2)+U_B(\theta_2)}{\sigma} \geq \frac{\sigma-\alpha_1-U_A(\theta_1)+U_B(\theta_1)}{\sigma-\alpha_1}$, inequalities in (A.30) and (A.31) imply that:

$$\frac{\pi_B(\theta_2)}{\sigma} > \frac{\sigma-U_A(\theta_2)+U_B(\theta_2)}{\sigma} \geq \frac{\sigma-\alpha_1-U_A(\theta_1)+U_B(\theta_1)}{\sigma-\alpha_1} > \frac{\pi_B(\theta_1)}{\sigma-\alpha_1}$$

But, from Remark 2, when DIC_B is binding, $\pi_B(\theta_1) > \pi_B(\theta_2)$ holds. Hence $\frac{\pi_B(\theta_2)}{\sigma} > \frac{\pi_B(\theta_1)}{\sigma-\alpha_1}$ is a contradiction and it is impossible that DIC_B is binding while UIC_B is slack. Hence, the program of firm B is compatible only with UIC_B binding.

A.6 Regime 2: only UIC_B is binding

In this regime, firm A solves an unconstrained program while firm B solves a program where DIC_B is slack while UIC_B is binding so that $\mu_{D_B} = 0$ while $\mu_{U_B} > 0$.

Hence, considering the program of firm A, P_A, the FOCs w.r.t. $x_A(\theta_j)$ are the same as under full information and the efforts are set at the efficient levels, $x_A^{**}(\theta_1) = \frac{k_A}{\theta_1} = x_A^{fb}(\theta_1)$ and

$x_A^{**}(\theta_2) = \frac{k_A}{\theta_2} = x_A^{fb}(\theta_2)$. From the FOCs w.r.t. $U_A(\theta_j)$ one observes that firm A's reaction functions are the same as under full information (see (A.11) and (A.12) and expressions (14) and (16)): $U_A(\theta_1) = \frac{k_A^2 + 2\theta_1(U_B(\theta_1)+\alpha_1-\sigma)}{4\theta_1}$ and $U_A(\theta_2) = \frac{k_A^2 + 2\theta_2(U_B(\theta_2)-\sigma)}{4\theta_2}$.

Considering the program of firm B, P_B, with $\mu_{D_B} = 0$ and $\mu_{U_B} > 0$, from (A.20) and (A.21)one can check that $x_B(\theta_2)$ is set at the efficient level, i.e. $x_B^{**}(\theta_2) = \frac{k_B}{\theta_2} = x_B^{fb}(\theta_2)$, while the FOCs w.r.t. $x_B(\theta_1)$ now writes:

$$\lambda_1 (\theta_1 x_B(\theta_1) - k_B) \frac{U_A(\theta_1) - U_B(\theta_1) + \alpha_1 - \sigma}{2(\sigma - \alpha_1)} + \mu_{U_B}(\theta_2 - \theta_1) x_B(\theta_1) = 0$$

hence

$$\lambda_1 (\theta_1 x_B(\theta_1) - k_B) \frac{U_A(\theta_1) - U_B(\theta_1) + \alpha_1 - \sigma}{2(\sigma - \alpha_1)} < 0$$

where $\sigma - \alpha_1 > 0$ holds while $U_A(\theta_1) - U_B(\theta_1) + \alpha_1 - \sigma < 0$ is the condition assuring that an interior solution for the marginal worker of type θ_1 exists or that $\hat{\gamma}_1 < 1$. Thus it must be $\theta_1 x_B(\theta_1) - k_B > 0$ or $x_B^{**}(\theta_1) > \frac{k_B}{\theta_1} = x_B^{fb}(\theta_1)$ meaning that the effort of high-types is upward distorted.

By substituting $x_B^{**}(\theta_2) = \frac{k_B}{\theta_2}$ and the reaction function $U_A(\theta_2) = \frac{k_A^2 + 2\theta_2(U_B(\theta_2)-\sigma)}{4\theta_2}$ in (A.23) one has:

$$U_B^{**}(\theta_2) = \frac{k_A^2 + 2k_B^2}{6\theta_2} - \sigma + \frac{4}{3\lambda_2}\sigma\mu_{U_B} > \frac{k_A^2 + 2k_B^2}{6\theta_2} - \sigma = U_B^*(\theta_2).$$

From (A.22):

$$\lambda_1 \frac{k_B x_B(\theta_1) - \frac{1}{2}\theta_1 x_B^2(\theta_1) - U_B(\theta_1)}{2(\sigma - \alpha_1)} - \lambda_1 \frac{\sigma - \alpha_1 - U_A(\theta_1) + U_B(\theta_1)}{2(\sigma - \alpha_1)} - \mu_{U_B} = 0.$$

Substituting the reaction function of firm A, $U_A(\theta_1) = \frac{k_A^2 + 2\theta_1(U_B(\theta_1)+\alpha_1-\sigma)}{4\theta_1}$, and rearranging:

$$U_B^{**}(\theta_1) < \frac{2}{3}\left(k_B x_B^{**}(\theta_1) - \frac{1}{2}\theta_1 \left(x_B^{**}(\theta_1)\right)^2\right) + \alpha_1 - \sigma + \frac{k_A^2}{6\theta_1}. \tag{A.32}$$

By solving for $U_B(\theta_1)$ the FOC (A.22) of the unconstrained program P_B where $\mu_{U_B} = \mu_{D_B} = 0$ and $x_B(\theta_j) = x_B^{fb}(\theta_j)$, $j = 1, 2$ one instead obtains:

$$U_B^*(\theta_1) = \frac{2}{3}\left(k_B x_B^{fb}(\theta_1) - \frac{1}{2}\theta_1 \left(x_B^{fb}(\theta_1)\right)^2\right) + \alpha_1 - \sigma + \frac{k_A^2}{6\theta_1}. \tag{A.33}$$

Because the surplus $k_B x_B(\theta_1) - \frac{1}{2}\theta_1 x_B^2(\theta_1)$ is maximized for $x_B^{fb}(\theta_1)$, comparing (A.32) and (A.33) one observes that $U_B^{**}(\theta_1) < U_B^*(\theta_1)$.

Indirect utilities $U_i(\theta_j)$ are strategic complements, hence $U_B^{**}(\theta_1) < U_B^*(\theta_1)$ implies $U_A^{**}(\theta_1) < U_A^*(\theta_1)$ and $U_B^{**}(\theta_2) > U_B^*(\theta_2)$ implies $U_A^{**}(\theta_2) > U_A^*(\theta_2)$. However, given that the slopes of the two reaction functions are $\frac{\partial U_A(\theta_1)}{\partial U_B(\theta_1)} = \frac{\partial U_B(\theta_1)}{\partial U_A(\theta_1)} = \frac{1}{2}$ (see (14) and (15)), the change in $U_A^{**}(\theta_j)$ is lower than the change in $U_B^{**}(\theta_j)$, $j = 1, 2$, and we can conclude that $\hat{\gamma}_1$ increases whereas $\hat{\gamma}_2$ decreases w.r.t. the full information equilibrium.

A.6.1 Numerical simulations under Regime 2

We present here some numerical simulations to show that all the neglected conditions and the omitted constraints are indeed met. Take the following parameter values: $k_A = 6.5$, $k_B = 4$, $\theta_1 = 1$, $\theta_2 = 1.2$, $\sigma = 6$, $\alpha_1 = 0.3$ and $\lambda_1 = \lambda_2 = 0.5$. These values assure that the solution is interior in the case of full information on ability because conditions $(k_A^2 - k_B^2 < 6\theta_1(\sigma - \alpha_1))$ and $(k_A^2 - k_B^2 < 6\theta_2\sigma)$ are met (see Proposition 1); in addition they satisfy conditions $\frac{\theta_2-\theta_1}{\theta_1} \leq \frac{k_A^2-k_B^2}{3k_A^2}$ and $0.14 = \alpha_1^b < \alpha_1 \leq \alpha_1^c = 1.43$ of Regime 2. Optimal contracts under full information on ability are the following: $\{x_A^*(\theta_1), U_A^*(\theta_1)\} = (6.5, 11.05)$; $\{x_A^*(\theta_2), U_A^*(\theta_2)\} = (5.416, 7.958)$; $\{x_B^*(\theta_1), U_B^*(\theta_1)\} = (3.999, 6.675)$; $\{x_B^*(\theta_2), U_B^*(\theta_2)\} = (3.333, 4.312)$; with $\hat{\gamma}_1^* = 0.872$ and $\hat{\gamma}_2^* = 0.804$. Under screening, optimal contracts become: $\{x_A^{**}(\theta_1), U_A^{**}(\theta_1)\} = (6.5, 10.912)$; $\{x_A^{**}(\theta_2), U_A^{**}(\theta_2)\} = (5.416, 8.090)$; $\{x_B^{**}(\theta_1), U_B^{**}(\theta_1)\} = (4.270, 6.399)$; $\{x_B^{**}(\theta_2), U_B^{**}(\theta_2)\} = (3.333, 4.576)$; with $\hat{\gamma}_1^{**} = 0.895$ and $\hat{\gamma}_2^{**} = 0.792$. Comparing the solution under screening with the one under full information on ability one observes that all the effort levels remain the same except $x_B^{**}(\theta_1)$ which is upward distorted. Indirect utilities decrease for high-types and increase for low-types with respect to full information contracts. As a result of the changes in indirect utilities, $\hat{\gamma}_1^{**} > \hat{\gamma}_1^*$ whereas $\hat{\gamma}_2^{**} < \hat{\gamma}_2^*$ hold.

The marginal workers' utilities under screening are above zero showing that the workers' participation constraints are slack. Finally, all profits per-worker are strictly positive and $0 < \pi_B(\theta_1) < \pi_B(\theta_2)$ holds.

Interestingly, an increase in the parameter α_1 (i.e. $\alpha_1 = 0.8$) leads to a corner solution with $\hat{\gamma}_1^{**} = 1$ in the equilibrium with screening but it is still compatible with an interior solution in the full information equilibrium.

A.7 Regime 1: DIC_A and UIC_B are binding

In this regime, firm A solves the program where DIC_A is binding while UIC_A is slack so that $\mu_{D_A} > 0$ while $\mu_{U_A} = 0$; while firm B solves the same program as before where DIC_B is slack while UIC_B is binding so that $\mu_{D_B} = 0$ while $\mu_{U_B} > 0$.

Let us start from firm A. From (A.9) and (A.10) we have that $x_A^{**}(\theta_1) = \frac{k_A}{\theta_1} = x_A^{fb}(\theta_1)$ while the following equation holds for $x_A(\theta_2)$:

$$\lambda_2(k_A - \theta_2 x_A(\theta_2)) \frac{U_A(\theta_2) - U_B(\theta_2) + \sigma}{2\sigma} - \mu_{D_A}(\theta_2 - \theta_1) x_A(\theta_2) = 0$$

or

$$\lambda_2(k_A - \theta_2 x_A(\theta_2)) \frac{U_A(\theta_2) - U_B(\theta_2) + \sigma}{2\sigma} > 0$$

where $\frac{U_A(\theta_2)-U_B(\theta_2)+\sigma}{2\sigma} = \hat{\gamma}_2 > 0$ in the case of interior solutions. Hence it must be that $x_A^{**}(\theta_2) < \frac{k_A}{\theta_2} = x_A^{fb}(\theta_2)$ meaning that the effort of low-types is downward distorted.

From FOCs (A.11) and (A.12), one has:

$$\lambda_1 \frac{k_A x_A(\theta_1) - \frac{1}{2}\theta_1 x_A^2(\theta_1) - U_A(\theta_1)}{2(\sigma - \alpha_1)} - \lambda_1 \frac{\sigma - \alpha_1 + U_A(\theta_1) - U_B(\theta_1)}{2(\sigma - \alpha_1)} + \mu_{D_A} = 0. \tag{A.34}$$

and

$$\lambda_2 \frac{k_A x_A(\theta_2) - \frac{1}{2}\theta_2 x_A^2(\theta_2) - U_A(\theta_2)}{2\sigma} - \lambda_2 \frac{\sigma + U_A(\theta_2) - U_B(\theta_2)}{2\sigma} - \mu_{D_B} = 0. \tag{A.35}$$

Recall that the surplus is $S_A(\theta_j) \equiv k_A x_A(\theta_j) - \frac{1}{2}\theta_j x_A^2(\theta_j)$ with $S_A^{**}(\theta_1) = S_A^{fb}(\theta_1)$ and $S_A^{**}(\theta_2) < S_A^{fb}(\theta_2)$ because $x_A^{**}(\theta_1)$ is at the efficient level whereas $x_A^{**}(\theta_2)$ is distorted. Substituting and rearranging the two previous FOCs, they respectively become:

$$S_A^{fb}(\theta_1) - 2U_A^{**}(\theta_1) + U_B^{**}(\theta_1) - \sigma + \alpha_1 < 0 \tag{A.36}$$

and

$$S_A^{**}(\theta_2) - 2U_A^{**}(\theta_2) + U_B^{**}(\theta_2) - \sigma > 0 \tag{A.37}$$

Also note that, in the unconstrained program of firm A where $\mu_{D_B} = \mu_{U_B} = 0$, the previous FOCs can be respectively written as:

$$S_A^{fb}(\theta_1) - 2U_A^{*}(\theta_1) + U_B^{*}(\theta_1) - \sigma + \alpha_1 = 0 \tag{A.38}$$

and

$$S_A^{fb}(\theta_2) - 2U_A^{*}(\theta_2) + U_B^{*}(\theta_2) - \sigma = 0. \tag{A.39}$$

As for type θ_1, putting together (A.36) and (A.38) one has:

$$S_A^{fb}(\theta_1) = 2U_A^{*}(\theta_1) - U_B^{*}(\theta_1) + \sigma - \alpha_1 <$$
$$2U_A^{**}(\theta_1) - U_B^{**}(\theta_1) + \sigma - \alpha_1$$

or

$$2U_A^{*}(\theta_1) - U_B^{*}(\theta_1) < 2U_A^{**}(\theta_1) - U_B^{**}(\theta_1) \Leftrightarrow$$
$$2\left(U_A^{**}(\theta_1) - U_A^{*}(\theta_1)\right) - \left(U_B^{**}(\theta_1) - U_B^{*}(\theta_1)\right) > 0 \tag{A.40}$$

As for type θ_2, from (A.37) and (A.39) one can write:

$$S_A^{fb}(\theta_2) = 2U_A^{*}(\theta_2) - U_B^{*}(\theta_2) + \sigma >$$
$$S_A^{**}(\theta_2) > 2U_A^{**}(\theta_2) - U_B^{**}(\theta_2) + \sigma$$

or

$$2U_A^{*}(\theta_2) - U_B^{*}(\theta_2) > 2U_A^{**}(\theta_2) - U_B^{**}(\theta_2) \Leftrightarrow$$
$$\left(U_B^{**}(\theta_2) - U_B^{*}(\theta_2)\right) - 2\left(U_A^{**}(\theta_2) - U_A^{*}(\theta_2)\right) > 0 \tag{A.41}$$

Moving to firm B, as in Regime 1 we have that $x_B^{**}(\theta_1) > \frac{k_B}{\theta_1} = x_B^{fb}(\theta_1)$ and that $x_B^{**}(\theta_2) = \frac{k_B}{\theta_2} = x_B^{fb}(\theta_2)$ so that the effort of high-types is upward distorted while the effort of low-types is set at the efficient level.

From FOCs (A.22) and (A.23), one has:

$$\lambda_1 \frac{k_B x_B(\theta_1) - \frac{1}{2}\theta_1 x_B^2(\theta_1) - U_B(\theta_1)}{2(\sigma - \alpha_1)} - \lambda_1 \frac{\sigma - \alpha_1 - U_A(\theta_1) + U_B(\theta_1)}{2(\sigma - \alpha_1)} - \mu_{U_B} = 0. \tag{A.42}$$

and

$$\lambda_2 \frac{k_B x_B(\theta_2) - \frac{1}{2}\theta_2 x_B^2(\theta_2) - U_B(\theta_2)}{2\sigma} - \lambda_2 \frac{\sigma - U_A(\theta_2) + U_B(\theta_2)}{2\sigma} + \mu_{U_B} = 0. \tag{A.43}$$

The surplus is $S_B(\theta_j) \equiv k_B x_B(\theta_j) - \frac{1}{2}\theta_j x_B^2(\theta_j)$, with $S_B^{**}(\theta_1) < S_B^{fb}(\theta_1)$ and $S_B^{**}(\theta_2) = S_B^{fb}(\theta_2)$ because $x_B^{**}(\theta_1)$ is distorted whereas $x_B^{**}(\theta_2)$ is at the efficient level. Substituting and rearranging the two previous FOCs, they respectively become:

$$S_B^{**}(\theta_1) + U_A^{**}(\theta_1) - 2U_B^{**}(\theta_1) - \sigma + \alpha_1 > 0 \tag{A.44}$$

and

$$S_B^{fb}(\theta_2) + U_A^{**}(\theta_2) - 2U_B^{**}(\theta_2) - \sigma < 0 \tag{A.45}$$

In the unconstrained program of firm B, the previous two FOCs can be written as:

$$S_B^{fb}(\theta_1) + U_A^{*}(\theta_1) - 2U_B^{*}(\theta_1) - \sigma + \alpha_1 = 0 \tag{A.46}$$

and

$$S_B^{fb}(\theta_2) + U_A^{*}(\theta_2) - 2U_B^{*}(\theta_2) - \sigma = 0. \tag{A.47}$$

As for type θ_1, putting together (A.44) and (A.46) one has:

$$S_B^{fb}(\theta_1) = -U_A^{*}(\theta_1) + 2U_B^{*}(\theta_1) + \sigma - \alpha_1 >$$
$$S_B^{**}(\theta_1) > -U_A^{**}(\theta_1) + 2U_B^{**}(\theta_1) + \sigma - \alpha_1$$

or

$$2U_B^{*}(\theta_1) - U_A^{*}(\theta_1) > 2U_B^{**}(\theta_1) - U_A^{**}(\theta_1) \Leftrightarrow$$
$$(U_A^{**}(\theta_1) - U_A^{*}(\theta_1)) - 2(U_B^{**}(\theta_1) - U_B^{*}(\theta_1)) > 0 \tag{A.48}$$

As for type θ_2, from (A.45) and (A.47) one can write:

$$S_B^{fb}(\theta_2) = -U_A^{*}(\theta_2) + 2U_B^{*}(\theta_2) + \sigma =$$
$$S_B^{**}(\theta_2) < -U_A^{**}(\theta_2) + 2U_B^{**}(\theta_2) + \sigma$$

or

$$2U_B^{*}(\theta_2) - U_A^{*}(\theta_2) < 2U_B^{**}(\theta_2) - U_A^{**}(\theta_2) \Leftrightarrow$$
$$2(U_B^{**}(\theta_2) - U_B^{*}(\theta_2)) - (U_A^{**}(\theta_2) - U_A^{*}(\theta_2)) > 0 \tag{A.49}$$

Now, for high-types consider (A.40) and (A.48) which together imply:

$$U_A^{**}(\theta_1) - U_B^{**}(\theta_1) > U_A^{*}(\theta_1) - U_B^{*}(\theta_1)$$

meaning that $\gamma_1^{**} > \gamma_1^{*}$. As for low-types, (A.41) and (A.49) together imply:

$$U_A^{*}(\theta_2) - U_B^{*}(\theta_2) > U_A^{**}(\theta_2) - U_B^{**}(\theta_2)$$

or $\gamma_2^{**} < \gamma_2^{*}$.

A.7.1 Numerical simulations under Regime 1

To show that all the neglected conditions and the omitted constraints are met, let us consider the following parameter values: $k_A = 6.5$, $k_B = 4$, $\theta_1 = 1$, $\theta_2 = 1.2$, $\sigma = 6$, $\alpha_1 = 0.12$ and $\lambda_1 = \lambda_2 = 0.5$. Again these values assure that the solution is interior in the case of full information on ability (see conditions $k_A^2 - k_B^2 < 6\theta_1(\sigma - \alpha_1)$ and $k_A^2 - k_B^2 < 6\theta_2\sigma$ in Proposition 1); in addition they now satisfy conditions $\frac{\theta_2-\theta_1}{\theta_1} \leq \frac{k_A^2-k_B^2}{3k_A^2}$ and $0 < \alpha_1 \leq \alpha_1^b = 0.14$ of Regime 1. Under screening we obtain the following contracts: $\{x_A^{**}(\theta_1), U_A^{**}(\theta_1)\} = (6.5, 10.866)$; $\{x_A^{**}(\theta_2), U_A^{**}(\theta_2)\} = (5.397, 7.952)$; $\{x_B^{**}(\theta_1), U_B^{**}(\theta_1)\} = (4.238, 6.294)$; $\{x_B^{**}(\theta_2), U_B^{**}(\theta_2)\} = (3.333, 4.497)$; $\hat{\gamma}_1^{**} = 0.888$ and $\hat{\gamma}_2^{**} = 0.787$. Now that also DIC_A is binding, both $x_A^{**}(\theta_2)$ and $U_A^{**}(\theta_2)$ decrease with respect to optimal contracts under Regime 2 and, as a result of the following adjustments in workers' rents, marginal types slightly decrease with respect to before. One can check that the difference $\hat{\gamma}_1 - \hat{\gamma}_2$ is the lowest under full information and the largest under Regime 2 where only UIC_B is binding.

Also under Regime 2 marginal workers' utilities are above zero showing that the workers' participation constraints are slack. All profits per-worker are strictly positive and $0 < \pi_B(\theta_1) < \pi_B(\theta_2)$ and $0 < \pi_A(\theta_2) < \pi_A(\theta_1)$ hold.

A.8 Regime 3: UIC_A and UIC_B are binding

Let us start from firm A which now solves the program where DIC_A is slack while UIC_A is binding so that $\mu_{D_A} = 0$ while $\mu_{U_A} > 0$. From (A.9) and (A.10) we now have that $x_A^{**}(\theta_2) = \frac{k_A}{\theta_2} = x_A^{fb}(\theta_2)$ while the following equation holds for $x_A(\theta_1)$:

$$\lambda_1 (k_A - \theta_1 x_A(\theta_1)) \frac{U_A(\theta_1) - U_B(\theta_1) + \sigma - \alpha_1}{2(\sigma - \alpha_1)} + \mu_{U_A}(\theta_2 - \theta_1) x_A(\theta_1) = 0$$

or

$$\lambda_1 (k_A - \theta_1 x_A(\theta_1)) \frac{U_A(\theta_1) - U_B(\theta_1) + \sigma - \alpha_1}{2(\sigma - \alpha_1)} < 0$$

where $\frac{U_A(\theta_1)-U_B(\theta_1)+\sigma-\alpha_1}{2(\sigma-\alpha_1)} = \hat{\gamma}_1 > 0$ because we have an interior solution. Hence it must be that $x_A^{**}(\theta_1) > \frac{k_A}{\theta_2} = x_A^{fb}(\theta_1)$ meaning that the effort of high-types is upward distorted.

Rearranging (A.11) and (A.12) with $\mu_{D_A} = 0$ while $\mu_{U_A} > 0$ and substituting for the surpluses $S_A^{**}(\theta_1) < S_A^{fb}(\theta_1)$ and $S_A^{**}(\theta_2) = S_A^{fb}(\theta_2)$ one finds:

$$S_A^{**}(\theta_1) - 2U_A^{**}(\theta_1) + U_B^{**}(\theta_1) - \sigma + \alpha_1 > 0 \tag{A.50}$$

and

$$S_A^{fb}(\theta_2) - 2U_A^{**}(\theta_2) + U_B^{**}(\theta_2) - \sigma < 0 \tag{A.51}$$

In the unconstrained program of firm A where $\mu_{D_B} = \mu_{U_B} = 0$, the previous FOCs can be respectively written as in (A.38) and (A.39).

As for type θ_1, putting together (A.50) and (A.38) one has:

$$S_A^{fb}(\theta_1) = 2U_A^{*}(\theta_1) - U_B^{*}(\theta_1) + \sigma - \alpha_1 >$$
$$S_A^{**}(\theta_1) > 2U_A^{**}(\theta_1) - U_B^{**}(\theta_1) + \sigma - \alpha_1$$

or

$$2U_A^{*}(\theta_1) - U_B^{*}(\theta_1) > 2U_A^{**}(\theta_1) - U_B^{**}(\theta_1) \Leftrightarrow$$
$$\left(U_B^{**}(\theta_1) - U_B^{*}(\theta_1)\right) - 2\left(U_A^{**}(\theta_1) - U_A^{*}(\theta_1)\right) > 0 \tag{A.52}$$

As for type θ_2, from (A.51) and (A.39) one can write:

$$S_A^{fb}(\theta_2) = 2U_A^{*}(\theta_2) - U_B^{*}(\theta_2) + \sigma =$$
$$S_A^{fb}(\theta_2) < 2U_A^{**}(\theta_2) - U_B^{**}(\theta_2) + \sigma$$

or

$$2U_A^{*}(\theta_2) - U_B^{*}(\theta_2) < 2U_A^{**}(\theta_2) - U_B^{**}(\theta_2) \Leftrightarrow$$
$$2\left(U_A^{**}(\theta_2) - U_A^{*}(\theta_2)\right) - \left(U_B^{**}(\theta_2) - U_B^{*}(\theta_2)\right) > 0 \tag{A.53}$$

Inequalities (A.48) and (A.52) together write

$$\left(U_B^{**}(\theta_1) - U_B^{*}(\theta_1)\right) - 2\left(U_A^{**}(\theta_1) - U_A^{*}(\theta_1)\right) > 0$$
$$\left(U_A^{**}(\theta_1) - U_A^{*}(\theta_1)\right) - 2\left(U_B^{**}(\theta_1) - U_B^{*}(\theta_1)\right) > 0$$

which lead to a contradiction unless $U_B^{**}(\theta_1) - U_B^{*}(\theta_1) = U_A^{**}(\theta_1) - U_A^{*}(\theta_1) < 0$. Hence, it must be $U_A^{**}(\theta_1) - U_B^{**}(\theta_1) = U_A^{*}(\theta_1) - U_B^{*}(\theta_1)$. Using (18) and (19) one has that $U_A^{**}(\theta_1) - U_B^{**}(\theta_1)$ $= U_A^{*}(\theta_1) - U_B^{*}(\theta_1) = \frac{k_A^2 - k_B^2}{6\theta_1} \geq 0$. Which implies that $\hat{\gamma}_1^{**} = \hat{\gamma}_1^{*} = \frac{1}{2} + \frac{k_A^2 - k_B^2}{12\theta_1(\sigma - \alpha_1)}$. To sum up, one has:

$$
\begin{aligned}
U_B^{**}(\theta_1) - U_B^*(\theta_1) &= U_A^{**}(\theta_1) - U_A^*(\theta_1) \\
U_B^{**}(\theta_1) &< U_B^*(\theta_1) \text{ and } U_A^{**}(\theta_1) < U_A^*(\theta_1) \\
U_A^{**}(\theta_1) - U_B^{**}(\theta_1) &= U_A^*(\theta_1) - U_B^*(\theta_1) = \frac{k_A^2 - k_B^2}{6\theta_1} \\
\hat{\gamma}_1^{**} &= \hat{\gamma}_1^* = \frac{1}{2} + \frac{k_A^2 - k_B^2}{12\theta_1(\sigma - \alpha_1)}.
\end{aligned}
$$

Repeating the same reasoning for types θ_2 one has that inequalities (A.49) and (A.53) together write

$$
\begin{aligned}
2\left(U_A^{**}(\theta_2) - U_A^*(\theta_2)\right) - \left(U_B^{**}(\theta_2) - U_B^*(\theta_2)\right) &> 0 \\
2\left(U_B^{**}(\theta_2) - U_B^*(\theta_2)\right) - \left(U_A^{**}(\theta_2) - U_A^*(\theta_2)\right) &> 0
\end{aligned}
$$

which lead to a contradiction unless $U_A^{**}(\theta_2) - U_A^*(\theta_2) = U_B^{**}(\theta_2) - U_B^*(\theta_2) > 0$. Hence, using (20) and (21) and rearranging, $U_A^{**}(\theta_2) - U_B^{**}(\theta_2) = U_A^*(\theta_2) - U_B^*(\theta_2) = \frac{k_A^2 - k_B^2}{6\theta_2} \geq 0$ must hold. Which implies that $\hat{\gamma}_2^{**} = \hat{\gamma}_2^* = \frac{1}{2} + \frac{k_A^2 - k_B^2}{12\sigma\theta_2}$. Hence, the following holds:

$$
\begin{aligned}
U_A^{**}(\theta_2) - U_A^*(\theta_2) &= U_B^{**}(\theta_2) - U_B^*(\theta_2) \\
U_A^{**}(\theta_2) &> U_A^*(\theta_2) \text{ and } U_B^{**}(\theta_2) > U_B^*(\theta_2) \\
U_A^{**}(\theta_2) - U_B^{**}(\theta_2) &= U_A^*(\theta_2) - U_B^*(\theta_2) = \frac{k_A^2 - k_B^2}{6\theta_2} \\
\hat{\gamma}_2^{**} &= \hat{\gamma}_2^* = \frac{1}{2} + \frac{k_A^2 - k_B^2}{12\theta_2\sigma}.
\end{aligned}
$$

We can conclude that workers' sorting is not affected by incentive constraints and $\hat{\gamma}_1^{**} = \hat{\gamma}_1^* \geq \hat{\gamma}_2^{**} = \hat{\gamma}_2^*$. If the two firms are identical then they equally share the workforce of each type $\hat{\gamma}_1^{**} = \hat{\gamma}_1^* = \hat{\gamma}_2^{**} = \hat{\gamma}_2^* = \frac{1}{2}$. Type-$\theta_1$ workers and the two firms are worse off while type-θ_2 workers are better off with respect to the market equilibrium under full information on ability.

A.8.1 Numerical simulations under Regime 3

Let us focus on a symmetric equilibrium because this is the unique regime that is compatible with the firms being identical and thus equally sharing the market. Consider the following parameters: $k_A = k_B = 5$, $\theta_1 = 1$, $\theta_2 = 0.2$, $\sigma = 3$, $\alpha_1 = 0.5$ and $\lambda_1 = \lambda_2 = 0.5$. Again these values assure that the solution is interior in the case of full information on ability (see conditions $k_A^2 - k_B^2 < 6\theta_1(\sigma - \alpha_1)$ and $k_A^2 - k_B^2 < 6\theta_2\sigma$ in Proposition 1); in addition the chosen parameters now satisfy conditions $\frac{\theta_2 - \theta_1}{\theta_1} \geq \frac{k_A^2 - k_B^2}{3k_A^2}$ and $\alpha_1 > \alpha_1^c = 0.08$ of Regime 3. Under full information, optimal contracts are: $\{x_A^*(\theta_1), U_A^*(\theta_1)\} = \{x_B^*(\theta_1), U_B^*(\theta_1)\} = (5, 10)$; $\{x_A^*(\theta_2), U_A^*(\theta_2)\} = \{x_B^*(\theta_2), U_B^*(\theta_2)\} = (4.17, 7.42)$, entailing $\hat{\gamma}_1^* = 0.5$ and $\hat{\gamma}_2^* = 0.5$.

Under screening, the effort of high-types is upward distorted whereas indirect utilities of high-types fall and the ones of low-types increase. In line with the theoretical predictions one obtains the following screening contracts: $\{x_A^{**}(\theta_1), U_A^{**}(\theta_1)\} = \{x_B^{**}(\theta_1), U_B^{**}(\theta_1)\} = (5.03, 9.97)$; $\{x_A^{**}(\theta_2), U_A^{**}(\theta_2)\} = \{x_B^{**}(\theta_2), U_B^{**}(\theta_2)\} = (4.17, 7.45)$; $\hat{\gamma}_1^{**} = 0.5$ and $\hat{\gamma}_2^{**} = 0.5$. The marginal workers' utilities are all above zero showing that the workers' participation constraints are slack and profits per-workers are strictly positive. Finally, condition $\frac{\sigma}{\sigma-\alpha_1} > \frac{\pi_A(\theta_2)}{\pi_A(\theta_1)} > 1$, necessary for Regime 3 to hold (see Corollary 4), is met.

A.9 A richer specification of the workers' utility function

Suppose now that both high and low-ability workers are concerned with coworkers' ability. The workers' utility functions write:

$$u_A(x_A, w_A; \theta_j, \gamma) = w_A(x_A) - \frac{1}{2}\theta_j x_A^2 - \gamma\sigma + \alpha_j(\hat{\gamma}_1 - \hat{\gamma}_2), \tag{A.54}$$

$$u_B(x_B, w_B; \theta_j, \gamma) = \underbrace{w_B(x_B) - \frac{1}{2}\theta_j x_B^2}_{\text{net compensation } U_i(\theta_j)} \underbrace{-(1-\gamma)\sigma}_{\text{mismatch disutility}} \underbrace{-\alpha_j(\hat{\gamma}_1 - \hat{\gamma}_2).}_{\text{concern for coworkers' ability}} \tag{A.55}$$

A first difference from before it that $\alpha_2 > 0$ meaning that, low-types also care for coworkers' ability. A second difference is that the low-type marginal worker, $\hat{\gamma}_2$, enters the premium for coworkers' ability too. Thus, CfCA here translates in a premium for the worker only if her employer is able to hire a larger share of high-ability than of low-ability workers $(\hat{\gamma}_1 - \hat{\gamma}_2 > 0)$ or if the worker belongs to the workforce characterized by the higher average ability. Instead, in the reduced model, the concern for coworkers' ability only depends on the relative share of high types. It may be natural to assume that $\alpha_1 \geq \alpha_2 > 0$: CfCA is lower for low-ability workers, for example because they care less for the "social status" of their firm or because they have less career opportunity outside the firm.

Marginal workers at the interior solution are defined as follows:

$$\widehat{\gamma}_1 = \frac{1}{2} + \frac{(\sigma + \alpha_2)(U_A(\theta_1) - U_B(\theta_1)) - \alpha_1(U_A(\theta_2) - U_B(\theta_2))}{2\sigma(\sigma - \alpha_1 + \alpha_2)} \tag{A.56}$$

$$\widehat{\gamma}_2 = \frac{1}{2} + \frac{(\sigma - \alpha_1)(U_A(\theta_2) - U_B(\theta_2)) + \alpha_2(U_A(\theta_1) - U_B(\theta_1))}{2\sigma(\sigma - \alpha_1 + \alpha_2)} \tag{A.57}$$

The SOCs require again that $\sigma - \alpha_1 > 0$ which we continue to assume.[15] Here again CfCA is not so strong to reverse the standard Hotelling "forces".

[15] The objective of this section is to compare the market equilibrium allocation obtained in the reduced model with the one obtained with this richer specification. Hence we disregard here both the welfare analysis and the SOCs required to obtain the efficient allocation.

Expressions (A.56) and (A.57) show that marginal workers are now interdependent: $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ depend both on $U_i(\theta_j)$, $i = A, B$, $j = 1, 2$ and on α_j, $j = 1, 2$. Comparing those expressions with (7) and (8) one observes the following. The difference $(U_A(\theta_1) - U_B(\theta_1))$ now enters both $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ and with a positive sign because, if that difference increases, more high-types are attracted in firm A and this accrues the premium for coworkers' ability that high and low-types employed by firm A receive via the parameters α_1 and α_2, respectively. In addition, both $\widehat{\gamma}_2$ and $\widehat{\gamma}_1$ are decreasing in the difference $(U_A(\theta_2) - U_B(\theta_2))$ because, if that difference increases, a larger share of low-types are attracted in firm A and this negatively affects the premium received by its employees via the parameters α_1 and α_2, respectively.

Even if expressions are richer, we show below that, with this specification, the basic intuitions and results are confirmed. To start with, Remark 2 continues to hold and, at the interior solution, it continues to be true that firm A is able to hire a better workforce only if it offers high-types a higher return to ability than its competitor.

A.9.1 Equilibrium contracts when workers' ability is observable

Plugging (A.56) and (A.57) together with first-best effort levels (13) into P_i, we solve firms' programs for indirect utilities. Here we can no longer solve two separated programs for low and high-ability workers because the labor demands of the two firms include indirect utilities of both workers' type. We derive firm i's expected profits with respect to $U_i(\theta_j)$, $j = 1, 2$, by taking $U_{-i}(\theta_j)$ as given and we find two reaction functions for each firm in which indirect utilities offered by one firm are a function of $U_i(\theta_{-j})$ and of the ones offered by the rival firm. Then we solve the system of the four reaction functions in four unknowns and find indirect utilities at the equilibrium: $U_i^e(\theta_i)$, $i = A, B$ and $j = 1, 2$, where the superscript e indicates 'extended' model. Substituting $U_i^e(\theta_i)$ in (A.56) and (A.57) we then obtain the expressions for marginal workers.

We do not report here expressions of indirect utilities because they are long and tedious. By plugging them into the expressions of marginal workers' utilities and computing their derivatives with respect to α_1 and α_2, one can check the following: the utility of high-type workers is increasing in both α_1 and α_2, whereas the utility of low-types is decreasing in both α_1 and α_2.

The expressions for marginal workers write:

$$\widehat{\gamma}_1^e = \frac{1}{2} + \frac{\left(k_A^2 - k_B^2\right)\left(\alpha_1\theta_1\lambda_1\lambda_2 - 3(\sigma+\alpha_2)\theta_2\lambda_1\lambda_2 - 2\alpha_2\theta_1\lambda_2^2\right)}{4\theta_1\theta_2[2\left(\alpha_1\lambda_1 + \alpha_2\lambda_2\right)^2 - 9\sigma\lambda_1\lambda_2(\sigma - \alpha_1 + \alpha_2)]} \tag{A.58}$$

$$\widehat{\gamma}_2^e = \frac{1}{2} + \frac{\left(k_A^2 - k_B^2\right)\left(2\alpha_1\theta_2\lambda_1^2 - 3(\sigma-\alpha_1)\theta_1\lambda_1\lambda_2 - \alpha_2\theta_2\lambda_1\lambda_2\right)}{4\theta_1\theta_2[2\left(\alpha_1\lambda_1 + \alpha_2\lambda_2\right)^2 - 9\sigma\lambda_1\lambda_2(\sigma - \alpha_1 + \alpha_2)]}. \tag{A.59}$$

Both (A.58) and (A.59) depend on α_1 and α_2. The difference between marginal workers is:

$$\widehat{\gamma}_1^e - \widehat{\gamma}_2^e \geq 0 \iff -\frac{\left(k_A^2 - k_B^2\right)(2(\alpha_1\lambda_1+\alpha_2\lambda_2)(\theta_1\lambda_2+\lambda_1\theta_2)+3\sigma\lambda_1\lambda_2(\theta_2-\theta_1))}{4\theta_1\theta_2[2(\alpha_1\lambda_1+\alpha_2\lambda_2)^2-9\sigma\lambda_1\lambda_2(\sigma-\alpha_1+\alpha_2)]} \geq 0. \tag{C1A}$$

Note that the denominator of the second terms of (A.58)-(A.59) and of Condition $C1A$ is the same. It goes to zero when $\alpha_1 = \tilde{\alpha}_1$, where

$$0 < \tilde{\alpha}_1 \equiv \frac{-\lambda_2(4\alpha_2 + 9\sigma) + 3\sqrt{\lambda_2\sigma(8\alpha_2 + (9 - \lambda_1)\sigma)}}{4\lambda_1} < \sigma.$$

In words, the functions $\widehat{\gamma}_1^e$, $\widehat{\gamma}_2^e$ and $(\widehat{\gamma}_1^e - \widehat{\gamma}_2^e)$ have the same vertical asymptote $\tilde{\alpha}_1$, with $0 < \tilde{\alpha}_1 < \sigma$. Studying $\widehat{\gamma}_1^e$, $\widehat{\gamma}_2^e$ and $(\widehat{\gamma}_1^e - \widehat{\gamma}_2^e)$ as a function of α_1 one can check that $\widehat{\gamma}_1^e$ and $(\widehat{\gamma}_1^e - \widehat{\gamma}_2^e)$ are monotonically increasing in α_1 and tend to infinity for $\alpha_1 \to \tilde{\alpha}_1$. Except for very low values of λ_1, $\widehat{\gamma}_2^e$ is instead monotonically decreasing in α_1 and tends to minus infinity for $\alpha_1 \to \tilde{\alpha}_1$; see Figure 2.

Even if expressions are here more complex, the pattern of $\widehat{\gamma}_1^e$ is the same as in the reduced model. When $k_A > k_B$ but firms' heterogeneity is not too high, starting from a value of α_1 close to zero and letting α_1 grow larger, an interior solution where $\widehat{\gamma}_1^e < 1$ first exists. Then $\widehat{\gamma}_1^e$ increases with α_1 and hits the corner solution $\widehat{\gamma}_1^e = 1$ on the left of $\tilde{\alpha}_1$ and remain 1 hereafter. The novelty here is $\widehat{\gamma}_2^e$ which now depends on α_1 as well. As mentioned before, except for extreme values of the distribution of θ, $\widehat{\gamma}_2^e$ decreases with α_1 and a corner solution such that $\widehat{\gamma}_2^e = 0$ is reached on the left of $\tilde{\alpha}_1$. As an intuition, the higher the concern for coworkers' ability and the larger the benefit from having more workers hired in firm A than in firm B because the share of workers enjoying the premium for coworkers' ability increases while the share of workers suffering the loss decreases. Hence an increase in α_1 has an effect that is similar to an increase in $k_A - k_B$, they both pushes towards a corner solution with $\widehat{\gamma}_1^e = 1$. In addition, whenever $\alpha_1 > \alpha_2$ the premium for high-types employed in firm A is larger than the loss suffered by low-types employed in firm B, hence as α_1 increases, the allocation also tends to a corner solution with $\widehat{\gamma}_2^e = 0$. Interestingly, when sorting is such that $\widehat{\gamma}_1^e = 1$ and $\widehat{\gamma}_2^e = 0$, a full market segmentation emerges with all high-types employed in the more efficient firm A and all low-types hired in firm B. A full market segmentation does not occur when $\alpha_1 = \alpha_2$ unless σ is very low (but, given the condition $\alpha_1 < \sigma$, this implies that α_1 is very low as well). In this case an interior solution with $\frac{1}{2} < \widehat{\gamma}_1^e < \widehat{\gamma}_2^e < 1$ is more likely.

The following proposition summarizes the main results for this specification of the model.

Proposition 6 ***Full information on ability (extended model).*** *(i) When ability is observable while mismatch disutility is the workers' private information, equilibrium contracts are the Nash equilibrium contracts* $\{x_i^e(\theta_j), U_i^e(\theta_j)\}_{j=1,2;\ i=A,B}$ *of the game in which firms compete in the utility space and are defined by efficient efforts* $x_i^e(\theta_j) = x_i^*(\theta_j)$*.*

(ii) When firms are identical ($k_A = k_B$) *they equally share the workforce of both types:* $\widehat{\gamma}_1^e = \widehat{\gamma}_2^e = \frac{1}{2}$ *and* $E_A^e(\theta) = E_B^e(\theta)$*.*

(iii) The problem is concave for $\alpha_1 < \sigma$*. The high-type marginal worker,* $\widehat{\gamma}_1^e$*, and the difference between marginal types* $(\widehat{\gamma}_1^e - \widehat{\gamma}_2^e)$ *are both increasing in* α_1*. Except for very low values of* λ_1*,* $\widehat{\gamma}_2^e$

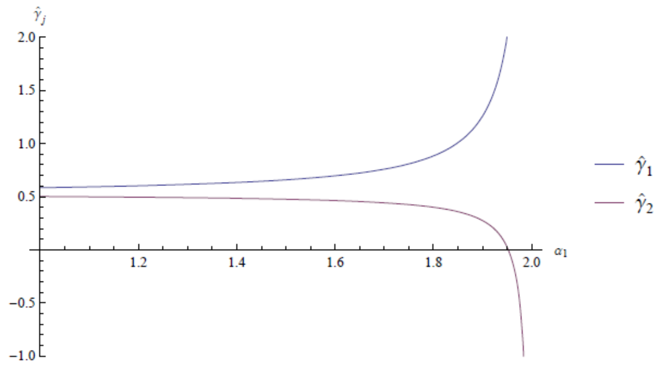

Figure 2: Market equilibrium in the extended model. Marginal workers $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ as a function of α_1 are shown when $\alpha_2 = \theta_1 = \Delta k = 1$, $\lambda_1 = 1/2$, $\sigma = 2$ and $\theta_2 = 2.5$.

is decreasing in α_1.
(iv) When $k_A > k_B$*, then firm A hires a larger share of both types and the better workforce:* $\widehat{\gamma}_1^e > \widehat{\gamma}_2^e > \frac{1}{2}$ *and* $E_A^e(\theta) < E_B^e(\theta)$.
(v) If $(k_A - k_B)$ *is sufficiently high and/or* α_1 *is sufficiently closer to* $\tilde{\alpha}_1$*, then a corner solution with* $\widehat{\gamma}_1^e = 1$ *for high-types and a corner solution with* $\widehat{\gamma}_2^e = 0$ *for low-types emerge. Corner solutions are less likely when* $\alpha_1 = \alpha_2$.
(vi) High-type's utility increases with both α_1 *and* α_2*. Low-type's utility decreases with both* α_1 *and* α_2.

Overall we can conclude that the reduced model analyzed in the main text is able to capture the main effects of CfCA on market equilibrium and on workers' sorting. When $\alpha_2 > 0$, CfCA empowers high-type workers via both α_1 and α_2. Low types, who where not affected by CfCA when $\alpha_2 = 0$, are impaired by both α_1 and α_2 when $\alpha_2 > 0$. In addition, CfCA may lead to a market equilibrium with full market segmentation.