WORKING PAPERS

N° 1257

Revised version
September 2023

"Shining with the stars: competition, screening, and concern for coworkers' quality"

Francesca Barigozzi and Helmuth Cremer

Toulouse School of Economics

Shining with the stars: competition, screening, and concern for coworkers' quality *

Francesca Barigozzi† Helmuth Cremer‡

March 2022, revised September 2023

Abstract

We study how workers' concern for coworkers' ability (CfCA) affects competition in the labor market. Two firms offer 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 their taste for being employed by any of the two firms. Workers receive a utility premium when employed by the firm hiring most high-ability workers and suffer a utility loss if hired by its competitor. These premiums/losses are endogenously determined.

We characterize contracts and workers' sorting into the two firms under complete and private information on workers' ability. We show that CfCA is detrimental to firms, but it benefits high-ability workers, especially when their ability is observable. In addition, CfCA exacerbates the existing distortion in high-ability workers' sorting into the two firms.

Keywords: Concern for Coworkers' Quality, Competition, Screening, Sorting.
***JEL*-Classification:** D82, L13, M54.

*We thank the Advisory Editor and the reviewers for their constructive and thoughtful comments. 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).

†University of Bologna, Italy, Email: francesca.barigozzi@unibo.it.
‡Toulouse School of Economics, France, Email: helmuth.cremer@tse-fr.eu.

1 Introduction

Consider newly graduated lawyers from a prestigious Law School facing a choice between applying for a position at either Cravath or Skadden, two renowned law firms. Their decision will naturally depend on factors like offered salaries and the amenities provided by each firm. However, it's also likely to be influenced by the number of high-profile "star" lawyers employed by each firm. The traditional economic literature focuses on job decisions based on applicants' preferences for the organization, and the associated monetary compensation. Our innovation lies in the assumption that workers' choices is also influenced by the quality of their coworkers. Specifically, we propose that an organization's attractiveness increases as it hires more top-tier employees, in other words, a higher number of "star" workers.

Why is workers' utility increasing in their coworkers' quality? First, working with top professionals may give preferential access to resources, opportunities, and general perks/benefits inside and outside the organization.[1] Second, top workers bring social status and reputation to the organization, and the latter may be a source of utility *per se* for coworkers.[2] Third, workers' utility may be increasing in the measure of high-ability coworkers because being employed in an organization that hires a qualified workforce increases the workers' future career prospects outside the firm.[3] Note that we disregard complementarities and possible spillovers in term of productivity, which have been considered before.[4]

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

[2] For example, though borrowed from a very selective market, the reputation of a football team is increasing in direct relation to the number of top-tier players it possesses. This correlation between the team's reputation and the presence of elite players subsequently elevates the social status of all other members within the team. Take, for example, FC Barcelona during the 2010–2011 season, which boasted three high-caliber players: Lionel Messi, Andrés Iniesta, and Xavi Hernández. These renowned players not only enhanced the team's overall standing but also positively influenced the social standing of their fellow teammates. Another example from a very different market is provided by best Masters Programs, an independent online guide that offers a ranking of the first 50 universities in the USA and Europe based on the total number of Nobel Prize winners (ranging from 12 to 151), including both faculty and alumni. They state: "A university that has produced a large quantity of Nobel Prize winners has done so due to its ability to hold its students and faculty to rigorous standards, encourage their peoples' talents, and offer facilities that allow students and faculty to explore and expand upon their theories. A supportive, well-funded institution gives its faculty and students the capacity to accomplish much more than a lower-level school, which in turn increases its ability to support and fund research. It's a cycle that rises higher and higher. (See https://www.bestmastersprograms.org, accessed in September 2023)

[3] Dustmann et al. (2016) and Glitz (2017) study peer effects in job search among former coworkers and find evidence that coworkers' networks help to reduce informational frictions in the labor market and lead to gains for workers and firms.

[4] See, among others, Au and Chen (2021), Lindquist et al. (2022) in the Economic Literature and Groysberg and Lee (2008), Ertug et al. (2018), Tan and Netessine (2019) in the Management Literature.

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 workers employed in the organization hiring most 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.[5] Specifically, they experience greater utility when employed by the firm that attracts the larger share of high-ability candidates from the labor market, while their utility decreases when they work for the firm that attracts the smaller share of high-ability candidates. Firms use nonlinear contracts to compete for workers. Optimal contracts are contingent on workers' ability and are designed in the *utility space* so that they are characterized by the (gross) indirect utility offered to the worker, i.e. his/her rent, 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. With some abuse, we interpret this case as representing a "senior" job-market where candidates' previous outcomes are observable (e.g. successful lawsuits for a lawyer and the publications list for a researcher). Here, equilibrium contracts entail efficient effort levels. 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-ability workers and, to a lower extent, also a larger share of low-ability workers. 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

[5] 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.

high-ability workers.

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 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. The case of private information on ability can be interpreted as a junior job-market where applicants had no opportunities yet to prove their talent in practice. We show that, if the two firms are identical and CfCA is sufficiently low, then the market allocation is incentive compatible. In case of positive but small difference in the two firms' productivities, the market allocation continues to be incentive compatible when CfCA is sufficiently low and workers' heterogeneity is large. Otherwise, the market allocation is not incentive compatible and, depending on which incentive constraints are binding, one of three different regimes emerges. Recall that, as the value of CfCA increases, firms compete more intensely for talented workers. The less efficient firm faces incentive constraints sooner, even for lower CfCA values, and must overincentivize high-skilled workers in all three regimes. On the other hand, the more efficient firm starts by underincentivizing low-skilled workers in Regime 1. As CfCA increases, the efficient firm becomes unconstrained in Regime 2, but it still needs to overincentivize high-ability workers like the less efficient firm in Regime 3.

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 (and thus between wages) that firms offer to the workers. Specifically, in Regime 1 and 2, the share of high-ability workers 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-ability 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 and wages. Low-ability workers on the other hand are better off because their utility increases. In Regime 1, results are ambiguous in this respect.

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, overincentivization of talented workers (in the form of countervailing incentives) partially erodes the additional surplus appropriated by high-ability workers in the full information equilibrium, and the more so when CfCA is high.

Recall that full information on ability can be interpreted as selection of senior job market candidates, while private information on ability might correspond to selection of junior candidates. The model delivers the following general results. CfCA empowers *senior* talented job market applicants, including the ones employed by the least efficient firm. However, *junior* talented applicants entering the labor market for the first time are not able to appropriate all the surplus from CfCA. Such surplus is substantially eroded by screening contracts which imply lower rent for and overincentivization of talented workers. This suggests that, in a labor market where CfCA increases competition for talented workers, employers struggle to balance the benefits and costs of high-powered incentives, notably in the case of junior job market applicants.[6]

As mentioned before, in the main text we study a specification of the model where only high-ability workers' utility increases with the measure of high-ability coworkers. This specification is tractable and intuitions are easy to grasp. In Appendix A.12, we study the market equilibrium and its welfare properties in an extension of the model with a richer specification of CfCA. Specifically, both high- and low-ability workers are concerned with coworkers' ability and the premium/disutility for CfCA depends on the measure of high- and low-ability coworkers. We show that our reduced-form model is able to capture the results on market equilibrium and on workers' sorting obtained with the richer specifications.

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

[6] This in turn is consistent with the observation that young talented workers in professional services suffer because of stress and burnout. According to Deloitte's 2015 external workplace well-being survey, 84 percent of millennials say they have experienced burnout at their current job, compared to 77 percent of all respondents. Nearly half of millennials say they have left a job specifically because they felt burned out, compared to 42 percent of all respondents. See https://www2.deloitte.com/us/en/pages/about-deloitte/articles/burnout-survey.html

the presence of both vertical and horizontal preference uncertainty.[7] 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.[8]

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 dependents upon a worker's ability. When motivation is observable, while ability is 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 workers. 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 upwards and the more efficient firm may or may not distort effort of low-skilled workers downwards.

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.[9] 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

[7] 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.

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

[9] 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.

Burani (2019) and in "taste for firms" in this papers. In the two papers both characteristics are the workers' private information and, in Barigozzi and Burani (2019), intrinsic motivation 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).

Our paper is also related to the matching literature applied to the labor market; seminal papers are Gale and Shapley (1962) and Kelso and Crawford (1982). The labor market is a typical many-to-one matching market where each firm may employ many workers, but each worker works for at most one firm. When workers are not only concerned with the firm they are matched to, but also with the other workers matched to the same firm, peer effects become significant; see Echenique and Yenmez (2007). This results in a more complex preference ordering for each agent and can create challenges for achieving market stability and developing efficient algorithms for identifying stable matchings; see Pycia (2012). Recently, Nax and Pradelski (2016) obtain convergence results for a class of matching markets that mirror key features of a decentralized and dynamic labor market.

Finally, social networks can directly affect job search activities and their outcomes. Empirical evidence indicates that personal contacts play a vital role in facilitating employment opportunities through word-of-mouth, serving as a potential alternative to more formal means of obtaining employment information. Two different theoretical approaches have been proposed to explain this evidence in the economic literature on Networks. Calvó-Armengol and Jackson (2004) model social networks as graphs, and assume that individuals exchange job information only with their strong ties, while weak ties can help them by providing job information only indirectly. In Zenou (2011, 2013), individuals belong to mutually exclusive two-person groups, referred to as dyads and two individuals belonging to the same dyad hold a strong tie to each other. In addition, he assumes that weak ties with people outside the dyad are superior to strong ties for providing support in getting a job.[10]

Our approach to the study of peer effects in the labor market differs significantly from the one used in the matching and network literature. In particular, we explicitly model the payoffs of both firms and workers, and use a noncooperative game-theoretical framework to derive the market equilibrium and its welfare properties. Our contribution lies in the application of multi-principal literature on competition and screening, which analyzes how principals compete in designing nonlinear contracts to attract agents of different types. Our approach is comple-

[10] A recent paper by Bolte et al. (2020) studies the consequences of homophily in the workplace. 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.

mentary to the one used in the matching and network literature. It allows for the understanding of the role of peer effects in the design of contracts with full and private information on workers' ability and sheds light on noncooperative firm-worker matching.

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

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 $k_A \geq k_B$.

Profit *per-worker*, conditional on the worker being hired, is 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 paid by firm i to the worker exerting effort x, with $x \geq 0$. The unit price of output is exogenous and set to 1. A firm's total profit depends on the measure of high and low-ability workers hired by the firm, as we will show below.

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 $\theta_2 > \theta_1$.[11] 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.

[11] Essentially, we have a model wherein workers differ in their marginal product but this marginal product is also firm specific. None of our results would change if rather than differentiating workers by cost of effort we consider different levels of productivity which may be scaled up or down when moving across firms.

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 most high-ability job market candidates and suffer a disutility otherwise. Premiums and losses increase continuously in the share of high-ability candidates hired by the employer.

Take firm A and the share of its high-ability 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 homophily among high-skill workers.[12] In the conclusion we discuss possible job-market mechanisms resulting in a workers' utility function which increases continuously with the quality of their coworkers.[13]

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

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

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

where the relevance of CfCA is represented by the parameter $\alpha_j \geq 0$ that we discuss below. 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-ability workers if their employer is able to hire a larger share of high-ability job market candidates than its competitor and in a utility loss suffered by high-ability workers employed by the firm hiring the lower share of them.[14]

[12] 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.

[13] In the last paragraph of the Conclusion, we discuss how we expect spillover effects on coworkers' productivity to affect our framework and results. Disregarding those effects may be seen as a limitation of the model. However, understanding the impact of star coworkers on market equilibrium and workers sorting in isolation is interesting *per se* and represents a first step in the study of competition for talent when peer effects matter. In addition, it is not clear whether productivity spillovers are positive or negative. For example, Tan and Netessine (2019) study how coworkers' sales ability affects other workers' sales performance in restaurants and find that "you do not necessarily learn to fly if you work with a superman." In a laboratory experiment, van Veldhuizen et al. (2018) study whether worker effort is positively related to the productivity of coworkers who observe them but do not find significant peer effects.

[14] 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.

Note that, when employed by firm B, a high-ability worker's premium for coworkers' ability is $+\alpha_1((1-\hat{\gamma}_1)-1/2) = -\alpha_1(\hat{\gamma}_1 - 1/2)$.

In reality, the importance of peer effects created by high-ability coworkers is likely to be determined by dynamic processes. However, given the complexity of the issue, we restrict ourselves to a static model and assume exogenous α_j.

When $\alpha_1 > \alpha_2 \geq 0$, 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 opportunities outside the firm. If instead $\alpha_2 \geq \alpha_1 > 0$, CfCA is higher for low-ability workers. Here, joining a firm with a high proportion of high-ability workers could be a strong indicator of an individual's ability level, which may hold more value for low-ability workers than for high-ability workers who have alternative methods to demonstrate their competence.

In the main text we analyze a reduced-form model with $\alpha_1 > \alpha_2 = 0$. In Appendix A.12, we present and discuss the richer specification with $\alpha_2 > 0$ and $\alpha_2 \gtreqless \alpha_1$ under full information on ability. At the end of Section 3, we compare market equilibria generated by the reduced-form model and the general model. At the end of Section 4, we compare the welfare properties displayed by the two models.

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.

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: $\mathrm{E}_A(\theta) = \mathrm{E}_B(\theta) = \mathrm{E}(\theta)$.

(*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: $\mathrm{E}_A(\theta) < \mathrm{E}_B(\theta)$.

(*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: $\mathrm{E}_A(\theta) > \mathrm{E}_B(\theta)$.

Workers' utilities (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 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 level 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

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

if employed by firm A and *total* indirect utility

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

if employed by firm B.

The participation constraints require that

$$\mathcal{U}_A(\theta_j, \gamma) \geq 0 \qquad \text{and} \qquad \mathcal{U}_B(\theta_j, \gamma) \geq 0,\ \text{for all } \theta_j \in \{\theta_1, \theta_2\},\ \gamma \in [0,1]. \tag{PC}$$

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 $-i$.[15]

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 $\mathcal{U}_A(\theta_j, \widehat{\gamma}_j) = \mathcal{U}_B(\theta_j, \widehat{\gamma}_j)$, $j = 1, 2$. 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", associated with mismatch costs, 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. When $\alpha_1 > \sigma$, CfCA dominates mismatch cost and there is a corner solution with all high-ability workers employed by one firm. This is in line with intuition, and we therefore relegate the formal proof to Appendix A.1.

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. This section not only provides a reference for the section where costs are not observable, but it is also interesting for its own sake. Roughly speaking, one can think of it as describing a senior job market where asymmetries of information on abilities are likely to be of second order importance. It shows how CfCA combined with competition affects sorting and efficiency. Similarly, we can think of Section 5 below as dealing with junior job markets where asymmetric information on abilities is likely to be more significant. Note, however, that we do not intent to compare junior and senior markets because information is not the only aspect in which they differ. In particular, CfCA is also likely to differ between these job markets.

[15] 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.

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

$$w_i(\theta_j) = U_i(\theta_j) + \frac{1}{2}\theta_j x_i^2(\theta_j). \tag{9}$$

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

$$\pi_i(\theta_j) = k_i x_i(\theta_j) - \frac{1}{2}\theta_j x_i^2(\theta_j) - U_i(\theta_j). \tag{10}$$

Each firm maximizes profits obtained by multiplying (10) with their workforce determined by expressions (7) and (8). Hence, firm A and B respectively solves the following program:

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

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. 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 $U_{-i}(\theta_{-j})$). Hence, firms maximize profits per-worker's for each *type,* and Program P_i can be decomposed into two programs:[16]

$$\begin{aligned}
\max_{\{x_A(\theta_j),U_A(\theta_j)\}} \pi_A(\theta_j) = \; & \left(\tfrac{1}{2} + \widehat{\gamma}_j\right)\left(k_A x_A(\theta_j) - \tfrac{1}{2}\theta_j x_A^2(\theta_j) - U_A(\theta_j)\right) \\
\max_{\{x_B(\theta_j),U_B(\theta_j)\}} \pi_B(\theta_j) = \; & \left(\tfrac{1}{2} - \widehat{\gamma}_j\right)\left(k_B x_B(\theta_j) - \tfrac{1}{2}\theta_j x_B^2(\theta_j) - U_B(\theta_j)\right)
\end{aligned} \tag{P_i'}$$

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

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

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

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

$$x_i^*(\theta_j) = x_i^{fb}(\theta_j) = \frac{k_i}{\theta_j}. \tag{11}$$

[16] See Appendix A.12 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.

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 first-best efforts (11) in firms' Programs P_i', 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. Reaction functions are:

$$U_i(\theta_j) = \frac{k_i^2 - 2\theta_j(\sigma - \alpha_j)}{4\theta_j} + \frac{1}{2}U_{-i}(\theta_j), \; i = A, B; \; j = 1, 2. \tag{12}$$

Expression (12) shows 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:

$$U_i^*(\theta_j) = \frac{2k_i^2 + k_{-i}^2}{6\theta_j} - \sigma + \alpha_j, \; i = A, B; \; j = 1, 2. \tag{13}$$

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 (13) one can also check that CfCA benefits high-ability workers who receive a larger $U_i^*(\theta_1)$ when $\alpha_1 > 0$ than when $\alpha_1 = 0$. Hence, the indirect utility of high-ability workers, independently of the firm hiring them, 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 and their difference:

$$\widehat{\gamma}_1^* = \frac{1}{2} + \frac{k_A^2 - k_B^2}{12\theta_1(\sigma - \alpha_1)} \geq \frac{1}{2}, \tag{14}$$

$$\widehat{\gamma}_2^* = \frac{1}{2} + \frac{k_A^2 - k_B^2}{12\theta_2\sigma} \geq \frac{1}{2}. \tag{15}$$

$$\widehat{\gamma}_1^* - \widehat{\gamma}_2^* = \frac{\left(k_A^2 - k_B^2\right)(2\alpha_1\theta_1 + \sigma(\theta_2 - \theta_1))}{12\theta_1\theta_2\sigma(\sigma - \alpha_1)} \geq 0. \tag{16}$$

Firm A, holding a competitive advantage, hires a larger share of high-ability workers and, to a lower extent, also a larger share of low-ability workers. As a result, the more efficient firm always hires the workforce characterized by the higher average ability. Moreover, marginal worker $\widehat{\gamma}_1^*$ is increasing in α_1.

From (14) and (15), an interior solution with $\frac{1}{2} < \widehat{\gamma}_2^* < \widehat{\gamma}_1^* < 1$ requires that $k_A^2 - k_B^2 < 6\theta_1(\sigma - \alpha_1)$ holds. Intuitively, firm B remains active in the market only if firm A's competitive advantage is not too high relative to workers' mismatch disutility. The condition for an interior $\widehat{\gamma}_1$ can be rewritten as

$$\widehat{\gamma}_1^* < 1 \quad \Leftrightarrow \quad \alpha_1 < \alpha_1' \equiv \sigma - \tfrac{k_A^2 - k_B^2}{6\theta_1}. \tag{17}$$

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 $\alpha_1 \geq \alpha_1'$.

Total utilities $\mathcal{U}_i^*(\theta_j, \widehat{\gamma}_j^*)$ of marginal workers in equilibrium are given by:

$$\mathcal{U}_i^* \left(\theta_j, \widehat{\gamma}_j^*\right) = \mathcal{U}_{-i}^* \left(\theta_j, \widehat{\gamma}_j^*\right) = \frac{k_i^2 + k_{-i}^2 - 6\theta_j \sigma}{4\theta_j} + \alpha_j, \ i = A, B; \ j = 1, 2. \tag{18}$$

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 (18) confirms that high-ability workers' payoff is increasing in the concern for coworkers' quality.

Note that having $U_A^*(\theta_j, \widehat{\gamma}_j^*) = U_B^*(\theta_j, \widehat{\gamma}_j^*) \geq 0$, $j = 1, 2$, 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 (see Footnote 15). Using (18) we conclude that full market coverage requires:

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

Thus, the condition for a fully covered market requires a σ sufficiently lower than $k_A^2 + k_B^2$, while, from (17), the condition for an interior solution requires a σ sufficiently larger than $k_A^2 - k_B^2$.[17]

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 (P_i') one can check that firm B earns positive profits and that firm A earns higher profits than B : $\pi_A^* > \pi_B^* > 0$. Interestingly, the derivative with respect to α_1 of the firms' profit 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 (17). 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*

[17] Putting all conditions together, market is fully covered and the solution is interior for workers θ_1 and θ_2 $\left(\widehat{\gamma}_j^* < 1 \text{ for } j = 1, 2\right)$ if:

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

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 (11) and by indirect utilities (13).*
(ii) When $k_A = k_B$, $\widehat{\gamma}_1^* = \widehat{\gamma}_2^* = \frac{1}{2}$ *and* $E_A^*(\theta) = E_B^*(\theta)$ *hold.*
(iii) When $k_A > k_B$, $\widehat{\gamma}_1^* > \widehat{\gamma}_2^* > \frac{1}{2}$ *and* $E_A^*(\theta) < E_B^*(\theta)$ *hold.*
(iv) $\widehat{\gamma}_1^*$ *increases with* α_1. *An interior solution,* $\widehat{\gamma}_1^* < 1$, *requires* $\alpha_1 < \alpha_1'$, *where* α_1' *is expressed in (17). Otherwise,* $\widehat{\gamma}_1^* = 1$. $\widehat{\gamma}_2^*$ *is independent of* α_1; *however, when condition* $\alpha_1 < \alpha_1'$ *is met,* $\widehat{\gamma}_2^* < 1$ *necessarily holds.*
(v) The concern for coworkers' quality benefits high-ability *workers (including the ones hired by firm B) but is detrimental to firms.*

Point *(iii)* shows that the firm employing a larger measure of high-ability job market candidates is also hiring a workforce with higher average quality. Hence, a larger market share of talented workers goes hand in hand with a better workforce. Note that, the difference $\widehat{\gamma}_1^* - \widehat{\gamma}_2^* > 0$ incorporates two effects that contribute to the overall equilibrium size and average ability of firm A's workforce. The first one is generated by the good matching between firm A and its high-ability workers which is more profitable than the one between the same firm and its low-ability workers. This effect occurs even absent CfCA and is documented by $\widehat{\gamma}_1^*|_{\alpha_1=0} - \widehat{\gamma}_2^* > 0$. The second effect is generated by CfCA and drives the marginal worker of high-ability to the right even more: such effect is measured by $\widehat{\gamma}_1^* - \widehat{\gamma}_1^*|_{\alpha_1=0} > 0$.

Let us consider point *(v)* of the above proposition. From expression (13) we observe that high-ability workers' indirect utility is increasing in α_1. Intuitively, high-ability workers hired by firm B must be compensated for the utility loss suffered because of CfCA. But, given that indirect utilities are strategic complements, workers employed by firm A also have to be compensated accordingly. From (9), wages follow the same pattern as indirect utilities and high-ability workers' compensation also increases with CfCA. By contrast, CfCA is detrimental to firms. Intuitively, α_1 decreases the mismatch disutility of high-ability workers and thus increases competition. As a result, to attract talented workers, firms must give up a larger share of surplus when $\alpha_1 > 0$ than when $\alpha_1 = 0$. Finally, note that low-type workers are not affected by CfCA.

In Appendix A.12, we present the analysis of the market equilibrium using a richer specification of workers' preferences and CfCA. In this specification, both types of workers may value CfCA, which means that α_2 may be positive. Additionally, the peer effect now depends on the measure of both types, decreasing with $\widehat{\gamma}_2$. These changes introduce new tradeoffs that may affect the sorting pattern. To understand these tradeoffs, we proceed in two steps. First, let's assume that the premium for CfCA is $\pm\alpha_1(\widehat{\gamma}_1 - \widehat{\gamma}_2)$ with $\alpha_2 = 0$. In this case, firms want to reduce the measure of low-ability workers to increase the utility premium they offer to high-ability workers. However, reducing the measure of low-ability workers lowers production and profits. Second, let's assume that $\alpha_2 > 0$, meaning that the premium for CfCA is $\pm\alpha_j(\widehat{\gamma}_1 - \widehat{\gamma}_2)$,

$j = 1, 2$. In this case, the externality generated by peer effects is larger than before, as it now accrues to low-ability workers as well. Therefore, reducing the measure of low-ability workers now implies that a lower share of employees can benefit from the premium. As α_2 increases, low-ability workers benefit more and more from the premium. Hence, to increase the premium for CfCA, it becomes more effective to increase $\widehat{\gamma}_1$ rather than decrease $\widehat{\gamma}_2$. Starting from $\alpha_2 = 0$ and following a continuity argument, firms still want to reduce the measure of low-ability workers when α_2 is positive but small. However, when α_2 becomes sufficiently large and $\alpha_2 > \alpha_1$ holds, firms start to prefer increasing the measure of low-ability workers instead (together with increasing the measure of high-ability workers).

Interestingly, these extra effects do not affect any of our qualitative results. Proposition 6 in Appendix A.12 summarizes the results characterizing the market equilibrium with the richer specification, which (except for some technicalities) is the same as Proposition 1, its counterpart in our simplified model. This means that our reduced-form model, while being tractable, is able to capture the main results on market equilibrium and workers' sorting obtained with the richer specification.

In the following section, we study the optimal allocation that 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 is efficient, one has to compare the equilibrium allocation with the one that maximizes total surplus. Recall that effort levels are set at their efficient levels in equilibrium; hence our welfare analysis focuses on workers' sorting.

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:

$$\begin{aligned} \max_{\{x_A(\theta_j), U_A(\theta_j), \hat{\gamma}_j\}_{j=1,2}} SW = \ & \textstyle\sum_{j=1}^{2} \int_0^{\widehat{\gamma}_j} \left[\pi_A(\theta_j) + \mathcal{U}_A(\theta_j)\right] dF(\gamma) \\ & + \textstyle\sum_{j=1}^{2} \int_{\widehat{\gamma}_j}^{1} \left[\pi_B(\theta_j) + \mathcal{U}_B(\theta_j)\right] dF(\gamma); \end{aligned} \tag{P_W}$$

where profits $\pi_i(\theta_j)$ are defined in (10) and workers' total utilities $\mathcal{U}_A(\theta_j)$ and $\mathcal{U}_B(\theta_j)$ are expressed in (5) and (6). Effort levels $\{x_A(\theta_j)\}_{j=1,2}$ are the efficient ones (see 11).

In Appendix A.2 we show that program P_W can be rewritten as:

$$\max_{\{\hat{\gamma}_j\}_{j=1,2}} SW = \frac{1}{2\theta_1\theta_2}\left[\left(k_A^2 - k_B^2\right)\left(\widehat{\gamma}_1\lambda_1\theta_2 + \widehat{\gamma}_2\lambda_2\theta_1\right) + k_B^2\left(\lambda_1\theta_2 + \lambda_2\theta_1\right)\right] + \\ - \frac{1}{2}\sigma\left[\lambda_1\hat{\gamma}_1^2 + \lambda_2\hat{\gamma}_2^2 + \lambda_1\left(1-\hat{\gamma}_1\right)^2 + \lambda_2\left(1-\hat{\gamma}_2\right)^2\right] \\ + 2\lambda_1\alpha_1\left(\hat{\gamma}_1 - \tfrac{1}{2}\right)^2. \tag{19}$$

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 received by high-ability workers because of their CfCA.

Specifically, the first line of (19) shows how social welfare is affected by firms' marginal productivity. When firm A has a competitive advantage, it hires a relatively larger workforce because this increases social welfare via the productivity gain deriving from the good matching between firms and workers. The second line of (19) reports total mismatch disutility. Finally, the third line of (19) indicates total premium from coworkers' quality accruing high-ability workers 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 $-\lambda_1(1-\hat{\gamma}_1)\alpha_1\left(\hat{\gamma}_1 - \frac{1}{2}\right)$). The last term of (19) is the only one which depends on α_1, and it suggests that social surplus might be increasing in CfCA. To see whether this is the case, let $\widehat{\gamma}_1^{fb}$ and $\widehat{\gamma}_2^{fb}$ denote the efficient marginal workers reported in (A.2) and (A.3) in Appendix A.3. By substituting $\widehat{\gamma}_1^{fb}$ and $\widehat{\gamma}_2^{fb}$ into SW and differentiating with respect to α_1, we formally establish that social surplus is increasing in CfCA.

Intuitively, 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 with $k_A > k_B$ is optimal because *i*) it increases overall productivity, and thus surplus, via the good matching between firm A and high-ability workers, and *ii*) it creates a net premium from CfCA. Benefits *i*) and *ii*) together are larger than the additional mismatch disutility generated by the asymmetric allocation.

The second order condition with respect to $\widehat{\gamma}_1$ requires that $\alpha_1 < \frac{1}{2}\sigma$, 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 $\alpha_1 < \frac{1}{2}\sigma$ holds.

In Appendix A.3, we fully characterize efficient sorting in Proposition 2. In the subsection below, we show that sorting in equilibrium is distorted and we explain why.

4.2 Inefficient sorting in the market equilibrium

Recall that, in 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

(13) into (7) and (8). In the first best, instead, marginal workers are such that the sum of firms' profits and workers' utilities is maximized.

Let us compare equilibrium marginal workers (14)–(15) with efficient marginal workers (A.2)–(A.3) in Appendix A.3. From Proposition 1 above and Proposition 2, illustrated in Appendix A.3, it follows:

Proposition 3 ***Welfare analysis****(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.
(iia) When condition $\alpha_1 < \alpha_1'$ holds, implying that an interior solution emerges for both marginal workers in equilibrium, 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 work-force hired by firm A is too low and the one of firm B is too high ($E_A^*(\theta) > E_A^{fb}(\theta)$ and $E_B^*(\theta) < E_B^{fb}(\theta)$).
(iib) When $\alpha_1 \in [\alpha_1', \alpha_1'')$, where α_1' and α_1'' are respectively defined by (17) and (A.5), a corner solution with $\widehat{\gamma}_1 = 1$ would be efficient but an interior solution with $\widehat{\gamma}_1 < 1$ emerges instead in equilibrium.

Interestingly, only high-ability workers appropriate the surplus generated by CfCA when firms are heterogeneous. 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 share of the increased return from the matching between high-ability workers and the more efficient firm. We further elaborate on that below.

As expressed in part *(iia)* of the proposition, too few high-ability and too few low-ability workers are employed by firm A in equilibrium ($\widehat{\gamma}_j^* < \widehat{\gamma}_j^{fb}$, $j = 1, 2$). This result in itself is not sufficient to compare the average ability of workers hired by the two firms. However, comparing expressions (16) and (A.4), one can easily check that $\widehat{\gamma}_1^* - \widehat{\gamma}_2^* < \widehat{\gamma}_1^{fb} - \widehat{\gamma}_2^{fb}$, meaning that average ability in firm A is inefficiently low. Finally, part *(iib)* of Proposition 3 is explained by the fact that threshold values for interior solutions are such that $\alpha_1' > \alpha_1''$. Hence, the region of the parameters such that $\widehat{\gamma}_1^* < 1$ is too large.

What is the source of the inefficient sorting observed in 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.4. We show that sorting obtained by a monopsonist is inefficient, but to a lesser extent than sorting in the market allocation:

$$\widehat{\gamma}_1^* < \widehat{\gamma}_1^M < \widehat{\gamma}_1^{fb}, \tag{20}$$

$$\widehat{\gamma}_2^* < \widehat{\gamma}_2^M < \widehat{\gamma}_2^{fb}. \tag{21}$$

Consequently, we conclude that strategic interaction and profit maximization jointly contribute to the distortion in workers' sorting. We provide intuitions below.

Market sorting is inefficient for three reasons that sum up and all contribute to the downward distortion of $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$. The first 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: firm B competes too aggressively while firm A 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 firm A and a negative one for workers hired by firm B that are only partially internalized in equilibrium. This channel pushes towards a lower $\widehat{\gamma}_1$ but does not affect $\widehat{\gamma}_2$. Let us consider the three distortions more in detail.

The effect of profit maximization 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 second term in the expression of the social welfare function (19)). 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 (20) and (21) hold.

Let us move to strategic interaction 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 (21). 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, note 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 through the externality introduced by workers' peer effects. Specifically, firm A disregards the utility loss suffered by high-ability workers employed by firm B while firm B disregards the premium accruing high-types hired by firm A. This further reduces $\widehat{\gamma}_1$ in equilibrium.

While we continue to study our simplified model in the main text, we also analyze the efficient solution and the welfare analysis for the richer model in Appendix A.12.3. Despite minor differences, our simplified model produces the same main qualitative results as the richer specification.[18] To verify this, compare Propositions 2 and 3 to Propositions 7 and 8, which are their counterparts in the richer model. When firms are heterogeneous, efficiency in both settings requires the more efficient firm to employ a larger share of high-ability workers and to hire the better workforce, that is the one characterized by the smaller average cost of effort. Moving to the welfare analysis, in both settings $k_A = k_B$ yields efficient sorting in equilibrium. Conversely, in both setting, $k_A > k_B$ entails that equilibrium sorting is inefficient because firm A hires too few high-ability workers.

More specifically, the marginal worker of high-ability is always downward distorted in the market equilibrium with both specifications. Regarding the marginal worker of low-ability, a difference between the two settings may occur when $\alpha_1 > \alpha_2 > 0$. In such a case, in the market equilibrium, the marginal worker of low-ability can be upward distorted in the richer model whereas it is always downward distorted in the simplified model. When instead $\alpha_2 > \alpha_1$, the two settings display the same welfare analysis for both marginal types.[19]

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-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

[18] Subject, once again, to some conditions on α_1 which are necessary for the second-order conditions.

[19] In more details, as α_1 grows larger, a corner solution becomes efficient and, depending on the size of α_2, it is either a monopoly for firm A or full market segmentation with firm A hiring all high-ability and firm B hiring all low-ability workers. The former corner solution becomes efficient when α_2 is large enough. The latter becomes efficient when α_2 is low or zero.

utilities) are given by[20]

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

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

$$\frac{1}{2}(\theta_2 - \theta_1)\, x_i^2(\theta_2) \leq U_i(\theta_1) - U_i(\theta_2) \leq \frac{1}{2}(\theta_2 - \theta_1)\, x_i^2(\theta_1), \qquad (22)$$

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 2 (see Appendix A.8), 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 of worker θ_j, taking as given the indirect utility $U_{-i}(\theta_j)$ 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.

We consider an equilibrium in which the screening entails $\widehat{\gamma}_1 \geq \widehat{\gamma}_2$ as under full information on ability. We verify *ex post* that this condition indeed holds in equilibrium.

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 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 $\{x_i^*(\theta_j), U_i^*(\theta_j)\}$, $i = A, B$, $j = 1, 2$,

[20] 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 $w_i(\theta_2) - \frac{1}{2}\theta_2 x_i^2(\theta_2) + \frac{1}{2}\theta_2 x_i^2(\theta_2) - \frac{1}{2}\theta_1 x_i^2(\theta_2) = U_i(\theta_2) + \frac{1}{2}(\theta_2 - \theta_1)\, x_i^2(\theta_2)$.

where $x_i^*(\theta_j) = x_i^{fb}(\theta_j) = k_i/\theta_j$ and $U_i^*(\theta_j)$ are described in (13) and substitute them into DIC_i and UIC_i, $i = A, B$. One immediately observes that UIC_B is always met (see Appendix A.5 for more details). Rearranging the other incentive constraints, one finds three thresholds that define the relevant regimes (see Figure 1):

$$UIC_B \text{ is slack if } \alpha_1 \leq \frac{\theta_2 - \theta_1}{6\theta_2\theta_1^2}\left[3k_B^2(\theta_2 - \theta_1) - \theta_1(k_A^2 - k_B^2)\right] \equiv \alpha_1^a, \tag{23}$$

$$DIC_A \text{ is slack if } \alpha_1 \geq \frac{\theta_2 - \theta_1}{6\theta_1\theta_2^2}\left[\theta_2(k_A^2 - k_B^2) - 3k_A^2(\theta_2 - \theta_1)\right] \equiv \alpha_1^b, \tag{24}$$

$$UIC_A \text{ is slack if } \alpha_1 \leq \frac{\theta_2 - \theta_1}{6\theta_2\theta_1^2}\left[\theta_1(k_A^2 - k_B^2) + 3k_A^2(\theta_2 - \theta_1)\right] \equiv \alpha_1^c. \tag{25}$$

Lemma 1 in the appendix states conditions under which full-information contracts are incentive compatible. We discuss and explain such conditions in Appendix A.5.

5.2 Screening contracts

We now turn to the case where 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 chain of inequalities in Footnote 17 is met.

In Remark 1 (see Appendix A.7), we derive the ranking of the three threshold values α_1^a, α_1^b and α_1^c defined in (23)–(25). This allows us to check which incentive constraints start to be binding when (A.10)–(A.12) are not met, and α_1 increases. Figure 1 below reports the two possible rankings depending on whether heterogeneity in workers' ability is larger or lower than firms' heterogeneity.

Lemma 2 (see Appendix A.8) 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 2 established the following proposition (see also Figure 1).

Proposition 4 *Under competition and screening, when conditions (A.10)–(A.12) 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 (23)-(25), the following three regimes become relevant in turn:

Regime 1 *Both UIC_B and DIC_A are binding for*

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

and H/peer effects/submissions/3 GEB/revision 2/paper/S86YNA00.wmf

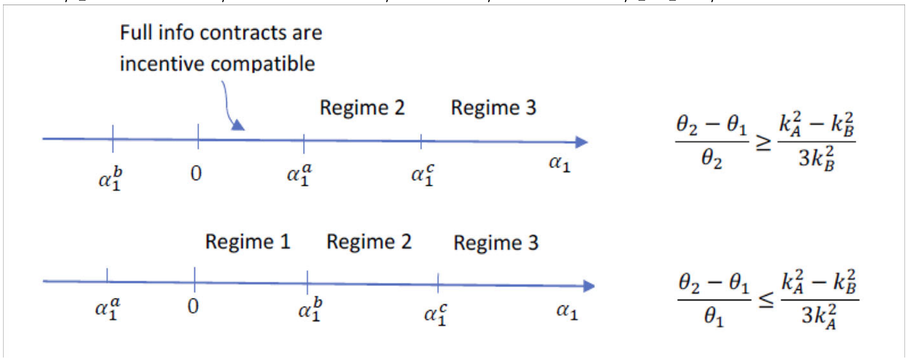

Figure 1: The different regimes according to the relevance of the concern for coworkers' ability.

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*

$$\alpha_1 > \alpha_1^c.$$

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 the two inequalities is necessary for having UIC binding and DIC slack for firm A (see Lemma 2). 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. Note that, since $\pi_A(\theta_2)/\pi_A(\theta_1) > 1$ holds for Remark 2, $\sigma/(\sigma-\alpha_1) \geq \pi_A(\theta_2)/\pi_A(\theta_1)$ is always satisfied for α_1 large enough.[21] Interestingly, Regime 3 is the only one that is compatible with the case of identical firms.

The intuition behind Figure 1 and Proposition 4 is the following. We observe that, in the three regimes, UIC_B is always binding while no incentive constraint at all or one of the two can be binding for firm A.

Let us first consider the case where heterogeneity in workers' ability is relatively larger than the one in firms' marginal productivity, i.e.$(\theta_2-\theta_1)/\theta_2 \geq (k_A^2-k_B^2)/3k_B^2$. Here, since workers are heterogeneous enough, mimicking is too costly for the workers to be attractive and full information contracts are incentive compatible for α_1 sufficiently low. As α_1 increases, competition for talented workers increases and firm B, the disadvantaged one, is the first to be incentive constrained. Indeed, firm A, relying on more resources, can increase low-ability workers' salary and discourage them from mimicking. On the contrary, firm B needs to solve the usual rent-extraction/efficiency trade-off by resorting to high-types' effort distortions: UIC_B starts to be binding and we enter Regime 2. It is only when α_1 becomes so large that the competition for high-ability workers increases substantially, that firm A also needs to resort to overincentivization of high-ability workers in order to prevent low-ability workers' mimicking: UIC_A starts to be binding and we enter Regime 3.

Let us move to the case where workers' heterogeneity is relatively lower than firms' heterogeneity, i.e.$(\theta_2-\theta_1)/\theta_1 \leq (k_A^2-k_B^2)/3k_A^2$. Here, since workers' types are similar, mimicking also occurs for no or very low α_1's. In addition, the competitive pressure exerted by firm B is relatively low. As a result firm A, like a monopolist, finds it convenient to pay its workers in such a way that high-ability workers want to mimic low-ability types (DIC_A is binding). For firm B instead, low-ability types are always the mimickers because they are attracted by the relatively higher salary of their high-ability colleagues (UIC_B is binding). This explains Regime 1. As α_1 grows larger, competition for talented workers increases and contracts offered to high-ability types in firm A improves. As a consequence, DIC_A ceases to be binding and we enter Regime 2. For even larger α_1, UIC_A starts to be binding and we enter Regime 3.

The following proposition, established in Appendix A.10, A.9 and A.11, 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:*

[21] In Appendix A.11.1, we derive a numerical solution under Regime 3 and we check *ex-post* that the condition $\sigma/(\sigma-\alpha_1) \geq \pi_A(\theta_2)/\pi_A(\theta_1) > 1$ is indeed met.

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 full information indirect utilities $U_i^{*}(\theta_j)$, $i = A, B$; $j = 1, 2$.*

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-ability workers' indirect utilities are lower than in the full information equilibrium $(U_i^{**}(\theta_1) < U_i^{*}(\theta_1))$ whereas low-ability workers' 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-ability workers' indirect utilities are lower than in the full information equilibrium $(U_i^{**}(\theta_1) < U_i^{*}(\theta_1))$ whereas low-ability workers' ones are higher $(U_i^{**}(\theta_2) > U_i^{*}(\theta_2))$, $i = A, B$.*

While points (i) in Proposition 5 directly follow from the binding incentive constraints, points (ii) deserve some explanations. First of all, recall that reaction functions (12) 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-ability 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, from (7) and (8), 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-ability 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 other. As a consequence $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ will move in the same direction but to a lower extent than under Regime 2. 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.

Total distortions in effort levels are larger in Regime 1 than in Regime 2 because, under the latter, only the effort level of high-ability 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 1 ***Sorting under screening.*** *Compared to full-information, workers' sorting is such that:*

Regime 1 *(i) The share of high-ability workers employed in firm A increases ($\widehat{\gamma}_1^{**} > \widehat{\gamma}_1^*$) whereas the share of low-ability workers 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-ability workers 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-ability 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-ability 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 distortion 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-ability types are better off, while both the firms and high-ability 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

Regime	Binding constraints	Effort distortions	Indirect utilities compared to full info	Sorting compared to full info
Regime 1	DIC_A and UIC_B	Downward distortion for low-ability workers in firm A; upward distortion for high-ability workers in firm B	In firm A, high-ability workers are better off, while low-ability workers are worse off; the opposite holds in firm B	It improves for high-ability workers; it worsens for low-ability workers
Regime 2	UIC_B	Upward distortion for high-ability workers in firm B	In both firms, high-ability workers are worse off, while low-ability workers are better off	Like in Regime 1, but the effects are larger
Regime 3	UIC_A and UIC_B	Upward distortion for high-ability workers in both firms	Like in Regime 2	Same sorting as under full information

Table 1: Summary of the main properties of screening contracts.

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-ability types are upward distorted. When instead firms differ ($k_A > k_B$), both workers' sorting and high-ability workers' effort levels are distorted.

Notably, Corollary 1 shows that either $\widehat{\gamma}_1$ moves on the right and $\widehat{\gamma}_2$ on the left or the two marginal workers do not change. This proves that our initial conjecture that $\widehat{\gamma}_1 > \widehat{\gamma}_2$ is verified in equilibrium.

Our main results are summarized in Table 1.

Our characterization of the screening equilibrium is qualitative and no closed-form solution is derived. Hence, for each possible regime, in the Appendices A.10.1, A.9.1, and A.11.1, we present a numerical simulation to show that the set of parameter values for which the omitted constraints are indeed satisfied is not empty. This ensures that all the three regimes exist.

To sum up, when ability is not observable and either Regimes 1 or Regime 2 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. 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-ability types both under Regime 2 and under Regime 3. Specifically, the effort exerted by high-ability 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, talented workers are worse off because their indirect utility is reduced. Conversely, low-ability types still exert the efficient level of effort and are better off because they receive a larger indirect utility than under full information. 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 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-ability 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-ability 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 all the surplus from CfCA.

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 candidate's 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 job market applicants, 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 are 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-ability workers face overincentivization 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 some 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 measure 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 a couple of examples.

First, 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-ability workers that one firm is able to hire. In turn, by hiring the larger share of high-ability 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.

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

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

[2] Au P.H. and B.R. Chen (2021), "Matching with peer monitoring," *Journal of Economic Theory*, 192, 105172.

[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] Calvò-Armengol A. and M. O. Jackson (2004), "The effects of social networks on employment and inequality", *American Economic Review* 94, 426–454.

[9] Dustmann C.; A. Glitz; U. Schönberg, and H. Brücker (2016), "Referral-based Job Search Networks," *Review of Economic Studies*, 83(2), 514–546.

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

[11] Echenique F. and M. Yenmez, (2007), "A solution to matching with preferences over colleagues," *Games and Economic Behavior*, 59(1), 46–71.

[12] Gale, D., Shapley, L.S. (1962), "College admissions and the stability of marriage," *Amer. Math. Monthly* 69 (1), 9–15.

[13] Glitz, A. (2017), "Coworker Networks in the Labour Market," *Labour Economics*, Vol. 44, 218–230.

[14] 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.

[15] 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.

[16] Kelso, A.S., Crawford, V.P., (1982), "Job matching, coalition formation, and gross substitutes," *Econometrica* 50, 1483–1504.

[17] 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.

[18] Lindquist, M.J., Sauermann, J. and Y. Zenou (2022), "Peer effects in the workplace. A network approach," mimeo.

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

[20] Nax, H.H., and B.S.R. Pradelski (2016), "Core Stability and Core Selection in a Decentralized Labor Matching Market," *Games* 7(2):10. https://doi.org/10.3390/g7020010.

[21] Pycia M. (2012), "Stability and preference alignment in matching and coalition formation," *Econometrica*, 80(1), 323–362.

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

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

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

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

[26] 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.

[27] van Veldhuizen R., H. Oosterbeek, J. Sonnemans (2018), "Peers at work: Evidence from the lab." PLoS ONE 13(2): e0192038. https://doi.org/10.1371/journal.pone.0192038

[28] Zenou, Y. (2015), "A Dynamic Model of Weak and Strong Ties in the Labor Market," *Journal of Labor Economics* 33, 891–932.

[29] Zenou Y. (2013), "Spatial versus social mismatch," *Journal of Urban Economics* 74, 113–132.

Appendix

A.1 Workers' sorting when $\alpha_1 > \sigma$

To study this case, use (5) to write the utility of the high-ability marginal worker when hired by firm A as:

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

From (A.1), when $\alpha_1 > \sigma$ we have that $\mathcal{U}_A(\theta_1, \widehat{\gamma}_1)$ 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 $\widehat{\gamma}_1$. Thus, if $U_A(\theta_1) > U_B(\theta_1)$, a corner solution exists with $\widehat{\gamma}_1 = 1$. If $U_A(\theta_1) < U_B(\theta_1)$ the corner solution entails $\widehat{\gamma}_1 = 0$.

A.2 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 (11) in the social welfare function the problem simplifies to

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

Solving the integral and rearranging yields the following expression

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

Rearranging the previous formulation of the SW and isolating its three components, one derives (19) the main text.

A.3 Efficient sorting

Efficient sorting entails:

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

$$\widehat{\gamma}_2^{fb} = \frac{1}{2} + \frac{k_A^2 - k_B^2}{4\theta_2\sigma} \geq \frac{1}{2} \tag{A.3}$$

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

Confirming that, when $k_A > k_B$, it is efficient that firm A hires a larger share of workers of each type and the workforce with the higher average ability $\frac{1}{2} < \widehat{\gamma}_2^{fb} < \widehat{\gamma}_1^{fb}$. We also observe that $\widehat{\gamma}_1^{fb}$ is monotonically increasing in α_1.

From (A.2) and (A.3), an interior solution with $\widehat{\gamma}_1^{fb} < 1$ and $\widehat{\gamma}_2^{fb} < 1$ requires that $k_A^2 - k_B^2 < 2\theta_1(\sigma - 2\alpha_1)$ holds, implying that firms' heterogeneity must be sufficiently low. Moreover, if we have an interior solution for $\widehat{\gamma}_1^{fb}$, we also have one for $\widehat{\gamma}_2^{fb}$. The interior condition for $\widehat{\gamma}_1^{fb}$ can also be written as

$$\widehat{\gamma}_1^{fb} < 1 \iff \alpha_1 < \alpha_1'' \equiv \tfrac{1}{2}\sigma - \tfrac{k_A^2 - k_B^2}{4\theta_1}. \tag{A.5}$$

Like in the market equilibrium, see expression (17), 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 $k_A = k_B$, $\widehat{\gamma}_1^{fb} = \widehat{\gamma}_2^{fb} = 1/2$ and $E_A^{fb}(\theta) = E_B^{fb}(\theta)$ hold.*
(ii) The concern for coworkers ability increases total surplus.
(iii) When $k_A > k_B$, $\widehat{\gamma}_1^{fb} > \widehat{\gamma}_2^{fb} > 1/2$ and $E_A^{fb}(\theta) < E_B^{fb}(\theta)$ hold
(iv) $\widehat{\gamma}_1^{fb}$ increases with α_1. An interior solution, $\widehat{\gamma}_1^{fb} < 1$, requires $\alpha_1 < \alpha_1''$, where α_1'' is expressed in (A.5). Otherwise, $\widehat{\gamma}_1^{fb} = 1$. $\widehat{\gamma}_2^{fb}$ is independent of α_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-ability 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 (A.2) when $\alpha_1 = 0$ with (A.3)). But now CfCA becomes also relevant. Specifically, a second benefit from moving $\widehat{\gamma}_1$ on the right arises because of the externality created by CfCA: a the larger share of high-ability workers employed by firm A can enjoy the premium. A third one arises because, as a result, there are fewer high-type workers employed by firm B who suffer the disutility from CfCA.

The following chain of inequalities holds: $\widehat{\gamma}_1^{fb} > \widehat{\gamma}_1^{fb}|_{\alpha_1=0} > \widehat{\gamma}_2^{fb} > 1/2$ 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 $(\widehat{\gamma}_1^{fb} - \widehat{\gamma}_2^{fb})$ and the average ability of the workforce in firm A both increase with α_1.

A.4 The multi-firm monopsonist

Notably, since efforts are set at the efficient level in the market allocation, the unique possible distortion is in workers' sorting to firms.

To understand why workers' sorting is inefficient in equilibrium, 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:

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

where $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ are given by (7) and (8) respectively, while $\pi_i(\theta_j)$ is defined by (10).

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

$$U_A(\theta_1)=\widehat{\gamma}_1\sigma-\alpha_1\left(\hat{\gamma}_1-\tfrac{1}{2}\right) \tag{A.6}$$

$$U_B(\theta_1)=(1-\widehat{\gamma}_1)\sigma+\alpha_1\left(\hat{\gamma}_1-\tfrac{1}{2}\right) \tag{A.7}$$

$$U_A(\theta_2)=\widehat{\gamma}_2\sigma \tag{A.8}$$

$$U_B(\theta_2)=(1-\widehat{\gamma}_2)\sigma \tag{A.9}$$

Substituting the first-best effort levels, $x_i^{fb}(\theta_j)$, in $\pi_i(\theta_j)$ and plugging the indirect utilities (A.6)-(A.9) into the expressions for $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ and $\pi_i(\theta_j)$, 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:

$$\begin{aligned} \widehat{\gamma}_1^M &= \frac{1}{2}+\frac{k_A^2-k_B^2}{8\theta_1(\sigma-\alpha_1)}\geq\frac{1}{2} \\ \widehat{\gamma}_2^M &= \frac{1}{2}+\frac{k_A^2-k_B^2}{8\theta_2\sigma}\geq\frac{1}{2} \\ \widehat{\gamma}_1^M-\widehat{\gamma}_2^M &= \frac{\left(k_A^2-k_B^2\right)(2\alpha_1\theta_1+\sigma(\theta_2-\theta_1))}{8\theta_1\theta_2\sigma(\sigma-\alpha_1)}\geq 0 \end{aligned}$$

Comparing the previous inequalities with (A.2)-(A.3) and with (14)-(15) shows that the ranking of marginal types is the one expressed in (20) and (21) in the main text. Hence, sorting designed by the monopsonist is not efficient but the distortion is lower than the one in 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.5 Incentive compatible first-best contracts

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

$$0\leq\alpha_1\leq\frac{(\theta_2-\theta_1)^2}{2\theta_2\theta_1^2}k^2=\alpha_1^a=\alpha_1^c. \tag{A.10}$$

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

$$0 \leq \alpha_1 \leq \alpha_1^a; \tag{A.11}$$

$$\frac{\theta_2 - \theta_1}{\theta_2} \geq \frac{k_A^2 - k_B^2}{3k_B^2}. \tag{A.12}$$

When $k_A = k_B = k$, DIC_i, $i = A, B$, are always satisfied and, UIC_i, $i = A, B$, are met if Condition (A.10) holds. 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.[22]

When $k_A > k_B$ conditions are more stringent. First, condition (A.11) confirms that the allocation characterized in Proposition 1 is incentive compatible only if α_1 is small enough. In addition, condition (A.12) states that heterogeneity in workers' ability $(\theta_2 - \theta_1)$ must be relatively larger than heterogeneity in firms' productivity $(k_A - k_B)$.

Conditions (A.10)–(A.12) 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.

A.6 Proof of Lemma 1

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:

$$3k_A^2(\theta_2 - \theta_1)^2 + 6\alpha_1\theta_1\theta_2^2 \geq \theta_2(\theta_2 - \theta_1)(k_A^2 - k_B^2) \text{ for } DIC_A \tag{A.13}$$

$$3k_A^2(\theta_2 - \theta_1)^2 + \theta_1(\theta_2 - \theta_1)(k_A^2 - k_B^2) \geq 6\alpha_1\theta_2\theta_1^2 \text{ for } UIC_A \tag{A.14}$$

$$3k_B^2(\theta_2 - \theta_1)^2 + \theta_2(\theta_2 - \theta_1)(k_A^2 - k_B^2) + 6\alpha_1\theta_2^2\theta_1 \geq 0 \text{ for } DIC_B \tag{A.15}$$

$$3k_B^2(\theta_2 - \theta_1)^2 \geq \theta_1(\theta_2 - \theta_1)(k_A^2 - k_B^2) + 6\alpha_1\theta_2\theta_1^2 \text{ for } UIC_B \tag{A.16}$$

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 (A.10) holds. This proves part (i) of Lemma 1.

[22] 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).

Solving (A.13), (A.14) and (A.16) for α_1 (we omit A.15 because DIC_B is always slack) one finds conditions (23)-(25) in the main text and the three relevant threshold values for α_1. Hence, if (23)-(25) 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 Lemma 1.

A.7 A first step to derive the relevant incentive constraints

Let us consider again conditions (23)–(25) and, starting from $\alpha_1 = 0$ and letting α_1 grow larger, let us check which constraint becomes binding first. 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 (A.10)-(A.12) 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)* By comparing the threshold values α_1^a, α_1^b and α_1^c defined in (23)-(25), we observe that $\alpha_1^c > \max\{\alpha_1^a, \alpha_1^b\}$ holds. Note that, for $\alpha_1 > \alpha_1^c > \alpha_1^b$, UIC_A binds while DIC_A is slack. In addition, $\alpha_1 > \alpha_1^c > \alpha_1^a$ implies that 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 workers' and firms' heterogeneity 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.8 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 2 *(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 2, 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

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

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

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

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 2. As mentioned in the main text we assume that, under asymmetric information on ability, $\hat{\gamma}_1 \geq \hat{\gamma}_2$ holds. We check ex-post that this is true in equilibrium.

- Firm A solves:

$$\begin{array}{rl} \max_{\{x_A(\theta_j),U_A(\theta_j)\}_{j=1,2}} E(\pi_A)= & \lambda_1\frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)}\left(k_Ax_A(\theta_1)-\frac{1}{2}\theta_1x_A^2(\theta_1)-U_A(\theta_1)\right) \\ & +\lambda_2\frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{2\sigma}\left(k_Ax_A(\theta_2)-\frac{1}{2}\theta_2x_A^2(\theta_2)-U_A(\theta_2)\right) \\ & \\ s.t. & U_A(\theta_1)-U_A(\theta_2)-\frac{1}{2}(\theta_2-\theta_1)x_A^2(\theta_2)\geq 0 \qquad (\mu_{D_A}) \\ & U_A(\theta_2)-U_A(\theta_1)+\frac{1}{2}(\theta_2-\theta_1)x_A^2(\theta_1)\geq 0 \qquad (\mu_{U_A}) \end{array} \qquad (P_A)$$

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:

$$\lambda_1\frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)}(k_A-\theta_1x_A(\theta_1))+\mu_{U_A}(\theta_2-\theta_1)x_A(\theta_1)=0; \qquad (A.17)$$

$$\lambda_2\frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{2\sigma}(k_A-\theta_2x_A(\theta_2))-\mu_{D_A}(\theta_2-\theta_1)x_A(\theta_2)=0. \qquad (A.18)$$

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

$$\frac{\lambda_1}{2(\sigma-\alpha_1)}\left(k_A x_A(\theta_1)-\frac{1}{2}\theta_1 x_A^2(\theta_1)-U_A(\theta_1)\right)-\lambda_1\frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)}+\mu_{D_A}-\mu_{U_A}=0; \tag{A.19}$$

$$\frac{\lambda_2}{2\sigma}\left(k_A x_A(\theta_2)-\frac{1}{2}\theta_2 x_A^2(\theta_2)-U_A(\theta_2)\right)-\lambda_2\frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{2\sigma}-\mu_{D_A}+\mu_{U_A}=0. \tag{A.20}$$

- When DIC and UIC are both binding, (22) 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.17) and (A.18) imply:

  $$\lambda_1\hat{\gamma}_1(k_A-\theta_1 x_A)+\mu_{U_A}(\theta_2-\theta_1)x_A=0; \tag{A.21}$$

  $$-\lambda_2\hat{\gamma}_2(k_A-\theta_2 x_A)+\mu_{D_A}(\theta_2-\theta_1)x_A=0. \tag{A.22}$$

  Summing up (A.21) and (A.22) 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.17) and (A.18) 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.19) and (A.20) 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}

  $$\frac{\lambda_1}{2(\sigma-\alpha_1)}\left(k_A x_A(\theta_1)-\frac{1}{2}\theta_1 x_A^2(\theta_1)-U_A(\theta_1)\right)-\lambda_1\frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)}>0; \tag{A.23}$$

  $$\frac{\lambda_2}{2\sigma}\left(k_A x_A(\theta_2)-\frac{1}{2}\theta_2 x_A^2(\theta_2)-U_A(\theta_2)\right)-\lambda_2\frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{2\sigma}<0; \tag{A.24}$$

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.23) and (A.33) 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.19) and (A.20) become:

$$\frac{\lambda_1}{2(\sigma-\alpha_1)}\left(k_A x_A(\theta_1) - \frac{1}{2}\theta_1 x_A^2(\theta_1) - U_A(\theta_1)\right) - \lambda_1\frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{2(\sigma-\alpha_1)} + \mu_{D_A} = 0; \tag{A.25}$$

$$\frac{\lambda_2}{2\sigma}\left(k_A x_A(\theta_2) - \frac{1}{2}\theta_2 x_A^2(\theta_2) - U_A(\theta_2)\right) - \lambda_2\frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{2\sigma} - \mu_{D_A} = 0. \tag{A.26}$$

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:

$$\frac{\sigma-\alpha_1+U_A(\theta_1)-U_B(\theta_1)}{\sigma-\alpha_1} > \frac{\pi_A(\theta_1)}{\sigma-\alpha_1} > \frac{\pi_A(\theta_2)}{\sigma} > \frac{\sigma+U_A(\theta_2)-U_B(\theta_2)}{\sigma}, \tag{A.27}$$

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.27) 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:

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

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

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:

$$\lambda_1 \left(\frac{\sigma-\alpha_1-U_A(\theta_1)+U_B(\theta_1)}{2(\sigma-\alpha_1)}\right)(k_B - \theta_1 x_B(\theta_1)) + \mu_{U_B}(\theta_2-\theta_1) x_B(\theta_1) = 0; \tag{A.28}$$

$$\lambda_2 \left(\frac{\sigma-U_A(\theta_2)+U_B(\theta_2)}{2\sigma}\right)(k_B - \theta_2 x_B(\theta_2)) - \mu_{D_B}(\theta_2-\theta_1) x_B(\theta_2) = 0. \tag{A.29}$$

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

$$\frac{\lambda_1}{2(\sigma-\alpha_1)}\left(k_B x_B(\theta_1) - \frac{1}{2}\theta_1 x_B^2(\theta_1) - U_B(\theta_1)\right) - \lambda_1 \frac{\sigma-\alpha_1-U_A(\theta_1)+U_B(\theta_1)}{2(\sigma-\alpha_1)} + \mu_{D_B} - \mu_{U_B} = 0; \tag{A.30}$$

$$\frac{\lambda_2}{2\sigma}\left(k_B x_B(\theta_2) - \frac{1}{2}\theta_2 x_B^2(\theta_2) - U_B(\theta_2)\right) - \lambda_2 \frac{\sigma-U_A(\theta_2)+U_B(\theta_2)}{2\sigma} - \mu_{D_A} + \mu_{U_A} = 0. \tag{A.31}$$

- 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.30) and (A.31):

$$\frac{\lambda_1}{2(\sigma-\alpha_1)}\left(k_B x_B(\theta_1) - \frac{1}{2}\theta_1 x_B^2(\theta_1) - U_B(\theta_1)\right) - \lambda_1 \frac{\sigma-\alpha_1-U_A(\theta_1)+U_B(\theta_1)}{2(\sigma-\alpha_1)} - \mu_{U_B} = 0; \tag{A.32}$$

$$\frac{\lambda_2}{2\sigma}\left(k_B x_B(\theta_2) - \frac{1}{2}\theta_2 x_B^2(\theta_2) - U_B(\theta_2)\right) - \lambda_2 \frac{\sigma-U_A(\theta_2)+U_B(\theta_2)}{2\sigma} + \mu_{U_A} = 0. \tag{A.33}$$

Substituting for per-workers profits $\pi_B(\theta_1)$ in (A.32) and (A.33) and dropping μ_{U_B} :

$$\frac{\pi_B(\theta_1)}{2(\sigma-\alpha_1)} > \frac{\sigma-\alpha_1-U_A(\theta_1)+U_B(\theta_1)}{2(\sigma-\alpha_1)}; \tag{A.34}$$

$$\frac{\pi_B(\theta_2)}{2\sigma} < \frac{\sigma-U_A(\theta_2)+U_B(\theta_2)}{2\sigma}. \tag{A.35}$$

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.34) and (A.35) 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.30) and (A.31):

$$\frac{\lambda_1}{2(\sigma-\alpha_1)}\left(k_B x_B(\theta_1) - \frac{1}{2}\theta_1 x_B^2(\theta_1) - U_B(\theta_1)\right) - \lambda_1 \frac{\sigma-\alpha_1-U_A(\theta_1)+U_B(\theta_1)}{2(\sigma-\alpha_1)} + \mu_{D_B} = 0; \tag{A.36}$$

$$\frac{\lambda_2}{2\sigma}\left(k_B x_B(\theta_2) - \frac{1}{2}\theta_2 x_B^2(\theta_2) - U_B(\theta_2)\right) - \lambda_2 \frac{\sigma-U_A(\theta_2)+U_B(\theta_2)}{2\sigma} - \mu_{D_A} = 0. \tag{A.37}$$

Substituting for profits in (A.36) and (A.37) and dropping μ_{D_B}:

$$\frac{\pi_B(\theta_1)}{\sigma-\alpha_1} < \frac{\sigma-\alpha_1-U_A(\theta_1)+U_B(\theta_1)}{\sigma-\alpha_1}; \tag{A.38}$$

$$\frac{\pi_B(\theta_2)}{\sigma} > \frac{\sigma-U_A(\theta_2)+U_B(\theta_2)}{\sigma}. \tag{A.39}$$

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.38) and (A.39) 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.9 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.19) and (A.20) and expression (12): $U_A(\theta_1) = \frac{k_A^2+2\theta_1(U_B(\theta_1)+\alpha_1-\sigma)}{4\theta_1}$ 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.28) and (A.29)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-ability 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.31) 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.30):

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

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:

$$U_B^{**}(\theta_1) < \frac{2}{3}\left(k_B x_B^{**}(\theta_1) - \frac{1}{2}\theta_1 (x_B^{**}(\theta_1))^2\right) + \alpha_1 - \sigma + \frac{k_A^2}{6\theta_1}. \tag{A.40}$$

By solving for $U_B(\theta_1)$ the FOC (A.30) 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:

$$U_B^*(\theta_1) = \frac{2}{3}\left(k_B x_B^{fb}(\theta_1) - \frac{1}{2}\theta_1 \left(x_B^{fb}(\theta_1)\right)^2\right) + \alpha_1 - \sigma + \frac{k_A^2}{6\theta_1}. \tag{A.41}$$

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.40) and (A.41) 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 (12), 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.

Since our characterization of the screening equilibrium is qualitative (no closed-form solution has been derived), for each possible regime, we present a numerical simulation to show that (at least) one solution exists such that the omitted constraints are indeed satisfied. Specifically, we must check that workers' participation constraints, the monotonicity conditions for effort levels and the nonbinding incentives constraints are all satisfied in the equilibrium with screening. In addition, we also check that firms' profits are non-negative, and the marginal workers are interior in equilibrium both under full information and with screening. Finally, in Regime 3 we check that $\frac{\sigma}{\sigma-\alpha_1} > \frac{\pi_A(\theta_2)}{\pi_A(\theta_1)} > 1$ holds.

A.9.1 Numerical simulations under Regime 2

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- ability types and increase for low-ability 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.10 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.17) and (A.18) 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

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

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-ability types is downward distorted.

From FOCs (A.19) and (A.20), one has:

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

and

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

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:

$$S_A^{fb}(\theta_1) - 2U_A^{**}(\theta_1) + U_B^{**}(\theta_1) - \sigma + \alpha_1 < 0; \tag{A.44}$$

and

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

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:

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

and

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

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

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

As for type θ_2, from (A.45) and (A.47) 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

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

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-ability types is upward distorted while the effort of low-ability types is set at the efficient level.

From FOCs (A.30) and (A.31), one has:

$$\lambda_1 \frac{k_B x_B(\theta_1) - \frac{1}{2}\theta_1 x_B^2(\theta_1) - U_B(\theta_1)}{2(\sigma - \alpha_1)} - \lambda_1 \frac{\sigma - \alpha_1 - U_A(\theta_1) + U_B(\theta_1)}{2(\sigma - \alpha_1)} - \mu_{U_B} = 0. \tag{A.50}$$

and

$$\lambda_2 \frac{k_B x_B(\theta_2) - \frac{1}{2}\theta_2 x_B^2(\theta_2) - U_B(\theta_2)}{2\sigma} - \lambda_2 \frac{\sigma - U_A(\theta_2) + U_B(\theta_2)}{2\sigma} + \mu_{U_B} = 0. \tag{A.51}$$

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:

$$S_B^{**}(\theta_1) + U_A^{**}(\theta_1) - 2U_B^{**}(\theta_1) - \sigma + \alpha_1 > 0; \tag{A.52}$$

and

$$S_B^{fb}(\theta_2) + U_A^{**}(\theta_2) - 2U_B^{**}(\theta_2) - \sigma < 0. \tag{A.53}$$

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

$$S_B^{fb}(\theta_1) + U_A^*(\theta_1) - 2U_B^*(\theta_1) - \sigma + \alpha_1 = 0; \tag{A.54}$$

and

$$S_B^{fb}(\theta_2) + U_A^*(\theta_2) - 2U_B^*(\theta_2) - \sigma = 0. \tag{A.55}$$

As for type θ_1, putting together (A.52) and (A.54) 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

$$2U_B^*(\theta_1) - U_A^*(\theta_1) > 2U_B^{**}(\theta_1) - U_A^{**}(\theta_1) \Leftrightarrow$$
$$\left(U_A^{**}(\theta_1) - U_A^*(\theta_1)\right) - 2\left(U_B^{**}(\theta_1) - U_B^*(\theta_1)\right) > 0. \tag{A.56}$$

As for type θ_2, from (A.53) and (A.55) 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

$$2U_B^*(\theta_2) - U_A^*(\theta_2) < 2U_B^{**}(\theta_2) - U_A^{**}(\theta_2) \Leftrightarrow$$
$$2\left(U_B^{**}(\theta_2) - U_B^*(\theta_2)\right) - \left(U_A^{**}(\theta_2) - U_A^*(\theta_2)\right) > 0. \tag{A.57}$$

Now, for high-ability types consider (A.48) and (A.56) 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-ability types, (A.49) and (A.57) together imply:

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

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

A.10.1 Numerical simulations under Regime 1

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.11 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.17) and (A.18) 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-ability types is upward distorted.

Rearranging (A.19) and (A.20) 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:

$$S_A^{**}(\theta_1) - 2U_A^{**}(\theta_1) + U_B^{**}(\theta_1) - \sigma + \alpha_1 > 0; \tag{A.58}$$

and

$$S_A^{fb}(\theta_2) - 2U_A^{**}(\theta_2) + U_B^{**}(\theta_2) - \sigma < 0. \tag{A.59}$$

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.46) and (A.47).

As for type θ_1, putting together (A.58) and (A.46) 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

$$2U_A^{*}(\theta_1) - U_B^{*}(\theta_1) > 2U_A^{**}(\theta_1) - U_B^{**}(\theta_1) \Leftrightarrow$$
$$(U_B^{**}(\theta_1) - U_B^{*}(\theta_1)) - 2(U_A^{**}(\theta_1) - U_A^{*}(\theta_1)) > 0. \tag{A.60}$$

As for type θ_2, from (A.59) and (A.47) 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

$$2U_A^{*}(\theta_2) - U_B^{*}(\theta_2) < 2U_A^{**}(\theta_2) - U_B^{**}(\theta_2) \Leftrightarrow$$
$$2(U_A^{**}(\theta_2) - U_A^{*}(\theta_2)) - (U_B^{**}(\theta_2) - U_B^{*}(\theta_2)) > 0. \tag{A.61}$$

Inequalities (A.56) and (A.60) together write

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

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 (13) 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.57) and (A.61) 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 (13) 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 $\hat{\gamma}_2^{**} = \hat{\gamma}_2^* = \frac{1}{2} + \frac{k_A^2 - k_B^2}{12\sigma\theta_2}$. Thus, 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.11.1 Numerical simulations under Regime 3

Recall that this is the unique regime that is compatible with the firms being identical and thus equally sharing the market. Let us focus on a symmetric equilibrium then. 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-ability types is upward distorted whereas indirect utilities of high-ability 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.12 Richer specification of the CfCA

Suppose that all workers, irrespective of their types, receive a utility premium (loss) that increases with the measure of high-ability coworkers and decreases with the measure of low-ability coworkers:

$$u_A(x_A, w_A; \theta_j, \gamma) = w_A(x_A) - \frac{1}{2}\theta_j x_A^2 - \gamma\sigma + \alpha_j(\hat{\gamma}_1 - \hat{\gamma}_2), \; j = 1, 2; \tag{A.62}$$

$$u_B(x_B, w_B; \theta_j, \gamma) = \underbrace{w_B(x_B) - \frac{1}{2}\theta_j x_B^2}_{\text{net compensation } U_i(\theta_j)} \underbrace{-(1-\gamma)\sigma}_{\text{mismatch disutility}} \underbrace{-\alpha_j(\hat{\gamma}_1 - \hat{\gamma}_2),}_{\text{concern for coworkers' ability}} \quad j = 1, 2. \tag{A.63}$$

Here, workers enjoy (suffer) a utility premium (loss) when they are employed in the workforce characterized by the higher (lower) average ability. Take firm A and the share of its high-ability and low-ability employees, namely $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$. When $\widehat{\gamma}_1 > \widehat{\gamma}_2$, firm A hires the workforce with the higher average ability. Thus, workers hired in firm A receive a premium, whereas workers hired by the competitor suffer a loss of the same amount. When $\widehat{\gamma}_1 < \widehat{\gamma}_2$ premium and loss are reversed. Note that, when employed by firm B, a worker's premium for coworkers' ability is $+\alpha_j((1-\widehat{\gamma}_1) - (1-\widehat{\gamma}_2)) = -\alpha_j(\widehat{\gamma}_1 - \widehat{\gamma}_2)$.

If $\alpha_1 > \alpha_2 \geq 0$, 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. The reduced-form model analyzed in the main text represents a special case of that.

If instead $\alpha_2 \geq \alpha_1 > 0$, CfCA is higher for low-ability workers. Here, joining a firm with a high proportion of high-ability workers could be a strong indicator of an individual's ability level, which may hold more value for low-ability workers than for high-ability workers who have alternative methods to demonstrate their competence.

A.12.1 Marginal workers

A worker of type (θ_j, γ) gets *total* indirect utility

$$\mathcal{U}_A(\theta_j) = U_A(\theta_j) - \gamma\sigma + \alpha_j(\hat{\gamma}_1 - \hat{\gamma}_2) \tag{A.64}$$

if employed by firm A and *total* indirect utility

$$\mathcal{U}_B(\theta_j) = U_B(\theta_j) - (1-\gamma)\sigma - \alpha_j(\hat{\gamma}_1 - \hat{\gamma}_2) \tag{A.65}$$

if employed by firm B.

Marginal workers at the interior solution are defined as follows:

$$\widehat{\gamma}_1 = \frac{1}{2} + \frac{(\sigma+\alpha_2)(U_A(\theta_1) - U_B(\theta_1)) - \alpha_1(U_A(\theta_2) - U_B(\theta_2))}{2\sigma(\sigma - \alpha_1 + \alpha_2)}; \tag{A.66}$$

$$\widehat{\gamma}_2 = \frac{1}{2} + \frac{(\sigma-\alpha_1)(U_A(\theta_2) - U_B(\theta_2)) + \alpha_2(U_A(\theta_1) - U_B(\theta_1))}{2\sigma(\sigma - \alpha_1 + \alpha_2)}. \tag{A.67}$$

The SOCs of the firms' problem require again that $\alpha_1 < \sigma$ which implies that the denominators in (A.66) and (A.67) are positives: as in the reduced model, CfCA is not strong enough to reverse the standard Hotelling "forces." In addition one can check that the SOCs of the social welfare's maximization problem require that $\alpha_1 < \frac{1}{2}\sigma$ which, as in the main text, we assume.

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 (A.66)-(A.67) with (7)-(8) one observes the following. The denominator $2\sigma(\sigma - \alpha_1 + \alpha_2)$ indicates that, while CfCA still increases competition for high-ability workers via the parameter α_1, now it simultaneously decreases competition for low-ability workers via the parameter α_2. 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-ability types are attracted in firm A and this accrues the premium for coworkers' ability that high and low-ability workers employed by firm A receive. In addition, $\widehat{\gamma}_1$ is decreasing in the difference $(U_A(\theta_2) - U_B(\theta_2))$ because, if that difference increases, a larger share of low-ability workers are attracted in firm A and this negatively affects the premium received by its high-ability employees. Instead, given that $\alpha_1 < \sigma$, $\widehat{\gamma}_2$ is increasing in the difference $(U_A(\theta_2) - U_B(\theta_2))$ like in a standard Hotelling model, despite low-ability workers' CfCA. In different words, the benefit from a relatively higher indirect utility more than compensate the negative impact of a larger $\widehat{\gamma}_2$ on the premium from CfCA accruing firm A's low-ability employees.

A.12.2 Equilibrium contracts when workers' ability is observable

Each firm maximizes profits obtained by multiplying (10) by their workforce. Hence, firm A and B respectively solves the following program:

$$\begin{aligned}
\max_{\{x_A(\theta_j),U_A(\theta_j)\}_{j=1,2}} \pi_A = & \lambda_1\widehat{\gamma}_1\left(k_A x_A(\theta_1) - \tfrac{1}{2}\theta_1 x_A^2(\theta_1) - U_A(\theta_1)\right) \\
& +\lambda_2\widehat{\gamma}_2\left(k_A x_A(\theta_2) - \tfrac{1}{2}\theta_2 x_A^2(\theta_2) - U_A(\theta_2)\right) \\
\max_{\{x_B(\theta_j),U_B(\theta_j)\}_{j=1,2}} \pi_B = & \lambda_1(1-\widehat{\gamma}_1)\left(k_B x_B(\theta_1) - \tfrac{1}{2}\theta_1 x_B^2(\theta_1) - U_B(\theta_1)\right) \\
& +\lambda_2(1-\widehat{\gamma}_2)\left(k_B x_B(\theta_2) - \tfrac{1}{2}\theta_2 x_B^2(\theta_2) - U_B(\theta_2)\right)
\end{aligned} \tag{P_i2}$$

where $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ are given by (A.66) and (A.67) and effort levels are the first-best ones in (11).

We solve firms' programs for indirect utilities. Specifically, 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. We obtain four reaction functions, two for each firm. Then we solve the system of the four reaction functions in four unknowns and find indirect utilities in equilibrium: $U_i(\theta_i)$, $i=A,B$ and $j=1,2$. Substituting $U_i(\theta_i)$ into (A.66) and (A.67) and rearranging, we obtain the expressions for marginal workers:

$$\widehat{\gamma}_1^{*e} = \frac{1}{2} + \frac{\left(k_A^2-k_B^2\right)\lambda_2\left(2\lambda_2\theta_1\alpha_2+\lambda_1\left(3\theta_2(\sigma+\alpha_2)-\alpha_1\theta_1\right)\right)}{4\theta_1\theta_2[9\lambda_1\lambda_2\sigma(\sigma-\alpha_1+\alpha_2)-2\left(\alpha_1\lambda_1+\alpha_2\lambda_2\right)^2]}; \tag{A.68}$$

$$\widehat{\gamma}_2^{*e} = \frac{1}{2} + \frac{\left(k_A^2-k_B^2\right)\left(2\alpha_1\theta_2\lambda_1^2-3(\sigma-\alpha_1)\theta_1\lambda_2\lambda_1-\alpha_2\theta_2\lambda_2\right)}{4\theta_1\theta_2[9\lambda_1\lambda_2\sigma(\sigma-\alpha_1+\alpha_2)-2\left(\alpha_1\lambda_1+\alpha_2\lambda_2\right)^2]}; \tag{A.69}$$

where the superscript $*e$ indicates the market equilibrium in this extended model. The premium/loss for CfCA writes:

$$\widehat{\gamma}_1^{*e} - \widehat{\gamma}_2^{*e} = \frac{\left(k_A^2-k_B^2\right)\left(2\left(\alpha_1\lambda_1+\alpha_2\lambda_2\right)\left(\theta_1\lambda_2+\lambda_1\theta_2\right)+3\sigma\lambda_1\lambda_2(\theta_2-\theta_1)\right)}{4\theta_1\theta_2[9\lambda_1\lambda_2\sigma(\sigma-\alpha_1+\alpha_2)-2\left(\alpha_1\lambda_1+\alpha_2\lambda_2\right)^2]}. \tag{A.70}$$

Expressions (A.68)-(A.70) all have the same vertical asymptote $\hat{\alpha}_1^e$:

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

When $\alpha_1 < \hat{\alpha}_1^e$ holds, then the denominators in (A.68)-(A.70) are all positive. The numerator of the second term in (A.68) is always positive because $(\sigma-\alpha_1+\alpha_2) > 0$ implies that $(3\theta_2(\sigma+\alpha_2)-\alpha_1\theta_1) > 0$. Thus, $\widehat{\gamma}_1^{*e} > \frac{1}{2}$ always holds for $\alpha_1 < \hat{\alpha}_1^e$. In addition, since the numerator of (A.70) is positive, $\widehat{\gamma}_1^{*e} - \widehat{\gamma}_2^{*e} > 0$ is satisfied for $\alpha_1 < \hat{\alpha}_1^e$. Finally, the numerator of the second term in (A.69) is positive for $\alpha_1 < \tilde{\alpha}_1^e$, where:

$$0 < \tilde{\alpha}_1^e \equiv \frac{\lambda_2\left(3\theta_1\sigma+\theta_2\alpha_2\right)}{3\lambda_2\theta_1+2\lambda_1\theta_2} < \hat{\alpha}_1^e.$$

Hence, from (A.69), $\widehat{\gamma}_2^{*e} \geq \frac{1}{2}$ holds for $0 < \alpha_1 \leq \tilde{\alpha}_1^e$ whereas $\widehat{\gamma}_2^{*e} < \frac{1}{2}$ holds for $\tilde{\alpha}_1^e < \alpha_1 < \hat{\alpha}_1^e$.

Notably, the threshold $\tilde{\alpha}_1^e$ is increasing in α_2. Hence, a sufficient condition to have $\tilde{\alpha}_1^e \geq \frac{1}{2}\sigma$ is $\tilde{\alpha}_1^e|_{\alpha_2=0} > \frac{1}{2}\sigma$. One can easily check that the previous inequality holds if:

$$\frac{\theta_1}{\theta_2} \geq \frac{2}{3}\frac{\lambda_1}{\lambda_2}. \tag{A.71}$$

In words, a low workers' heterogeneity and/or a share of high-ability workers not too high represent sufficient conditions for $\tilde{\alpha}_1^e \geq \frac{1}{2}\sigma$. This in turn implies that $\widehat{\gamma}_2^{*e} \geq \frac{1}{2}$ always holds in the parameters' range we are considering ($\alpha_1 < \frac{1}{2}\sigma$).

To sum up, (A.71) is a sufficient condition for $\widehat{\gamma}_1^{*e} - \widehat{\gamma}_2^{*e} > 0$ and $\widehat{\gamma}_1^{*e} > \widehat{\gamma}_2^{*e} > \frac{1}{2}$ to hold when the necessary condition for a concave social welfare function is met.

By plugging indirect utilities into the expressions of marginal workers' utilities we obtain:

$$\mathcal{U}_A^{*e}(\theta_1, \widehat{\gamma}_1^{*e}) = \mathcal{U}_B^{*e}(\theta_1, \widehat{\gamma}_1^{*e}) = \frac{k_A^2 + k_B^2 - 6\theta_1\sigma}{4\theta_1} + \alpha_1 + \alpha_2\frac{\lambda_2}{\lambda_1}; \tag{A.72}$$

$$\mathcal{U}_A^{*e}(\theta_2, \widehat{\gamma}_2^{*e}) = \mathcal{U}_B^{*e}(\theta_2, \widehat{\gamma}_2^{*e}) = \frac{k_A^2 + k_B^2 - 6\theta_2\sigma}{4\theta_2} - \alpha_2 - \alpha_1\frac{\lambda_1}{\lambda_2}. \tag{A.73}$$

Since other employees' payoff is strictly larger than the one of marginal workers, the preceding expressions demonstrate that CfCA enhances the utility of high-ability workers through α_1 and α_2, while simultaneously reducing the utility of low-ability workers.

Recall that, in the reduced-form model, α_1 was accruing only high-ability workers' utility; see (13). The mechanism behind expressions (A.72)-(A.73) is that α_1 and α_2 simultaneously enhance competition for high-ability workers (especially when the share of low-ability types, λ_2, is high), and weaken competition for low-ability workers (especially when the share of high-ability types, λ_1, is high).

Results so far are summarized below:

Proposition 6 ***Full-information equilibrium in the richer specification.*** *(i) When ability is observable while mismatch disutility is the workers' private information, equilibrium contracts are the Nash equilibrium contracts of the game in which firms compete in the utility space. Effort levels are the efficient ones reported in (12).*
(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) When $k_A > k_B$ and (A.71) holds, then $\widehat{\gamma}_1^{*e} > \widehat{\gamma}_2^{*e} > \frac{1}{2}$, $\widehat{\gamma}_1^{*e} - \widehat{\gamma}_2^{*e} > 0$ and $E_A^{*e}(\theta) < E_B^{*e}(\theta)$.
(iv) $\widehat{\gamma}_1^{e*}$ increases with α_1. When α_2 is small, $\widehat{\gamma}_2^{fbe}$ decreases with α_1. When α_2 is large, $\widehat{\gamma}_2^{fbe}$ increases with α_1 but to a lower extent than $\widehat{\gamma}_1^{e*}$.
(v) The concern for coworkers' quality benefits high-ability workers and firm B while it is detrimental to low-ability workers and firm A.

Points (*iii*) in the proposition above is valid within the range of parameters, $\alpha_1 < \frac{1}{2}\sigma$, where the necessary condition for the social welfare program to be concave is satisfied. All results (*i*)-(*iii*) confirm Proposition 1 in the main text.

Point (iv) shows additional results. Specifically, in this richer model, CfCA negatively affects low-ability workers because it reduces competition for this specific workers' type. Moreover, while CfCA continues to negatively affect the efficient firm A, i.e. the firm employing most high-ability types, it now positively affects firm B. Indeed, by hiring more low- than high-ability workers, firm B more than compensates the profits' loss on high-ability types (that are paid relatively more when $\alpha_2 > 0$ than when $\alpha_2 = 0$) with the gain on low-ability types (that are paid less when $\alpha_2 > 0$ than when $\alpha_2 = 0$). The opposite holds for firm A.

A.12.3 Welfare analysis in the richer specification

With this specification, program P_W can be rewritten as:

$$\begin{aligned}\max_{\{\widehat{\gamma}_j\}_{j=1,2}} SW' = & \tfrac{1}{2\theta_1\theta_2}\left[\left(k_A^2 - k_B^2\right)\left(\lambda_1\widehat{\gamma}_1\theta_2 + \lambda_2\widehat{\gamma}_2\theta_1\right) + k_B^2\left(\lambda_1\theta_2 + \lambda_2\theta_1\right)\right] \\ & -\tfrac{1}{2}\sigma\left[\lambda_1\left(\widehat{\gamma}_1\right)^2 + \lambda_2\left(\widehat{\gamma}_2\right)^2 + \lambda_1\left(1-\widehat{\gamma}_1\right)^2 + \lambda_2\left(1-\widehat{\gamma}_2\right)^2\right] \\ & +2\left(\widehat{\gamma}_1 - \widehat{\gamma}_2\right)\left[\lambda_1\alpha_1\left(\widehat{\gamma}_1 - \tfrac{1}{2}\right) + \lambda_2\alpha_2\left(\widehat{\gamma}_2 - \tfrac{1}{2}\right)\right];\end{aligned} \tag{A.74}$$

the difference with respect to the reduced model is in the third term of (A.74) that previously was $+2\lambda_1\alpha_1\left(\widehat{\gamma}_1 - \frac{1}{2}\right)^2$ and is now $+2\left(\widehat{\gamma}_1 - \widehat{\gamma}_2\right)\left[\lambda_1\alpha_1\left(\widehat{\gamma}_1 - \frac{1}{2}\right) + \lambda_2\alpha_2\left(\widehat{\gamma}_2 - \frac{1}{2}\right)\right]$. The interpretation, however, remains the same as in (A.74) in the main text: the third line of (A.74) indicates total premium from coworkers' quality accruing workers employed in A (given by $(\lambda_1\widehat{\gamma}_1\alpha_1 + \lambda_2\widehat{\gamma}_2\alpha_2)\left(\widehat{\gamma}_1 - \widehat{\gamma}_2\right)$) net of the disutility experienced by workers hired by firm B (given by $-(\lambda_1\left(1-\widehat{\gamma}_1\right)\alpha_1 + \lambda_2\left(1-\widehat{\gamma}_2\right)\alpha_2)\left(\widehat{\gamma}_1 - \widehat{\gamma}_2\right)$).

Solving (A.74) for marginal workers one obtains the efficient sorting and the efficient premium for CfCA:

$$\widehat{\gamma}_1^{fbe} = \frac{1}{2} + \frac{\lambda_2\left(k_A^2 - k_B^2\right)\left[\alpha_2\left(\lambda_1\theta_2 + \lambda_2\theta_1\right) + \lambda_1(\theta_2(\sigma + \alpha_2) - \alpha_1\theta_1)\right]}{4\theta_1\theta_2\left[2\lambda_1\lambda_2\sigma\left(\frac{1}{2}\sigma - \alpha_1 + \alpha_2\right) - (\lambda_2\alpha_2 + \lambda_1\alpha_1)^2\right]}; \tag{A.75}$$

$$\widehat{\gamma}_2^{fbe} = \frac{1}{2} + \frac{\lambda_1\left(k_A^2 - k_B^2\right)\left[\lambda_2\left(\sigma\theta_1 + \theta_2\alpha_2\right) - \alpha_1\left(2\lambda_2\theta_1 + \lambda_1\theta_2\right)\right]}{4\theta_1\theta_2\left[2\lambda_1\lambda_2\sigma\left(\frac{1}{2}\sigma - \alpha_1 + \alpha_2\right) - (\lambda_2\alpha_2 + \lambda_1\alpha_1)^2\right]}; \tag{A.76}$$

$$\widehat{\gamma}_1^{fbe} - \widehat{\gamma}_2^{fbe} = \frac{\left(k_A^2 - k_B^2\right)\left[\left(\lambda_1\alpha_1 + \lambda_2\alpha_2\right)\left(\lambda_2\theta_1 + \lambda_1\theta_2\right) + \sigma\lambda_1\lambda_2(\theta_2 - \theta_1)\right]}{4\theta_1\theta_2\left[2\lambda_1\lambda_2\sigma\left(\frac{1}{2}\sigma - \alpha_1 + \alpha_2\right) - (\lambda_2\alpha_2 + \lambda_1\alpha_1)^2\right]}. \tag{A.77}$$

Functions (A.75)-(A.77) have all the same vertical asymptote $\hat{\alpha}_1^{fbe}$, with:

$$0 < \hat{\alpha}_1^{fbe} \equiv \frac{-\lambda_2(\alpha_2 + \sigma) + \sqrt{\lambda_2\sigma(2\alpha_2 + \sigma)}}{\lambda_1} < \tfrac{1}{2}\sigma.$$

When $\alpha_1 < \hat{\alpha}_1^{fbe}$ holds, then the denominators in (A.75)-(A.77) are all positive. The numerator of the second term in (A.75) is positive because $\left(\frac{1}{2}\sigma - \alpha_1 + \alpha_2\right) > 0$ implies that $(\theta_2(\sigma + \alpha_2) - \alpha_1\theta_1) > 0$. Thus, $\widehat{\gamma}_1^{fbe} > \frac{1}{2}$ holds for $\alpha_1 < \hat{\alpha}_1^{fbe}$. In addition, since the numerator of (A.77) is positive, $\widehat{\gamma}_1^{fbe} - \widehat{\gamma}_2^{fbe} > 0$ is satisfied for $\alpha_1 < \hat{\alpha}_1^{fbe}$. Finally, the numerator of (A.76) is positive for $\alpha_1 < \tilde{\alpha}_1^{fbe}$, where:

$$0 < \tilde{\alpha}_1^{fbe} \equiv \frac{\lambda_2 (\sigma\theta_1 + \theta_2\alpha_2)}{2\lambda_2\theta_1 + \lambda_1\theta_2} < \hat{\alpha}_1^{fbe}.$$

Hence, from (A.76), $\widehat{\gamma}_2^{fbe} \geq \frac{1}{2}$ holds for $0 < \alpha_1 \leq \tilde{\alpha}_1^{fbe}$ whereas $\widehat{\gamma}_2^{fbe} < \frac{1}{2}$ for $\tilde{\alpha}_1^{fbe} < \alpha_1 < \hat{\alpha}_1^{fbe}$.

Our main results are summarized below:

Proposition 7 ***Efficient sorting in the richer specification.*** *(i) When firms are identical ($k_A = k_B$) they equally share the workforce of both types: $\widehat{\gamma}_1^{fbe} = \widehat{\gamma}_2^{fbe} = 1/2$ and $E_A^{fbe}(\theta) = E_B^{fbe}(\theta)$.*
(ii) When $k_A > k_B$ and $\alpha_1 < \frac{1}{2}\sigma$, then firm A hires a larger share of high-ability workers ($\widehat{\gamma}_1^{fbe} > \frac{1}{2}$) and the best workforce: $\widehat{\gamma}_1^{fbe} - \widehat{\gamma}_2^{fbe} > 0$ and $E_A^{fbe}(\theta) < E_B^{fbe}(\theta)$ hold.
(ii) If $\alpha_1 \in [0, \tilde{\alpha}_1^{fbe})$, $\widehat{\gamma}_1^{fbe} > \widehat{\gamma}_2^{fbe} > \frac{1}{2}$ holds irrespective of α_2.
(iii-a) $\widehat{\gamma}_1^{fbe}$ increases with α_1 and, as α_1 grows larger and gets closer to $\hat{\alpha}_1^{fbe}$, the corner solution $\widehat{\gamma}_1^{fbe} = 1$ is reached.
(iii-b) When α_2 is small, $\widehat{\gamma}_2^{fbe}$ decreases with α_1. Thus, as α_1 enters the interval $[\tilde{\alpha}_1^{fbe}, \hat{\alpha}_1^{fbe})$, $\widehat{\gamma}_2^{fbe} < \frac{1}{2}$ starts to hold. As α_1 grows larger and gets closer to $\hat{\alpha}_1^{fbe}$, first the corner solution with $\widehat{\gamma}_1^{fbe} = 1$ is hit and then the one with $\widehat{\gamma}_2^{fbe} = 0$.
(iii-c) When α_2 is large, $\widehat{\gamma}_2^{fbe}$ increases with α_1 but the function $\widehat{\gamma}_2^{fbe}$ is flatter than the function $\widehat{\gamma}_1^{fbe}$. Thus, as α_1 grows larger and gets closer to $\hat{\alpha}_1^{fbe}$, first the corner solution with $\widehat{\gamma}_1^{fbe} = 1$ is hit and then the one with $\widehat{\gamma}_2^{fbe} = 1$.

To summarize, in equilibrium, $\widehat{\gamma}_1$ is always increasing in α_1 but, if α_2 is high enough (and $\alpha_2 > \alpha_1$ holds), $\widehat{\gamma}_2$ switches from being decreasing to being increasing in α_1. This holds both for the market equilibrium and the efficient sorting. In different words, the functions describing marginal workers of high and low-ability in equilibrium ($\widehat{\gamma}_1^{*e}, \widehat{\gamma}_2^{*e}$) are shaped as their efficient counterparts ($\widehat{\gamma}_1^{fbe}, \widehat{\gamma}_2^{fbe}$). However, three different distortions of marginal workers sum up when solving for the market equilibrium (see the discussion below Proposition 3 in Section 4): profit maximization, firms' competition, and limited internalization of the positive externality generated by the premium for CfCA. And such distortions all imply that $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ react more to changes of α_1 and α_2 in the efficient allocation than in the equilibrium allocation. As a result, when we constrain our analysis to the parameters' configuration such that the efficient and the equilibrium allocations can be compared (imposing condition A.71), we always obtain an interior solution in equilibrium entailing $\widehat{\gamma}_1^{*e} > \widehat{\gamma}_2^{*e} > \frac{1}{2}$. Whereas, a corner solution can emerge in the efficient allocation as α_1 grows larger.

Let us move to the welfare analysis. Comparing Propositions (6) and (7), we obtain the following results about market distortions in this richer model.

As in the reduced-form model, $\widehat{\gamma}_1^{*e} < \widehat{\gamma}_1^{fbe}$ holds $\forall \alpha_1, \alpha_2$, meaning that the share of high-ability types hired by firm A is inefficiently low. In addition, when α_2 is small (and $\alpha_1 > \alpha_2$),

and H/peer effects/submissions/3 GEB/revision 2/paper/S86YNA01.wmf

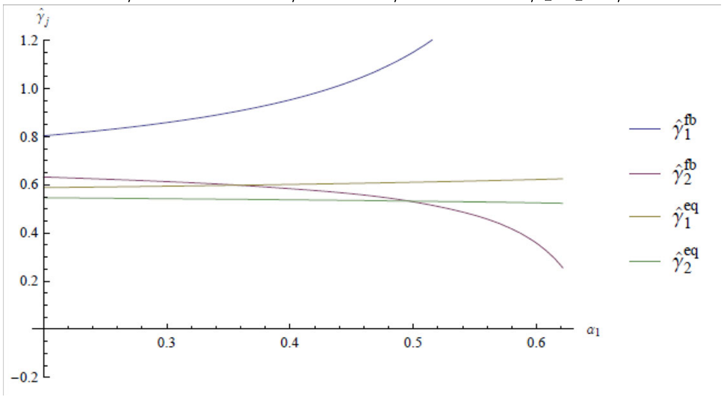

Figure 2: Marginal workers when $\alpha_1 > \alpha_2 > 0$: comparison between the efficient marginal workers, $\hat{\gamma}_j^{fb}$, and the equilibrium ones, $\hat{\gamma}_j^{eq}$. Parameters are $\alpha_2 = 0.2$, $\sigma = \theta_1 = 1.5$, $\theta_2 = 2$, $\lambda_1 = 1/2$, and $k_A - k_B = 2$.

$\widehat{\gamma}_2^{*e} < \widehat{\gamma}_2^{fbe}$ holds (as in the reduced-form model) for low values of α_1 whereas, for high values of α_1, the opposite holds and $\widehat{\gamma}_2^{*e} > \widehat{\gamma}_2^{fbe}$. This latter case, with α_2 small and α_1 sufficiently high, is the only situation in which the distortion in low-ability workers' sorting is different from the one in the reduced-form model. In different words, for α_2 small and α_1 sufficiently high, it would be efficient to decrease the share of low-ability types in firm A (and reach market segmentation with $\widehat{\gamma}_1 = 1$ and $\widehat{\gamma}_2 = 0$ eventually). However, firm A hires too many low-ability workers. This case is depicted in Figure 2 comparing marginal workers of both types in the efficient allocation and in equilibrium when α_2 is small. The figure shows that, for $\alpha_1 > 0.5$, $\widehat{\gamma}_2^{*e}$ lies above $\widehat{\gamma}_2^{fbe}$.

Finally, for α_2 large (and $0 < \alpha_1 < \alpha_2$), $\widehat{\gamma}_2^{*e} < \widehat{\gamma}_2^{fbe}$ always holds as in the reduced-form model. In this case, all distortions observed in the reduced-form model are confirmed in the richer model. In different words, for high values of α_1, it would be efficient to reach a monopoly with $\widehat{\gamma}_1 = \widehat{\gamma}_2 = 1$, but firm A continues to hire too few workers of both types. This case is depicted in Figure 3 comparing marginal workers of both types when sorting is efficient and in equilibrium when α_2 is large.

Our welfare analysis is summarized in the proposition below.

Proposition 8 ***Welfare analysis in the richer model.*** *(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 overall effect of the concern for coworkers' ability on total surplus is ambiguous. It benefits high-ability workers and firm B but it impairs low-ability workers and firm A.

(ii-a) Market sorting is inefficient because the share of high-ability workers employed by firm

and H/peer effects/submissions/3 GEB/revision 2/paper/S86YNA02.wmf

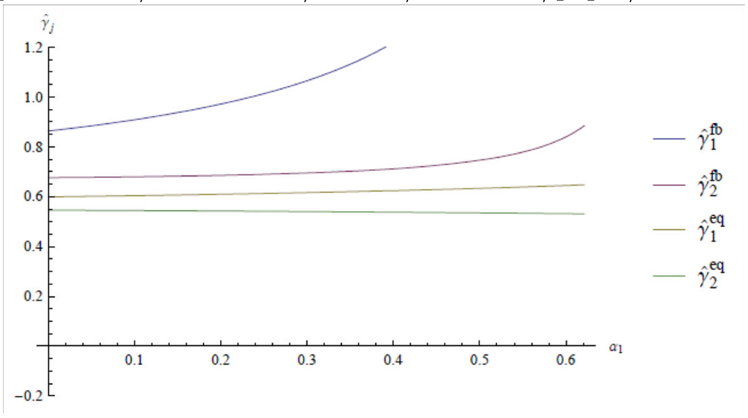

Figure 3: Marginal workers when $0 < \alpha_1 < \alpha_2$: comparison between the efficient marginal workers, $\hat{\gamma}_j^{fb}$, and the equilibrium ones, $\hat{\gamma}_j^{eq}$. Parameters are $\alpha_2 = 1.2$, $\sigma = \theta_1 = 1.5$, $\theta_2 = 2$, $\lambda_1 = 1/2$, and $k_A - k_B = 2$.

A is too low. In addition, when α_2 is high, the share of low-ability workers employed by firm A is always too low as well; see Figure 3. However, when α_2 is low, the distortion in sorting of low-ability workers depends on the magnitude of α_1; see Figure 2. Specifically, when α_1 is low, the share of low-ability workers employed by firm A is still too low. Conversely, when α_1 is sufficiently large, the share of low-ability workers employed by firm A becomes too high.

(ii-b) Under condition (A.71), an interior solution always emerges for both marginal workers in equilibrium. However, as α_1 grows larger and gets closer to $\hat{\alpha}_1^{fbe}$, a corner solution with either $\widehat{\gamma}_1 = \widehat{\gamma}_2 = 1$ or $\widehat{\gamma}_1 = 1$ and $\widehat{\gamma}_2 = 0$ would be optimal.