NICEP Working Paper: 2025-11

Discrimination spillovers, glass ceilings and pay gaps

Gianni De Fraja and József Sákovics

Nottingham Interdisciplinary Centre for Economic and Political Research

School of Economics, University of Nottingham, Sir Clive Granger Building,
University Park, Nottingham, NG7 2RD
ISSN 2397-9771

Discrimination Spillovers, Glass Ceilings and Pay Gaps[*]

Gianni De Fraja[†]
University of Nottingham and CEPR

József Sákovics[‡]
Universitat de les Illes Balears

May 2, 2025

Abstract

This paper is a theoretical analysis of the consequences of workplace discrimination. We prove that discrimination against a group at lower levels of the hierarchy affects the pay of members of the same group at higher levels, leading to a "pay gap" relative to non-discriminated workers. These spillovers in turn induce firms to alter the match between workers and jobs for the discriminated group, potentially leading to a "glass ceiling". The phenomenon can occur even in firms where "equal pay for equal jobs" appears to be adhered to. The explanation is based on the standard participation and incentive constraints: the need to compensate workers for the direct discrimination they suffer, to induce them to work, and the need to maintain pay differentials between job levels, to provide effort incentives. We end the paper showing that neither competition among workers, nor competition among firms for workers eliminates these spillovers.

JEL Numbers: J71, J70, K38, M52, M14, D82.

Keywords: Discrimination, Inequality, Minorities, Glass ceiling, Pay gap, Earnings differentials, Labour market bias, Principal agent, Mechanism Design.

[*]We would like to thank Carmen Matutes, Guido Friebel, Jack Fanning, Jonathan Thomas, Paul Seabright, Thomas De Fraja, and participants at a seminar at UIB for useful comments. We are grateful to the University of Nottingham for providing funding under project A7C134 which led to the development of this paper. Sákovics thanks the Spanish government for support through a Beatriz Galindo grant (BG20/00079) and grant PID2020-115018RB-C33 funded by MCIN/AEI/10.13039/501100011033.

[†]Nottingham School of Economics, Sir Clive Granger Building, University Park, Nottingham, NG7 2RD, UK, and CEPR, 33 Great Sutton Street, London EC1V 0DX, UK; email: `gianni.defraja@nottingham.ac.uk`.

[‡]Departament d'Economia de l'Empresa, Edifici Jovellanos, Universitat de les Illes Balears, Carretera de Valldemossa km 7.5, 07122 Palma de Mallorca, Spain; email: `j.sakovics@uib.eu`.

1 Introduction

This paper identifies discrimination spillovers as a significant factor affecting workers' pay, status, and well-being. Our motivation stems from the observation that workplace discrimination varies in intensity across different levels of an organisation's hierarchy. Our theoretical model shows that an organisation's optimal response to a pattern of direct discrimination faced by lower-level employees might be to let it spill over and trickle up to affect the pay of workers from the same group at higher levels, who may suffer little or no direct discrimination themselves. The "glass ceiling" and the "pay gap" may thus arise solely from differences in discrimination across different hierarchical levels.

There is ample evidence of pay and achievement gap, for example between men and women, which cannot be explained by observable characteristics, other, obviously, than gender (Gunderson, 1989; Wood et al., 1993; Bertrand, 2018). In her Nobel Lecture, Goldin (2024) attributes the gender pay gap to firms' practice "to disproportionately reward individuals who work long hours and who work particular hours" (Goldin, 2014, p. 1092): the willingness to work these long hours does not increase individual productivity, as it is "often the case that hours alone get rewarded", not ability or past success.[1] To the extent that women are less willing to apply for jobs with unpleasant hours, the pool of applicants for these jobs will include fewer qualified women, inevitably ensuring that even fair and impartial appointment and retention systems will end up with a gender imbalance.

Our theoretical analysis also attributes different outcomes to groups having unequal costs of working in unpleasant conditions. However, we show that this

[1]These are the "greedy" occupations, where "earnings are convex with respect to hours (Goldin, 2024, p. 1532)". This explanation of course complements the existing literature's which attributes the portion of the pay gap that is not due to human capital, measured by years and quality of education and accumulated labour market experience, to a variety of factors, ranging from a group's lower ability to bargain, or lesser desire to compete, to employers' differential promotion standards which may in turn be due to real or perceived differences in the groups' probability of leaving, to of course direct discrimination, implicit or blatant (Goldin, 2014, pp. 1093-94).

effect is even more pervasive and far reaching. The novelty of our contribution is to highlight that factors which affect job matches at the lower levels reverberate and amplify higher up in the hierarchy. The subtle nature of this spillover may make it difficult to detect: a pay gap between equally productive workers of different groups may emerge even when workers receive equal pay for equal jobs. It also makes it insidious to combat: even when the ablest members of a group may face no direct obstacles preventing them from accessing top jobs in their organisation, racial, gender, and other imbalances at lower hierarchical levels trickle up the ladder, with the obvious implication that any organisation wishing to eradicate these imbalances must fight discrimination and other obstacles to inclusion at *all* levels of the hierarchy. This paper's message to policy makers is that to erase pay gaps at the top they must tackle the rot in the foundations.

To obtain our results, we adapt a standard principal-agent model where employers design the compensation package to induce potential workers to apply for the job levels appropriate to their productivity. Higher level jobs are more productive but they require more effort, and so they have a higher utility cost, in the guise, for example, of higher stress and responsibility. This model is suitable to the study the hierarchical structure of organisations in view of two of its conclusions. First, all workers, except those at the very bottom of the hierarchy, obtain some rent, that is a utility from working which strictly exceeds their outside option, and, second, that higher ability workers select higher level jobs, requiring more effort but still leading to higher utility, as well as higher wages.

In order to analyse discrimination spillovers, we modify this textbook model in two key ways. First, we account for the reality that workers experience not only the disutility associated with the effort required to produce output, but also an additional utility cost unrelated to their performance. We assume that this additional utility cost may vary by job level and affects individuals differently based on their group identity. We discuss this utility cost in detail in Section 2.2, but, in brief, it captures the many aspects of workplace discrimination, from conscious or unconscious bias in appointments and promotions, to unnecessary long and unsociable hours, to co-workers' unkindness, incivility, and harassment. Second, we allow employers to offer tailored compensation packages to workers in different groups.[2] Even when forbidden by law from explicit discrimination, firms may in practice find ways to design compensation schemes so as to reward groups differently for jobs at the same level.[3] In large organisations with several career paths, firms may set differentiated pay scales for roles that, while equivalent in terms of responsibility and level, may attract distinct groups of workers based on their idiosyncratic preferences.[4]

These two modifications, together with our theoretical set-up of constant returns to scale, which we posit to abstract from technological constraints to employment, lead naturally to firms being willing to compensate workers who suffer additional utility costs due to their group identity. Firms wish to employ as many workers as there are available, up to the point where the marginal profit of employing an additional worker is zero. If a worker turns up at the firm's door who requires no compensation for discrimination and so is less expensive than an

[2]In Proposition 4 below, we provide conditions such that our results hold even when the firm allows workers from the discriminated group to choose from either pay scale, their own, or that intended for the non-discriminated groups.

[3]The response of society clearly depends on the perceived beneficiaries of such perks: as examples, compare free childcare and free golf club memberships, not to mention the type of corporate entertainment which Martin Scorsese depicted being offered by Jordan Belfort's stockbroking firm to its employees in the 2013 feature film *The Wolf of Wall Street*.

[4]A major UK court case illustrates. In Brierley-vs-Asda, the UK Court of Appeal ruled on March 26th 2021 that the UK subsidiary of Wal-Mart had acted unlawfully by offering differential pay scales to warehouse and shop-floor employees, thereby allowing the latter, who are more predominately female, to sue for compensation for the lower pay received over the years. Plausibly a firm offering different scales for, say, human resources and estate maintenance, might be engaging in the same type of discrimination. These examples justify the scepticism with which some view anti-discrimination legislation (Posner, 1989). Stylised facts and some evidence (De Fraja et al., 2019) suggest that professorial pay in universities tends to be higher in some academic disciplines, economics, finance, and computer sciences among them, than in others such as education or English literature, where female representation is higher (Fagan and Teasdale, 2021). This is consistent with implicit or indirect discrimination, though we are not aware of any instances where this has been claimed to be the case.

equally productive worker already employed by the firm, the firm will employ this new worker *as well as*, not *instead of*, the currently employed worker.[5] Separate modelling of monetary pay and targeted in-kind compensation is clearly possible, but it would lead to unrewarding complications, and so, for the sake of simplicity, we assume firms offer different pay scales to different groups. We do not view differential compensation as discriminatory behaviour in itself, but rather as a profit maximising response to direct discrimination against some workers.

The additional utility costs due to discrimination affect the compensation offered by the firm in the canonic way prescribed by the revelation principle of the mechanism design literature (Laffont and Martimort, 2002). The workers' information advantage – they know their idiosyncratic suitability to a given job level – forces employers to offer workers a contract which induces them to "self-select" into the job level chosen by the firm for them. They achieve this by offering them a pay package which they prefer to mimicking a "worse" type and choosing a job level lower down the hierarchy, where stress is lower and responsibilities are fewer. But, and here is our key insight, if there are additional reasons for members of one group to try to avoid jobs lower down the hierarchy, then a *lower* monetary inducement suffices to entice workers in a discriminated group into the higher level jobs. For them, a promotion from the shop floor to the office up-stairs will be good news even if it does not come with as good a pay rise as their co-workers', because it offers the chance to leave an environment where harassment is rife for those of certain genders, races, religions, sexualities, disabilities, and so on. In other words, discrimination at the lower levels of the hierarchy lets the firm compensate workers, at least in part, with a work environment in the higher parts of the hierarchy, where the disutility costs of workplace discrimination are lower. These workers are willing to accept a "pay gap" when they are promoted to a higher job level. On the

[5] Also note that, if anything, our benchmark set-up where firms are willing and able to compensate members of the discriminated group for their additional disutility at the workplace stacks the deck *against* the emergence of lower pay for them.

other hand, when direct discrimination is increasing in job level, it may make it too expensive for the firm to assign workers to their efficient job level. In that case the job assignment is distorted, with overqualified people filling some mid-level jobs. When this happens at the top, these distortions may generate a “glass ceiling”.

In our model, the firm takes the extent of discrimination faced by its workers as given. Since discrimination increases the cost of employing some workers, one might hold that profit maximisation would over time find ways to whittle away this cost increasing obstacle. There are two routes this may conceivably happen. The first is internal: since discrimination is, at least partly, caused by their employees and stakeholders, would not firms at least try to reduce its extent? As we show in a simple numerical example in Section 2.6, discrimination may lead to profit losses which are both small relative to the overall profit the firm makes, and a small fraction of the utility loss suffered by the employees. It follows that, even when eradicating discrimination would carry a small cost, essentially all of the benefit would accrue to the employees, and so a profit maximising firm may lack sufficient internal incentives to incur this cost.

The second route is via competitive pressure. If there were several firms, all operating with the same constant return to scale technology, would each not try to increase its own profit by attempting to attract discriminated workers employed elsewhere? This view is in line with the influential intellectual tradition in economics which views competition as a powerful mechanism shaping economic interactions (Becker, 1957). Therefore, in Section 3, we embed the isolated firm studied in Section 2 into a competitive labour market. We show there that the distortions and the group differences of the equilibrium pay scale we find in the first part of the paper are left in essence unaltered by competition. Since competition for workers does not eliminate the requirement to provide incentives for truthful revelation of workers’ type (Epstein and Peters, 1999), all pay scales are shifted parallelly up, thus maintaining any differences between groups that exist in the isolated firm. Our theoretical analysis tallies with the scarce empirical evidence we found (Winter-Ebmer, 1995; Cooke et al., 2019; Siddique et al., 2023), which indeed suggests a limited role for competition to lessen the effects of discrimination.

Our main contribution is to the literature on discrimination in labour markets and employment contracts (a recent excellent survey is Onuchic (2022)). The kind of discrimination we describe differs from statistical discrimination (Phelps, 1972; Aigner and Cain, 1977), whereby workers of one group are rewarded on the basis of the distribution of some relevant group characteristics, which may be endogenously determined to differ by group in equilibrium, even when *ex-ante* they are equal (Coate and Loury, 1993). It also differs from taste-based discrimination (Becker, 1957; Guryan and Charles, 2013), where employers irrationally dislike members of some groups. The self-selection of workers on the available pay scales determines their choice of effort, and hence their labour supply, and so our model provides a theoretical foundation to experimental work on the effect of pay differences (Breza et al., 2018), especially when explicitly discriminatory (Gagnon et al., 2025), on labour supply, and empirical work on the effort choice in the presence of discrimination (Glover et al., 2017). Our modelling of heterogeneous discrimination across hierarchical levels is reminiscent of Hurst et al. (2024), which studies the variation of racial discrimination across occupations. They reduce the dimensionality of the problem by looking at generic tasks, allowing for heterogeneity of discrimination across these. As these tasks are not carried out in the same firm, externalities across tasks do not arise, except possibly at the occupation selection level, which we abstract from. Theoretical economic analyses of pay gaps and glass ceilings are on the other hand notable for their paucity.[6] We believe to

[6]One exception is Grout et al. (2009), whose story, unlike the present paper, has group differences in productivity “at the heart” (p 2): the glass ceiling emerges because the groups differ in their outside options, and so firms find they can “promote women more ‘cheaply’ than men” (p 5). Firms find this in our model as well, but this is due to one group experiencing intense discrimination at lower job

be the first to explore the consequences of differences in discrimination at different levels of a hierarchy.

Our paper also speaks to the literature on identity and the economics of organizations (Akerlof and Kranton, 2005). At the core of this line of research is the theme, also central to our analysis, that monetary reward are but a part of what motivates employees to exert non-contractible effort: workers' self-image as jobholders also provides powerful work incentives. This viewpoint maps naturally into our discussion of discrimination in Section 2.2 as the component of job utility which is directly affected by social intercourse with other people, rather than being confined to physical or mental interactions with the material or the intellectual world.

The paper is organised as follows: in Section 2 we propose a theoretical model of an isolated firm whose workers are discriminated against. We derives the firm's profit maximising pay and employment schedule, and end by showing that with very simple and plausible functional forms the model can replicate stylised aspects of discrimination in practice. Section 3 embeds the model into an imperfectly competitive labour market, and a brief conclusion is presented in Section 4. An appendix collects the proofs.

2 The model

In this section, we study the equilibrium of a one-shot interaction between a profit maximising firm and a pool of workers.

2.1 Workers and jobs

The firm's output, sold at a unit price of 1, is given by the sum of its workers' output, that is, there are no externalities in production among workers.[7] There is

levels which makes its members willing to accept a lower pay increase for promotion to a level where discrimination is less severe.

[7] That is, a worker's output is independent of the type and number of any other workers employed by the firm. This allows us to study the firm's problem of designing job scales in isolation, and differs from models of production in teams (Costrell and Loury, 2004; Herkenhoff et al., 2024), where co-

a continuum of job levels $\ell \in \mathbb{R}_+$, and workers in higher level jobs produce more output. Output is contractible, and we normalise it to be equal to the job level: ℓ denotes both the job level and the output of a worker at that level. There is a continuum of workers, who differ in their privately observed "type", denoted by θ, representing their idiosyncratic utility cost of effort (higher type, higher cost). Workers also differ according to their observable identity, their "group", denoted by $g \in G$. Groups partition the set of workers, that is, each worker belongs to one and only one group. While the notation allows for any number of groups, little is lost by focusing on the specific case of two groups, only one of which suffers from workplace discrimination. We normalise θ to be distributed in $[0,1]$, according to the distribution function $\Phi(\theta)$, the same for every group,[8] with $\phi(\theta) = \Phi'(\theta) > 0$. As standard (Laffont and Martimort, 2002), we impose monotonicity of the (inverse of the) hazard rate:[9] $h(\theta) = \frac{\Phi(\theta)}{\phi(\theta)}$ satisfies $h'(\theta) > 0$.

The workers of group g who are not employed obtain a common reservation utility, u_g.[10] When employed, workers' utility is measured by a quasi-linear utility function, given by the difference between their wage and the monetary equivalent of the disutility associated with having to work. We divide the utility cost of working into two conceptually distinct parts, $\psi(\ell, \theta)$ and $\delta_g(\ell)$, so that the utility enjoyed by a type $\theta \in [0,1]$ worker from group $g \in G$, working at job level $\ell \in \mathbb{R}_+$, and

workers' characteristics affect a person's outcomes.

[8]The distribution of θ could differ among groups, but this has in general allocative consequences (De Fraja, 2005), which might confound the role played by discrimination. The assumption of the same function Φ for all groups, therefore focuses our analysis on the consequences of discrimination. It also eliminates any possible influence on pay differences among groups of differences in the distribution of preference parameters, be they for commuting, flexibility, and the like (Wiswall and Zafar, 2018; Petrongolo and Ronchi, 2020; Cook et al., 2021). The reduction in the importance of physical strength for productivity implies that differences in the distributions of θ among groups are nowadays much smaller than they might have been in the past.

[9]This is satisfied by practically all the distributions functions routinely used in theoretical and empirical work (Bagnoli and Bergstrom, 2005).

[10]While the reservation utility may in general be type dependent, as long as it decreases slower with type than the equilibrium utility in employment, only its value for the marginal worker is relevant, so it is innocuous to assume that it is constant in θ.

receiving salary $w > 0$ is given by

$$w - \psi(\ell, \theta) - \delta_g(\ell). \tag{1}$$

2.2 Formalising workplace discrimination

In this subsection we provide the rationale for keeping the δ component of a worker's disutility conceptually separate from ψ, and we detail its interpretation. While not strictly necessary, intuition is assisted by associating to ψ and δ distinct sources of disutility. We view ψ as the classical cost of effort, intrinsically linked to the worker's job level: stress boredom and frustration, feelings of inadequacy, foregone leisure, exertion of mental or physical effort, and so on. These are unaffected by other people's attitude or behaviour. In contrast, δ captures discrimination: the disutility associated with the inevitable human interactions with other members of the organisation, fellow co-workers, superiors, and subordinates, as well as with clients, customers, and suppliers. This interpretation justifies naturally our modelling choices for the arguments of these two functions: we write $\psi(\ell, \theta)$ to indicate that, conditional on the job level, the cost captured by ψ depends only on the worker's "ability" θ, not the worker's group identity, g. Similarly, writing δ as $\delta_g(\ell)$, captures our key modelling innovation: the idea that, conditional on the worker's group identity, the cost it measures varies with the job level, and that conditional on ℓ, it is independent of the worker's ability θ.

Thus, we view discrimination, as measured by $\delta_g(\ell)$, as a black box containing everything that creates a utility cost for a worker in job level ℓ which depends only on group identity, that is on job irrelevant observable characteristics common to all workers in the group. To fix ideas, and without being exhaustive, three separate sources are obvious contributors to δ: the appointment and promotion process, the organisational culture, and interactions with fellow workers, superiors, subordinates, and suppliers and clients.

In our model, the firm posts a set of pay scales and hires anyone who turns up at its gates wishing to work for a contract chosen from those scales. Of course, in practice appointments are the conclusion of a selection process where human input and subjective preference are pervasive, in the selectors (Schaerer et al., 2023), the candidates (Exley and Kessler, 2022), and their referees (Eberhardt et al., 2023). Among the large variety of formalisations to include irrational discriminatory preferences in our model, a very simple one has the appointment process as a two stage procedure: once an offer is accepted in the first step, details of the work environment are hammered out in the second step. Individual workers have idiosyncratic preferences over these details, in the manner explored convincingly in horizontal product differentiation (Beath and Katsoulacos, 1991): in a chain store, or a hotel group, workers rank stores or hotels in different ways, perhaps depending on where their partners or friends and relations live. Similarly for different sections of a factory or departmental offices with different degrees of telephone interaction with clients and suppliers, or different potential for remote working. In academia, the teaching pattern and timetable affects the job satisfaction, but is rarely an integral part of contractual negotiations. A formal protocol that captures real life informal negotiations for the second stage could run as follows. There are n positions available and n appointees at each level ℓ, and a worker in group g derives disutility $\delta_{g,j}(\ell)$ from being in position $j = 1, \ldots, n$. Appointees choose in randomly allocated turns the position they wish to have. With probability $q_g(\ell)$, the firm (the line manager for the chosen position) has a bias, rational or irrational, conscious or unconscious (Hebl et al., 2018; Bertrand and Duflo, 2017), and will veto the appointment to this position of a worker of group g, forcing the appointee to choose another position, unless it is the last one.[11] Hence a type θ worker in

[11] An additional protocol may be formalised for the eventuality of an appointee selected in the first stage ending up with no match, to capture the possibility of workers of some groups not being

group g may be unable to work at the chosen position and so be forced to accept a different one, say a store less conveniently placed for where they live, a section of the production line which is too hot, or too smelly, or too noisy for them, or an office where there is too little or too much interaction with clients, or a position where there is too much or too little remote working, or be landed with a very inconvenient timetable. The expected disutility of this job level will clearly be increasing in $q_g(\ell)$.

The second component of δ, the "long hour culture" (Goldin, 2014) of the organisation, is easily formalised as a requirement to spend unproductive hours in a specified place, and to the extent that the opportunity disutility cost of spare time varies by group, it affects differently different groups. It is widely accepted that organizational culture can be influenced by management, but it cannot be controlled (Hermalin, 2001); even when it can be affected, any change is likely to be meaningful only in the long run.[12]

Lastly, a recent literature[13] highlights how the incivility or toxicity of the work environment, that is, peers', superiors', and subordinates' comments and behaviour, should be viewed as discrimination, targeting characteristics that are almost always group specific: gender, ethnicity, sexual orientation, disability, and so on. Its very nature is to affect groups differently, and so fits well into our definition of δ_g.

A stylised example may help clarify the distinction between ψ and δ. Consider

---

appointed at all with positive probability and the consequent waste of time and effort in engaging into a new job search.

[12]This line of thought tallies with the sociological approach of Small and Pager (2020) on "institutional discrimination".

[13]This follows Cortina's seminal contribution (2008), and identifies the toxicity of the work environment, for example the prevalence of sexual harassment (Folke and Rickne, 2022), or the extent to which bullying or racists (Kern and Grandey, 2009), sexist (Cortina et al., 2013), or homophobic (Di Marco et al., 2018) remarks and jokes are tolerated or encouraged, as an important aspect of a job which affects different groups differently, and is therefore discriminatory. Kabat-Farr et al. (2020) discuss recent literature on selective incivility in the workplace; they note that "this selective incivility may provide an explanatory mechanism for the lower rates of White women and women of color in the upper echelons of organizations".

two sources of disutility for a female chief financial officer (CFO):[14] (i) obstacles and difficulties in the signing of a major contract, and (ii) a colleague's unwanted flirting. The former contributes to $\psi(\theta, \ell)$: it depends on the job level as it stands to reason that it would cause similar stress and frustration to a male board member of the same type θ, but would cause little concern to a checkout operator in the same organisation, be they male or female. Unwanted flirting, on the other hand, depends very much on the group: if it happens at all to a male board member, it would in most cases leave him indifferent, but would likely annoy a female checkout operator just as it does the female CFO. As well as with the group it also varies with the position in the hierarchy, not because it is perceived differently by a CFO and a checkout operator, but because the frequency and obnoxiousness of any sexual attention are likely to differ substantially in these two work environments.[15] Note that it is not the "firm", its owners and directors, that is racist, sexist, or homophobic. Rather, it is the firm's employees, whose behaviour lowers the job quality for fellow employees.

2.3 Technical preliminaries

To impose some structure on our disutility functions, we normalise θ to indicate types with higher cost of exerting a given effort level, so that the disutility of work increases in it: $\frac{\partial \psi(\ell,\theta)}{\partial \theta} > 0$. It is also plausible that a higher job level/output should require a higher cost of effort: $\frac{\partial \psi(\ell,\theta)}{\partial \ell} > 0$. These features imply that, in equilibrium, more productive[16] workers are placed in higher level jobs. Finally, it is also natural

[14] CFO is a shorthand for "person whose type θ is such that their optimal job assignment in the firm's hierarchy is at levels corresponding to CFO". And similarly for "checkout operator", the other actor in our example.

[15] As with the first component of δ, there are several plausible ways to formalise this component: one possibility would be for it to depend on the proportion of workers of the various groups at job level ℓ: readers familiar with the television series *Mad Men* can visualise Peggy Olson's experience as she moved from the secretarial pool to the managerial level.

[16] As the disutility cost of a job increases in the worker's type, a lower θ worker can perform the same functions as a higher θ worker in a less costly, that is more efficient, manner, and in this sense lower type workers are more "productive".

that for a given increase in type, the cost of effort increases by more in a higher level job: $\frac{\partial^2 \psi(\ell,\theta)}{\partial\theta\partial\ell} > 0$. We capture all this in a simple way by positing that ψ is an increasing and convex function of the *sum* of the type and the job level: $\psi(\ell,\theta) = \psi(\ell+\theta)$. In addition, as standard in similar set-ups (Laffont and Tirole, 1993, p. 66), we take ψ to have a non-negative third derivative: this last condition ensures global concavity of the firm's problem, ruling out the need to employ random mechanisms. Formally, for every $\ell \in \mathbb{R}_+$, for every $\theta \in [0,1]$, we have: $\psi'(\ell+\theta) > 0$, $\psi''(\ell+\theta) > 0$, $\psi'''(\ell+\theta) \geq 0$.

We impose mild restrictions on $\delta_g(\ell)$: it is non-negative, and, while it can be concave, to ensure that the equilibrium job level matching be continuous in θ, we require the derivative of $\delta_g(\ell)$ not to decrease too fast as the job level increases. We formalise these assumptions as follows. Let $L \subset \mathbb{R}_+$ denote the set of job levels available in the firm.

Assumption 1. *The functions δ_g are non-negative, twice differentiable, and satisfy, for every $\ell \in L$, for every $\theta \in [0,1]$, and for every $g \in G$:*[17]

(i) $\delta_g'(\ell) > -\psi'(\ell+\theta)$;

(ii) $\delta_g''(\ell) > -\left(\psi''(\ell+\theta) + h(\theta)\psi'''(\ell+\theta)\right)$.

Note that Assumption 1 allows $\delta_g'(\ell)$ and $\delta_g''(\ell)$ to take negative values. Next, we rewrite the δ_g functions in a manner that will be useful for comparing different patterns of discrimination later on.

Let $L_g^m = \left\{\ell \in L | \delta_g(\ell) \leq \delta_g(\ell') \forall \ell' \in L\right\}$ and let $\ell_g^m = \min\left\{\ell \in L_g^m\right\}$. That is, ℓ_g^m is the lowest job level where $\delta_g(\ell)$ takes its lowest value. Then we can rewrite $\delta_g(\ell)$ as

$$\delta_g(\ell) = \alpha_g + \beta_g f_g(\ell), \tag{2}$$

[17] Assumption 1 avoids corner solutions, bunching and "holes" in the job level-wage mapping, which, while interesting from a theoretical viewpoint (Mussa and Rosen, 1978; Maskin and Riley, 1984), would distract from the focus of our analysis. Note that Assumption 1.(ii) is implied by $\delta_g''(\ell) > -\psi''(\ell+\theta)$, given the assumption that $\psi'''(\ell+\theta) \geq 0$ posited above.

where $\alpha_g = \delta_g(\ell_g^m)$, $\beta_g = 1$ and $f_g(\ell) = \delta_g(\ell) - \delta_g(\ell_g^m)$. Note that f_g is a function satisfying $f_g(\ell_g^m) = 0$, and $f_g(\ell) > 0$ for $\ell \notin L_g^m$.

Rewriting $\delta_g(\ell)$ as (2) provides two parameters $\alpha_g, \beta_g \in \mathbb{R}_+$ that lend themselves naturally to comparative statics analysis: an increase in α_g is a uniform worsening of the conditions for all workers in group g, while an increase in β_g is a worsening of the their job environment which becomes progressively more severe the higher $\delta_g(\ell)$ is, but leaves the discrimination suffered by group g workers at job level(s) in L_g^m unchanged. Thus, we will refer to α_g as group discrimination, and to β_g as job-level sensitivity of discrimination.

2.4 The optimal pay scales

The firm's problem is one of mechanism design. It chooses the pay scales it offers to its potential workers to maximise its profit. As group identity is observable, the firm may offer different scales to different groups, as in (Stiglitz, 1973, p. 294ff.). Having seen these scales, each worker chooses a point on the scale they are offered, or to remain unemployed. As discussed in the introduction, this may happen even where the law forbids explicit differential treatment based on irrelevant characteristics.[18]

To determine the optimal set of pay scales $\omega_g^*(\ell)$, we invoke the revelation principle (Myerson, 1979) and model the game as a direct revelation mechanism. That is, we suppose that, having committed to job level and wage functions of the reported type, $\ell_g(\theta)$ and $w_g(\theta)$ the firm "asks" workers to report their type θ. It chooses these schedules to maximise its profit, subject to the participation constraint that requires employed workers to have a weakly higher utility than

[18]See footnote 4 for an example where a firm attempted to offer group specific pay scales. As a mirror example, affirmative action policies in university admissions are often allowed, provided they are not explicit: in 1978 the US Supreme Court upheld the University of California policy on university student admissions which was formally open to all applicants, while designed to help minority ones, but, in 1995, struck down the policy designed by the School of Law at the University of Maryland, which explicitly favoured certain identifiable groups. This was reaffirmed in June 2023, when the court ruled that giving consideration to race at in the admission process violates the fourteenth amendment to the Constitution, as claimed by Students for Fair Admissions against Harvard and the University of North Carolina.

their reservation utility, and to the truth telling constraint that no worker has an incentive to report a θ different from their own.

We begin the characterisation of the optimal direct revelation mechanism, that is $\ell_g^*(\theta)$ and $w_g^*(\theta)$, by noting that, given job level and wage schedules $\ell_g(\theta)$ and $w_g(\theta)$, the utility of a worker from group g as a function of their truthfully reported type is given by:

$$U_g(\theta) = w_g(\theta) - \psi(\ell_g(\theta) + \theta) - \delta_g(\ell_g(\theta)). \tag{3}$$

We derive the implications of the truth-telling, or incentive compatibility, constraint first. The proofs of all the results are in the Appendix.

Lemma 1. *Incentive compatibility requires*[19]

$$\dot{\ell}_g(\theta) \leq 0, \tag{4}$$

$$\dot{w}_g(\theta) = \dot{\ell}_g(\theta)\left[\psi'(\ell_g(\theta) + \theta) + \delta_g'(\ell_g(\theta))\right]. \tag{5}$$

According to Lemma 1, to induce truth-telling, two conditions need to be met. First, the job assignment must rank workers by the reverse of their cost of effort θ. Second, the rate of change in the wage function must be adjusted to account for the change in disutility. Notice that the term in square brackets in (5) is positive by Assumption 1.(i), implying that lower θ workers will be paid higher salaries, to go with their higher level jobs.

The proof of Lemma 1 also shows that lower θ workers enjoy higher utility: this is natural, as a lower type could always mimic a higher one, and receive strictly higher utility than the latter. Consequently, the participation constraint for group g is given by the "individual rationality" requirement that the highest type willing to seek employment, $\bar{\theta}_g$, is just indifferent between doing so and receiving the

[19] We denote with a dot over a function of θ its derivative with respect to θ.

reservation utility:

$$U_g(\bar{\theta}_g) = u_g. \tag{6}$$

If the firm employs all type θ workers from group g in $[0, \bar{\theta}_g]$, then its profit from employing them is given by

$$\pi\left(\ell_g(\theta), w_g(\theta)\right) = \int_0^{\bar{\theta}_g} \left(\ell_g(\theta) - w_g(\theta)\right) \phi(\theta) \mathrm{d}\theta. \tag{7}$$

We are now ready to characterise the optimal policy.

Theorem 1. *The job level schedule $\ell_g^*(\theta)$ that maximises (7) subject to* (4), (5), *and* (6) *satisfies*

$$1 = \psi'\big(\ell_g^*(\theta) + \theta\big) + h(\theta)\psi''\big(\ell_g^*(\theta) + \theta\big) + \delta_g'\big(\ell_g^*(\theta)\big). \tag{8}$$

Moreover, the types in group g employed at the profit maximising solution are $[0, \bar{\theta}_g]$, where $\bar{\theta}_g$ is given by:

$$\ell_g^*(\bar{\theta}_g) = u_g + \psi\big(\ell_g^*(\bar{\theta}_g) + \bar{\theta}_g\big) + h(\bar{\theta}_g)\psi'\big(\ell_g^*(\bar{\theta}_g) + \bar{\theta}_g\big) + \delta_g\big(\ell_g^*(\bar{\theta}_g)\big). \tag{9}$$

The proof of this result also establishes that the equilibrium $\ell_g(\theta)$ schedule is *strictly* decreasing. We state this as a separate corollary.

Corollary 1. *There is no pooling. Higher types are assigned to strictly lower level jobs:*

$$\frac{\mathrm{d}\ell_g^*(\theta)}{\mathrm{d}\theta} < 0. \tag{10}$$

From (8) it is clear thhttps://www.gruppoindia.it/chi-siamo/at the presence of discrimination affects the job matching only via its derivative, and therefore it does not depend on the discrimination suffered at a specific job level, rather on the *relative* discrimination as compared to neighbouring job levels. A further implication is that any difference in job allocation of members of group g relative to a non-discriminated group is a consequence of *heterogeneous* discrimination: a parallel shift of the $\delta_g(\ell)$ function, that is the same change in discrimination of all workers in group g, does not change their job allocation. In contrast, by (5), the marginal worker is determined solely by the discrimination directly suffered, and therefore it is affected by a shift in $\delta_g(\ell)$.

Once we have the optimal job level schedule, the corresponding wage schedule can be derived from the incentive compatibility constraint by integrating (5), using (6) in combination with (3) as the boundary condition:[20]

Corollary 2.

$$w_g^*(\theta, \bar{\theta}_g) = u_g + \psi\big(\ell_g^*(\theta) + \theta\big) + \delta_g\big(\ell_g^*(\theta)\big) + \int_\theta^{\bar{\theta}_g} \psi'\big(\ell_g^*(x) + x\big)\mathrm{d}x. \tag{11}$$

There are four components of the optimal wage, as highlighted in (11): (i) compensation for giving up the reservation utility, (ii) compensation for the effort required at the job level chosen, (iii) any additional compensation due to discrimination affecting workers of group g at the chosen job level, and (iv) the information rent (see, for example, Laffont and Martimort, 2002), the incentive required to induce truth telling. This last term is the discrimination spillover:[21] its value for a type θ worker depends on the job assignments of all workers with a higher type than θ in group g, which in turn depends on the shape of the function δ_g, as we established in Theorem 1.

We can now translate the optimal direct revelation mechanism back into a pay scale. Let $\Theta_g^*(\ell)$ be the inverse function of $\ell_g^*(\theta)$, which Corollary 1 ensures to be invertible. Then

$$\omega_g^*(\ell) = w_g^*\big(\Theta_g^*(\ell), \bar{\theta}_g\big), \qquad \text{for} \quad \ell \in \Big[\ell_g^*(\bar{\theta}_g), \ell_g^*(0)\Big], \tag{12}$$

[20] We highlight in (11) the dependence of the wage on the type of the marginal worker.

[21] In a related context, (Kawamura and Sákovics, 2014) investigate the spillovers of equal treatment for mid-productivity workers onto personalised wages at higher levels.

is the firm's profit maximising choice of the pay scale to be offered to workers in group $g \in G$.

2.5 The consequences of discrimination

In this subsection we study how differences in discrimination among and within groups affect the equilibrium outcome, including the relative position of workers in different groups within the firm.[22] This is where the convenience of writing $\delta_g(\ell)$ as we did in (2) comes into play. We called α_g group discrimination: changes in its value capture changes in discrimination that affects all members of a group equally. Conversely, changes in β_g, the job-level sensitivity of discrimination, affect the relative discrimination between job levels.

We begin with the study of how discrimination affects the allocation of workers to jobs. For completeness, we include the groups' reservation utility as well, but the main focus is on differences in discrimination.

Proposition 1. *(i) Changes in α_g and u_g leave ℓ_g^* unchanged.*

(ii) $\frac{\mathrm{d}\ell_g^*(\theta)}{\mathrm{d}\beta_g} \gtreqless 0$ according to whether $\delta_g'\big(\ell_g^*(\theta)\big) \lesseqgtr 0$.

Thus, as we have anticipated in our discussion of Theorem 1, if a worker remains employed, a change in group discrimination has no effect on their job allocation, and neither does a change in reservation utility. This is because the incentive constraint depends on the difference in payoff between truth-telling and mimicking a different type and these are unaffected by changes in the intercept of the utility function or in the outside option.

On the other hand, following an increase in the job-level sensitivity of discrimination, a type θ worker will be working in a higher (lower) level job depending on whether the slope of the discrimination function is negative (positive) at their original job level. To see why this should be the case, note that an increase in β_g

[22]See Costrell and Loury (2004) for an analysis of the role of job assignments for the distribution of earnings.

changes the relative desirability of jobs close to $\ell_g^*(\theta)$, and thus changes the cost of inducing truth telling by type θ. When, say, $\delta_g'\big(\ell_g^*(\theta)\big) > 0$, an increase in β_g makes the jobs just below $\ell_g^*(\theta)$ relatively better than $\ell_g^*(\theta)$, since their discrimination increases by less than it does for $\ell_g^*(\theta)$. This, in turn, increases worker θ's temptation to misreport their type, and so the firm can reduce the cost of inducing truth telling by worker θ by allocating them to a lower job level.

From Proposition 1 we can immediately determine how discrimination affects the job level assigned to the most productive worker, type $\theta = 0$. Define group $0 \in G$ as a group whose members suffer no discrimination at any job level: $\delta_0(\ell) = 0$, for every $\ell \in \mathbb{R}_+$.

Corollary 3. *(Glass ceiling). The most productive discriminated worker has a lower (resp. a higher, resp. the same) level job than the most productive non-discriminated worker if and only if $\delta_g'(\ell_0(0)) > 0$ (resp. $\delta_g'(\ell_0(0)) < 0$, resp. $\delta_g'(\ell_0(0)) = 0$).*

The corollary identifies that the main determinant of the most able worker's job assignment, that is whether a glass ceiling is present, is the slope of the discrimination function δ_g near the top. If $\ell_0(0)$ is the job level assigned to the most productive worker in group 0, then the glass ceiling for the discriminated group arises if (and only if) job level $\ell_0(0)$ entails more discrimination for members of group g than job levels immediately below it. A possible way to map this formal corollary into real life is to consider that $\delta_g'(\ell_0(0)) > 0$ "translates" into the observation that the public scrutiny and sniping by envious rivals are more severe for the CEO than for other board members, not to mention suspicions of having being promoted in abeyance to DEI policies. A different, but not alternative scenario leading to $\delta_g'(\ell_0(0)) > 0$ at the top is that a CEO is required to work proportionally longer and more unsociable unnecessary hours than other board members (Goldin, 2014). Profit maximisation may induce the firm to design pay scales such that the compensation offered to a type 0 member of a discriminated group to select themselves into the CEO level is insufficient to compensate them for the disutility cost of the increased scrutiny and gossip attached to that level, and induce them to choose to remain a board member instead.[23]

Next, we turn to the overall employment in each group. This is determined by the marginal worker's type, obtained from (9).

Proposition 2. *The marginal worker's type $\bar{\theta}_g$ is reduced by increases in u_g, or α_g; if $\ell_g^*\left(\bar{\theta}_g\right) \notin L_g^m$, then it is also reduced by an increase in β_g.*

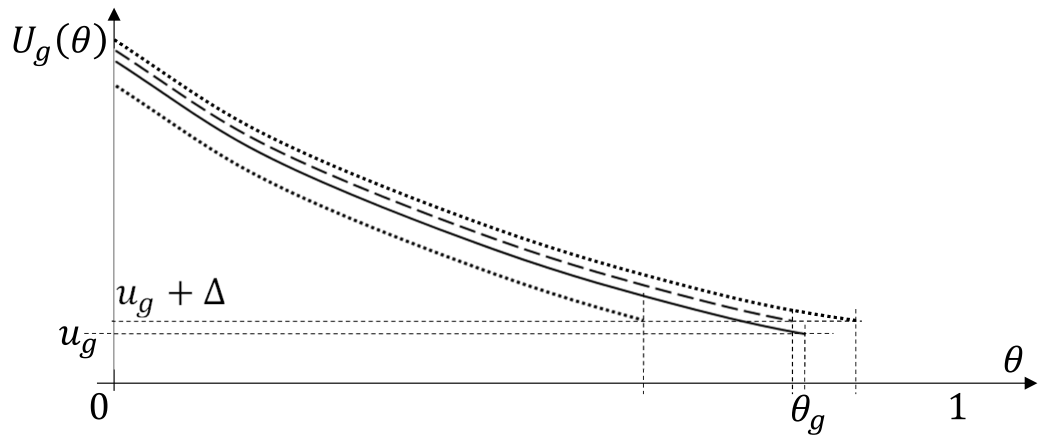
That is, discrimination against a group reduces the total employment of its members. This is relatively straightforward: by definition, the marginal worker, $\bar{\theta}_g$, is indifferent between working and not working. An increase in any of the three exogenous parameters, α_g, u_g, or β_g, tilts the balance either by making not working more attractive, the case of an increase in u_g, or by making working less attractive, increases in both α_g and β_g. As output is unchanged, the firm, which prior to the increase was also indifferent between employing and not employing the $\bar{\theta}_g$ worker, will not be willing to increase the wage to keep them employed.

In fact, the marginal employed worker is determined taking into account that increasing the marginal worker's pay not only increases the cost of employing them, but it also leads to higher informational rents to inframarginal workers (see (11)). Consequently, if any of α_g, β_g, or u_g increases, employment is lowered, not just because the marginal worker needs to be paid more, but also because increasing their pay increases everybody else's pay as well.

We continue the analysis with the study of the consequences of discrimination on the equilibrium salary scale and overall workers' utility. We begin with an increase in the reservation utility u_g and of an increase in the group g discrimination, α_g. By Proposition 2, these have identical effects on the the type of the marginal

[23]Pascual-Fuster et al. (2024) report evidence of female board members suffering this fate in Spanish firms during the period 2004-2012.

Figure 1: Utility as a function of type



Note: The solid curve is the utility received by a type θ worker at the firm's optimal schedule. When u_g increases, the utility curve shifts in a parallel way, and may in principle move to any of the shown positions: Propositions 2 and 3 show that it moves to a position similar to the dashed line.

worker, which in turn influences the information rents of all inframarginal workers as it is the upper limit of the integral in (11).

Proposition 3. *(i) An increase in u_g or α_g lowers* information rent *for all the workers. (ii) It increases* wage $w_g^*(\theta, \bar{\theta}_g)$ *by less than its full amount:* $\frac{\mathrm{d}w_g^*(\theta,\bar{\theta}_g)}{\mathrm{d}u_g} = \frac{\mathrm{d}w_g^*(\theta,\bar{\theta}_g)}{\mathrm{d}\alpha_g} \in (0,1)$. *(iii) An increase in u_g increases the* utility *of employed workers of type $\theta \in [0, \bar{\theta}_g)$, $U_g(\theta)$, by less than the increase in u_g, and (iv) an increase in α_g reduces the* utility *of all employed workers in group $g \in G$.*

That is, a uniform increase in discrimination leads to a higher wage of all workers, but this improvement falls short of the amount that the disutility cost of discrimination increases by. While workers are fully compensated for the increase in direct discrimination they suffer, the information rent they can collect shrinks with the reduced employment, so overall, they lose out. An analogous argument holds for an increase in reservation utility u_g. Figure 1 may provide some intuition for Propositions 2 and 3. If the reservation utility increases, two things happen: workers need to be paid more to be attracted, what increases their pay, but at the same time, because some workers do not stay in the labour market, the information rent of the workers who do stay is lower, as they have fewer workers "below" them. The balance of these two effects on workers' utility is in principle ambiguous, and Figure 1 depicts how both alternatives could in theory happen. The solid curve is the utility schedule when reservation utility is u_g and employed workers are at their equilibrium job levels, $\ell_g^*(\bar{\theta}_g)$. If the reservation utility increased to $u_g + \Delta$, the type $\bar{\theta}_g$ workers would no longer be willing to work at their current salary. The salary schedule would need to change: by Proposition 1, it shifts in a parallel fashion, but it could move to any of the three dashed or dotted positions in the picture; it could even, in a knife edge case, stay where it is, just be shortened. Proposition 3.(iii) shows that it shifts up, and, since $\bar{\theta}_g$ (who now earns $u_g + \Delta$) decreases, as shown in Proposition 2 , it does so by less than Δ.

We finally move to the effects on the salary schedule of changes in the job-level sensitivity of discrimination, as measured by β_g. Differentiating (11) we obtain

$$\begin{aligned}\frac{\mathrm{d}w_g^*(\theta,\bar{\theta}_g)}{\mathrm{d}\beta_g} &= \left(\psi'\left(\ell_g^*\left(\theta\right)+\theta\right)+\beta_g f'\left(\ell_g^*\left(\theta\right)\right)\right)\frac{\mathrm{d}\ell_g^*\left(\theta\right)}{\mathrm{d}\beta_g}+f_g\left(\ell_g^*\left(\theta\right)\right)\\ &\quad+\int_\theta^{\bar{\theta}_g(u_g)}\psi''\left(\ell_g^*\left(x\right)+x\right)\frac{\mathrm{d}\ell_g^*\left(x\right)}{\mathrm{d}\beta_g}\mathrm{d}x+\frac{\mathrm{d}\bar{\theta}_g\left(u_g\right)}{\mathrm{d}\beta_g}\psi'\left(\ell_g^*\left(\bar{\theta}_g\right)+\bar{\theta}_g\right). \quad (13)\end{aligned}$$

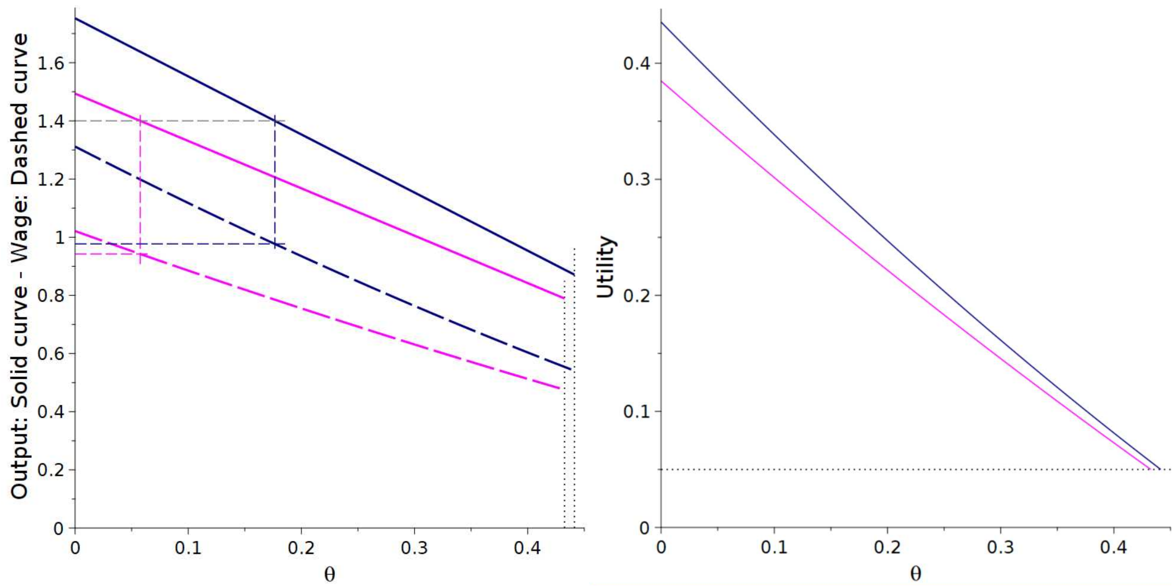
The sign of the change in wage of type θ as a result of an increase of the job-level sensitivity of discrimination is in general ambiguous. To see this, note that both the first and the second line of (13) can have either sign. The first line captures the variation due to the change in the compensation for the costs incurred: the first factor of its first term on the first line is positive, by Assumption 1i, so its sign is that of the change in the job level assigned. The sign of the latter depends on the derivative of the δ_g function at θ, by Proposition 1.(ii). The second term is non-negative, by definition. In the second line of (13), which represents the variation of the information rent, the sign of the integral term depends again on the derivative of the job-assignment function, but now at all the types above θ, rather than at θ itself. Note that this is a pure spillover effect. Finally, the last term is non-positive, as shown by Proposition 2. In sum the sign of the expression depends in general on the precise functional forms. In Section 2.6 we present examples with simple functional forms to show how the the firm's optimal choices vary with some of the parameters and to obtain a characterisation of the wage effects.

Even though they are compensated for the direct discrimination suffered, it is not immediately obvious that all members of the discriminated group prefer their own pay scale to those available to non-discriminated workers. Nonetheless, our final result shows that, if discrimination is high enough, then this is the case.

Proposition 4. *There exists $m \geq 0$ such that if $\alpha_g > m$, then members of group g strictly prefer a contract on the pay scale designed for group g to any contract designed for group 0.*

To end this section, we note that we have taken workplace discrimination as an immutable fact. As discrimination is a pure cost, the disutility it causes to some workers benefiting no one, would the firm not try, were it possible to reduce it in order to reduce the cost of compensating discriminated workers? To formalise this possibility, one route would be to posit a set of "discrimination reduction" increasing functions $\rho_g : \mathbb{R}_+ \to [0,1]$, with $\rho_g(0) = 0$, such that by investing amount x_g, the firm can reduce the disutility cost of group $g \in G$ from $\delta_g(\ell)$ to $\left(1 - \rho_g\left(x_g\right)\right)\delta_g(\ell)$. The focus of this paper is on the effects of the discrimination spillovers, and therefore we do not investigate the characteristics of the firm's optimal investment in discrimination reduction. However, it seems natural that, at least for some groups, the profit maximising level of investment x_g^* is such that $\rho_g\left(x_g^*\right) < 1$: full elimination of discrimination for group g is too costly, at least in cases such as those in Section 2.6, where the a firm's return on this investment can be very small indeed. To sum up, the extent of discrimination that we take as exogenously given can be interpreted as the amount that remains once the firm has

Figure 2: Example: the firm's optimal policy.



Note: The LHS depicts the output produced by (solid curves) and the wage received by (dashed curves) a worker in the discriminated group (pink curves) and the non-discriminated group (navy curves) as a function of their type θ. The thin dashed line illustrate the job levels and the pay for workers of the two groups employed at job level $\ell = 1.4$. On the RHS we plot the utility as a function of type θ for the two groups.

reduced it optimally.

2.6 Examples

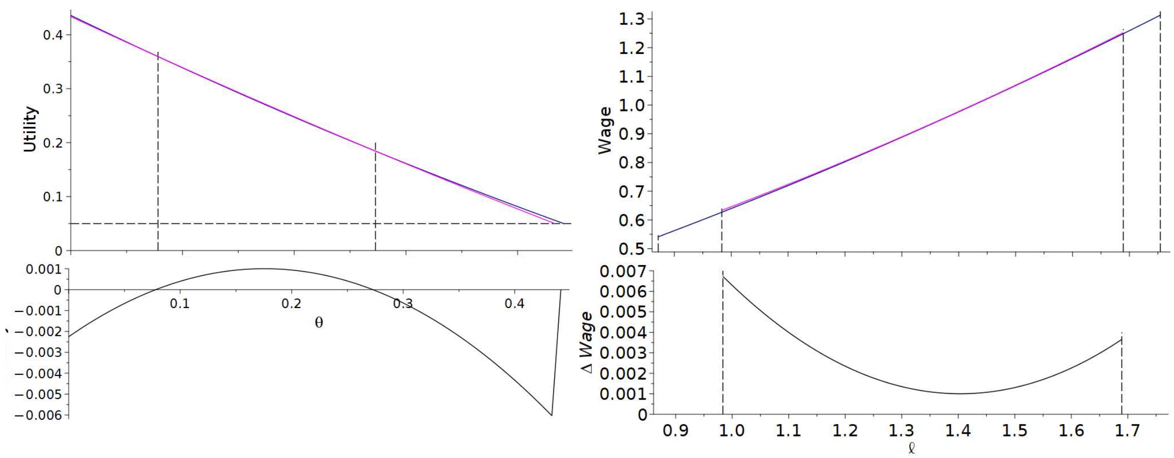
In this subsection, we use a simple set of functional forms for the general model developed above to illustrate two among the several possible types of equilibrium configurations that may emerge. We compare the firm's offer to workers in two groups: one, labelled by N, suffers next to no discrimination; some workers in the other group, labelled by D, suffer a strictly higher discrimination $\delta_D(\ell) > \delta_N(\ell)$, for at least some values of $\ell \in L$.[24] We posit quadratic formulations for the disutility functions: $\psi(\ell+\theta) = a(\ell+\theta)^2$, and $\delta_g(\ell) = \alpha_g + \beta_g(2a\ell - \gamma_g)^2$, $g = N, D$, and a uniform distribution of θ. The parameters $a > 0$ and $\gamma_g \in [0,1]$ measure the

[24]The Maple file used to compute the numerical examples is available in the online appendix.

importance of the cost of effort and determine the job level where discrimination is at its lowest. The higher γ_D the higher the job level where discrimination is lowest. The two cases we consider differ in the value of γ_D. In both examples, $a = 0.285$, $\beta_D = 0.2$, $\beta_N = 0.001$, and $u_D = u_N = 0.05$. In addition, $\alpha_D = \alpha_N = 0$: this ensures that there is a job level where no discrimination exists, putting the emphasis on the heterogeneity of discrimination.

In the first example we set $\gamma_D = 0.2$, this means that the lowest discrimination is at the lower levels of the hierarchy, and, consequently, discrimination is increasing in ℓ the relevant range. The solid curves in the LHS panel in Figure 2 show that at all ability levels, a worker in the discriminated group is allocated to a significantly lower job level. This happens because the positive slope of δ increases the information rent they need in order to report truthfully. These workers have an extra incentive to report a worse type, that is to avoid increased discrimination, so the firm adjusts by lowering the job level, and with it the associated information rent. Although workers are fully compensated for the discrimination they endure, *ceteris paribus* increasing their wage, as a result of their underemployment their cost of effort decreases, and as a result of the underemployment of their fellow group members their information rent decreases as well. The latter effects dominate: as shown by the comparison of the two dashed curves in the panel, the pay of the discriminated group is lower. In fact, even conditional on getting the same job, discriminated workers earn less, since they are overqualified: their cost of effort is lower than that of their non-discriminated peers at the same job level. As an example, the same panel illustrates the employees from the two groups who choose job level $\ell = 1.4$: these are type $\theta = 0.058$ for group D, and type $\theta = 0.177$ for group N, as shown by the horizontal thin dashed grey line. Their pay can be obtained by going down vertically to the dashed $w_g(\theta)$ schedules: the thin pink and navy lines illustrate the pay gap: the discriminated group's pay is lower than

Figure 3: Utility and pay scale.



Note: The LHS depicts the workers' utility as a function of their type θ in the discriminated group (pink curve) and in the non-discriminated group (navy curve). In the lower part the difference between the two is depicted. The dashed lines correspond to the intersection points (the vertical ones), and the reservation utility (the horizontal one). One the RHS, we show the pay scale offered by the firm to the two groups of workers. The vertical lines denote the upper and lower job level offered to workers in the two groups. The lower part of the diagram shows the wage premium offered to the discriminated group for the range of job levels available to both groups: job levels outside this range are not available to the discriminated group.

the non-discriminated group's: in the example $0.942 < 0.977$.

In addition to the glaring glass ceiling, note that, although at the lowest ability levels only the non-discriminated group is employed, the lowest job levels are exclusively filled by the discriminated group. Given the above discussion it is not surprising that utility is lower for the discriminated workers, as illustrated by the RHS panel in Figure 2.

In Figure 3, we posit a different set of parameters, to illustrate a different possibility. Here we set $\gamma_D = 0.8$, so that discrimination is worse at lower job levels (the non-discriminating job level is at $\ell \approx 1.4$): this example shows a subtler consequence of discrimination spillovers, which underlines the importance to analyse the entire job structure, in addition to comparing workers doing the same job. The diagram on the LHS of Figure 3 depicts the utility schedules of the various types θ

in the two groups, that is, it corresponds to the RHS in Figure 2. A hurried glance at it, and in particular at the difference between the two schedules in the lower part of the panel might reasonably give the impression that all is well: workers of the same type in the two groups receive roughly the same utility, some in the discriminated group are even strictly better off than their equal θ peers from group N. This impression would appear to be confirmed by a more careful analysis of this hierarchy's pay structure. The RHS of Figure 3 suggests that the pay received at any given job level is approximately the same for the two groups: if anything, those in the discriminated group are paid marginally more, as evidenced by the lower diagram on the RHS.[25] This is a consequence of their being compensated for the direct discrimination they suffer.

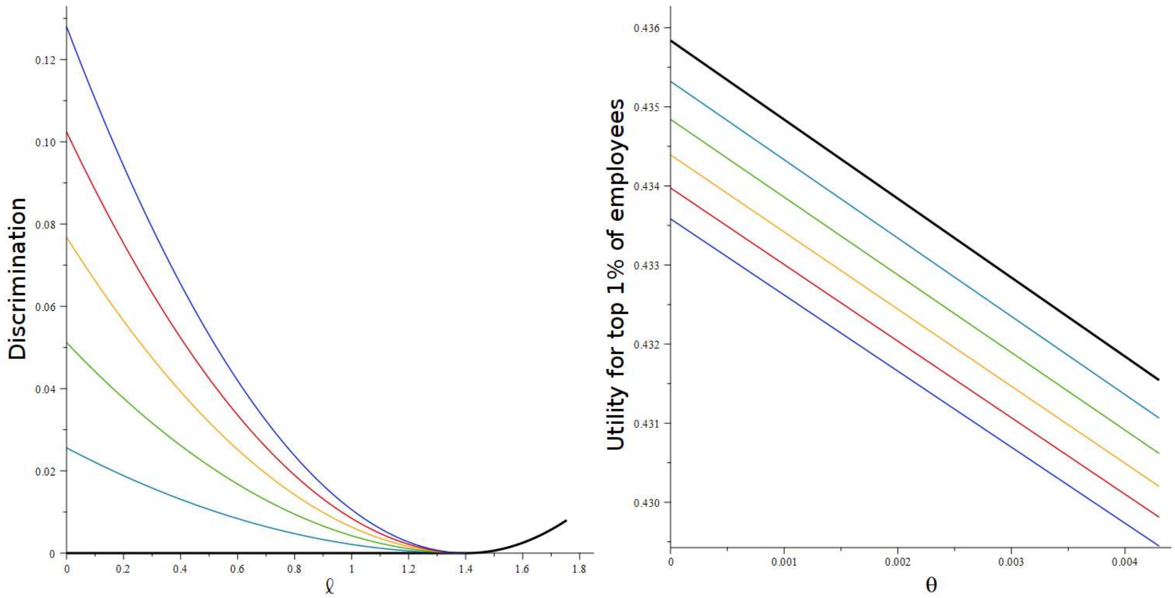
The elephant in the room is the job allocation: despite the tiny magnitude of the average utility loss of workers due to discrimination, their job allocation is heavily distorted, as the RHS panel in Figure 3 shows: the horizontal axis shows clearly that the highest job level available to members of the discriminated group is considerably lower that the highest job level that can be chosen by members of group N. We have a significant glass ceiling. In other words, "equal pay for equal jobs" goes hand in hand with the distortions in the job allocation caused by the discrimination spillovers. Workers in low level jobs face decreasing discrimination,[26] and as a result they are assigned to higher level jobs than their non-discriminated equivalents, as they are cheaper to induce into taking on more responsibility. The firm can further reduce the wage bill by creating a glass ceiling: at the top of the hierarchy, the discriminated workers are assigned to a lower level and thus exert less effort than their non-discriminated peers of the same type θ.[27]

[25]This type of observed pay schedules are precisely what has been reported in tightly controlled environments, such as the academic job market in Italy, where pay and promotions are very precisely determined (De Paola et al., 2018).

[26]This is consistent with evidence suggesting that there is more discrimination at the lowest levels of a hierarchy, for example Bircan et al. (2024).

[27]Unlike in the example in Figure 2, there is type θ_e such that type θ_e workers in the two groups

Figure 4: The spillover effect illustrated.



Note: The RHS of the figure shows the utility obtained by workers in the top 1% of the type distribution corresponding to different levels of discrimination suffered by workers in their group at lower levels of the organisational job ladder. See the text for details.

This example highlights how even clear-cut evidence of a perfect respect of the principle of "equal pay for equal work", long the topic of economists' attention (Fawcett, 1918; Edgeworth, 1922), hard fought (Cohen, 2012; Meehan and Kahn, 2016; Engstrom, 2018), only relatively recently turned into law (Brady, 1947; Elisburg, 1978) and not yet fully implemented in practice even in Scandinavian countries (Meyersson Milgrom et al., 2001) does not *per se* constitute evidence of a non-discriminatory economic environment.

We can use variations of the same example to highlight how changes in discrimination in the lower part of the hierarchy can affect workers at the top of the hierarchy. The LHS of Figure 4 depicts different, exogenously given, extent

have the same job level. This is the type who is assigned to the job level that is not discriminated against and where the slope of the $\delta_D(\ell)$ function is 0. Relative to this type θ_e, those with lower (higher) θ have lower (higher) job level if they belong to the discriminated group. In the numerical case depicted in Figure 3, type θ_e workers in the discriminated group are paid a shade more, because their information rent increases: the fact that higher types are assigned to higher job levels increases their information rent by more than the decrease due to the lower marginal type.

of the discrimination in the lower part of the organisational ladder: in each case discrimination is the same in the upper part of the pay scale. The purple line is the last example, while the thick black line is the case of no discrimination except a the top. The RHS shows the corresponding utility levels received by workers in the discriminated group with the best 1% of the values of θ, corresponding to, say, board members. Their utility, and in fact that of all the workers who are employed above the job level where discrimination is lowest, increases as discrimination at low level jobs decreases, even though their own job levels do not change.

As argued at the end of Section 2.5, the firm incentive to invest in discrimination reduction may not offer a sufficiently incentive to invest to eliminate this component of workers' disutility. This is the case in this example: even in the extreme case of complete elimination of discrimination, that is, an investment x_g^*, such that $\rho_g\left(x_g^*\right) = 0$ the increase in profit is quantitatively negligible: assuming two equal sized groups, in the example in Figure 2, it is just under 1.2%, in that in Figure 3 is even lower, at less than 0.1%.

3 Competition for labour

3.1 A Salop model of a labour market

Few, if any, firms operate in the perfect isolation posited above, and so a pertinent question arises naturally: would exposing the firm in the model in Section 2 to competition for its workers from other firms change the qualitative features of our findings? In this section we answer this question in the negative. Few would argue that the paradigm of perfect competition is an apt description for the modern labour markets, and for this reason we embed the single firm of Section 2 into an *imperfectly competitive* labour market.[28] To this end, we adapt the the well-known Salop (1979) "circular city" model of imperfect product market competition, a

[28]Recall that, on the output side, we have assumed perfect competition.

natural extension to N firms of the two-firm Hotelling (1929) price competition model. Formally, we study a market with N firms, each identical to the firm we examined in Section 2, located at evenly spaced locations, "towns", along a circle of unit length. We take the viewpoint of the firm located in $t = 0$. Letting $t \in [0, 1)$ be the clockwise distance from 0, firms are at locations $t \in \left\{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}\right\}$.

Workers, just like Salop's consumers, live in a continuum of "villages" evenly distributed along the circle. At each location t there is a measure 1, in $\mathbb{R}$, of workers of group $g \in G$. The distribution of workers is the same at every location, and as before, it is the same in every group and given by $\Phi(\theta)$. So the total measure, in $\mathbb{R}^2$, of workers is given by $|G| \times \Phi(1)$. If workers are employed, they need to travel to a town, unless of course they live there. The unit cost of travel for a worker is strictly positive and is denoted by $c > 0$.[29] We assume that the workers' location is unobservable: workers simply turn up at the gates.

3.2 Analysis of the competition case

While the model proposed above is a useful characterisation of labour market competition, and it can deliver interesting insights into the operation of these markets (as indicated by, among others, Bhaskar et al. (2002)), we limit ourselves here to use it to convince the reader that the qualitative features of the equilibrium obtained for an isolated firm remain unaltered when firms can compete for workers instead.

We model the competition between firms as a simultaneous choice static game and we therefore look for a Nash equilibrium. We focus on symmetric equilibria, implying that we can analyse only the competition between our reference firm, located in 0, and the firm located in $1/N$: competition with the firm at $\frac{N-1}{N}$ is its mirror case. Given the absence of externalities among groups, we consider a single

[29] From a technical viewpoint, the equilibrium set has a discontinuity at $c = 0$, when *all* workers are indifferent among all firms. Otherwise, for any $c > 0$, that is if travel is not costless, then workers prefer, *ceteris paribus*, employment at a nearer firm, and so for given wage schedule only a measure 0, in $\mathbb{R}^2$, of workers is indifferent between any two firms.

one, which allows us to lighten notation by dropping the subscript g when doing so generates no ambiguity. Because location is not observable and it is costless for workers to misreport theirs, truth telling can only be achieved by not conditioning contracts on location. That is, each firm offers a single pay scale to each group: if two workers of the same type and the same group from two different villages are both employed by the same firm, then they are allocated to the same job level and are paid the same. The firm's strategy in (12) in Section 2 was the choice of a pay scale $\omega(\ell)$ and the type of the marginal worker $\bar{\theta}_g$, which, given the workers' utility function, maps into the job allocation and salary functions $\{\ell(\theta), w(\theta)\}$, with $\ell : [0, \bar{\theta}_g] \rightarrow \mathbb{R}_+$ and $w : [0, \bar{\theta}_g] \rightarrow \mathbb{R}_+$ and the interval of types who select an available job level, $[0, \bar{\theta}_g]$. This remains the case in this set-up. However, the interval of employed workers may in general depend of the location where they live. To formalise this, in what follows we denote by $\bar{\theta}_c(t)$ the function determining the type of the marginal worker in village t when the unit travel cost is c. As shown by Epstein and Peters (1999), the revelation principle extends to competition, as long as the workers do not hold relevant private information, for example, a competitor's offers. We assume this to be the case here, and so the revelation principle still applies: firms still need to provide the incentive for truthful revelation of the workers' type. This means that versions of (8) and (11) still hold, implying that the relative payoffs between different types remain the same as in the isolated firm case. Thus, if a worker is employed, then their job level is still determined by (8) in Theorem 1. This in turn implies that each firm's strategy can be reduced to the choice of two variables. One is the highest worker type who accepts an offer. This worker lives at $t = 0$, and so has type $\bar{\theta}_c(0)$. With a slight abuse of notation we write it as $\bar{\theta}_c \in [0, 1]$. The other variable, which we denote by $\varepsilon_w \in \mathbb{R}$, is the change in every worker's salary relative to the one they would be paid by an isolated firm where the marginal worker employed has type $\bar{\theta}_c$. Hence the salary paid to worker type $\theta \in [0, \bar{\theta}_c]$ is $w^*(\theta, \bar{\theta}_c) + \varepsilon_w$, where $w^*(\theta, \bar{\theta}_c)$ is obtained by setting $\bar{\theta}_g = \bar{\theta}_c$ in (11). To sum up, the pay scale posted by the firm located in 0 is $\omega^*(\ell) = w^*(\Theta^*(\ell), \bar{\theta}_c) + \varepsilon_w$, for $\ell \in \left[\ell_g^*(\bar{\theta}_g), \ell_g^*(0)\right]$. Here $\Theta^*(\ell)$ is, as in (12), the inverse function of $\ell^*(\theta)$, in turn derived in (8) in Theorem 1.

The following summarises how the equilibrium changes as c varies. In the statement, we consider only the interval $\left[0, \frac{1}{2N}\right]$, symmetry ensuring that this is repeated along the circle.

Proposition 5. *There exist $c^C, c^H > 0$, with $c^C \leq c^H$, such that, at a symmetric Nash equilibrium, all firms choose a wage schedule satisfying the following.*

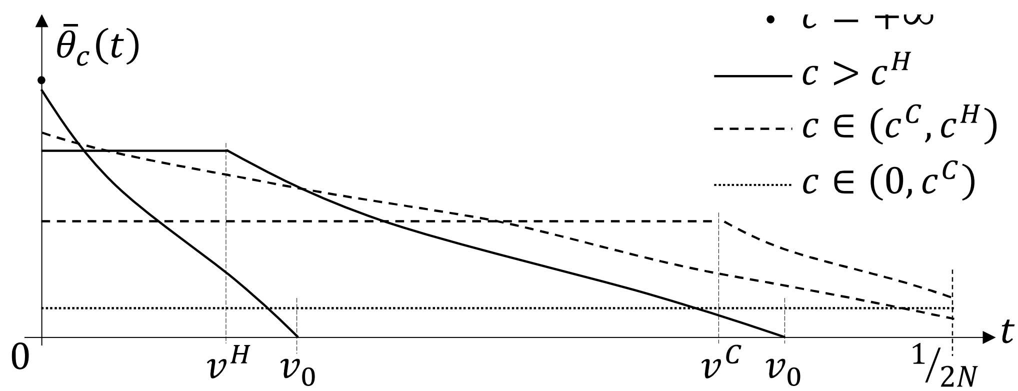
(i) ***Distance does not matter:*** *If $c \in (0, c^C]$, then the wage schedule is an upward parallel shift of $\omega^*(\ell)$, defined in (12), for $\ell \geq \ell^*(\theta_c)$, where $\bar{\theta}_c(t) < \bar{\theta}$; moreover $\bar{\theta}_c'(t) = 0$, for all $t \in \left[0, \frac{1}{2N}\right]$.*

(ii) ***Employment is higher near cities:*** *If $c \in (c^C, c^H]$, then there are some workers employed in every village; moreover there exists $v^C \in \left[0, \frac{1}{2N}\right]$ such that $\bar{\theta}_c'(t) = 0$ if $t \in [0, v^C]$, and $\bar{\theta}_c'(t) < 0$ if $t \in \left(v^C, \frac{1}{2N}\right]$.*

(iii) ***There are idle villages:*** *If $c > c^H$, then there exists $v^H, v_0 \in \left[0, \frac{1}{2N}\right]$, with $v_0 > v^H$, such that village t has workers employed by firm in 0 if and only if $t \in [0, v_0]$; moreover $\bar{\theta}_c'(t) = 0$ if $t \in [0, v^H]$, and $\bar{\theta}_c'(t) < 0$ if $t \in (v^H, v_0]$.*

Proposition 5 is illustrated in Figure 5, which depicts the function $\bar{\theta}_c(t)$ in the various possible scenarios identified in the statement. On the horizontal axis are the locations from 0 to $\frac{1}{2N}$ and the vertical axis measures the corresponding type of the marginal worker as c decreases from ∞, when nobody travels, to 0. For c high, that is above c^H, the solid lines depict cases when in some villages nobody commutes to work (Proposition 5.(iii)): two different patterns can emerge, as shown by the two solid curves. When the line is strictly decreasing, that is when $v^H = 0$, marginal workers in all villages obtain only their reservation utility, after paying for travel. If instead $v^H = 0$, so that there is a horizontal segment in a neighbourhood of 0, there

Figure 5: The marginal worker in the villages to the right of the city.



Note: The type of the highest θ worker employed at village t, $t \in [0, \frac{1}{2N}]$, as c varies. The village locations v^H, v_0, and v^C, are defined in Proposition 5.

are workers who live close enough to the city so that they would be willing to work in job levels lower than the ones available at the firm, but the firm chooses not to have jobs at these lower levels. As c decreases further, fewer and fewer villages have no one working, until, when $c = c^C$, Proposition 5.(ii) holds, and the function $\bar{\theta}_c(t)$ meets the horizontal axis at $t = \frac{1}{2n}$. For c lower than this threshold, two patterns are again possible depending whether $v^C = 0$ or $v^C > 0$. In both cases, employment in the mid-point village, $t = \frac{1}{2N}$ is strictly positive and strictly less than it is at $t = 0$. Here, the top worker in the middle village, $t = \frac{1}{2n}$, whose type is $\theta = 0$, is indifferent between working at either firm in $t = 0$, or firm in $t = \frac{1}{n}$, or remaining unemployed.

When $\bar{\theta}'_c(t) = 0$ for all $t \in \left[0, \frac{1}{2N}\right]$, that is when $c \leq c^C$, the distribution of the workers employed who live in villages $t \in \left[0, \frac{1}{2N}\right]$ is a scaled up version of those who live in $t = 0$. It follows that its hazard rate is exactly $h(\theta)$, and hence the job allocation for the workers employed is also the same as in isolated firm case. The worker's compensation differs from this case for two reasons. These work in opposite directions: first, competition ensures that workers living far are partly compensated for their commuting cost, which, pay being independent of location, also benefits workers with shorter commutes. Second, the information rent decreases, as some high θ workers are no longer employed. Both reasons cause a parallel shift in the wage function $w(\theta)$ and therefore the change is the same for all workers in a group.

Corollary 4 collects a partial characterisation of the effects of changes in competition on the outcome for workers in this low travel cost case. Recall that an increase in competition corresponds to a decrease in $\frac{c}{N}$, and so the statement shows the effects of a reduction in competition.

Corollary 4. *For $c < c^C$, at the Nash equilibrium, $\frac{\mathrm{d}\bar{\theta}_c}{\mathrm{d}\frac{c}{N}} < 0$ and $\frac{\mathrm{d}\varepsilon_w}{\mathrm{d}\frac{c}{N}} > 0$. Moreover, for every employed worker $\theta \in \left[0, \bar{\theta}_c\right]$, $\frac{\mathrm{d}w^*(\theta,\theta_c)}{\mathrm{d}\frac{c}{N}} = 0$, and, for workers in villages $t \in \left[0, \frac{1}{2N}\right]$, $\frac{\mathrm{d}U(\theta)}{\mathrm{d}\frac{c}{N}} = -t$.*

In this case of $c < c^C$, when travel cost becomes lower, competition among firms becomes more intense, and the marginal worker' type and employment both increase. This in turn increases all the inframarginal workers' information rent, but this, in the "knife-edge" linear travel cost case considered in the corollary, is exactly compensated by a change in the "wage premium" ε_w, with the consequence that overall salary does not change. Nevertheless, workers benefit from the increase in competition, unless they live in the city, at $t = 0$, as they can travel to work more cheaply.

The same conclusion applies when travel cost is sufficiently high that the marginal worker's type is not the same in every village. As shown in (8), the job allocation schedule depends on the hazard rate of the distribution of the workers employed, and therefore when, c is higher that c^C, it is in general different from $h(\theta)$. This so because in villages sufficiently distant, higher θ individuals are not employed. However, as Lemma 3 in the appendix shows, the resulting hazard rate shares the monotonicity property posited for $h(\theta)$. Thus, while the job allocation and wage function may vary with c, their qualitative nature is the same as that obtained in Section 2.

The broad conclusion one can draw from this discussion is that the characteristics of the pay scales and of the job allocation patterns are altered only minimally as the parameter measuring competition, c, or more precisely the number of firms relative to the the travel cost, $\frac{N}{c}$, moves from perfect monopsony to near perfect competition. Indeed, the firm's hierarchy "returns" to the exact shape each firm would choose in perfect isolation as competition becomes sufficiently intense, possibly having changed more substantially along the way, for intermediate levels of competition.

Any discrimination spillover, pay gaps, and glass ceiling which we identified in our analysis of an isolated firm in Section 2 are therefore replicated when firms compete for workers. And if the internal pressure to lower the disutility cost $\delta_g(\ell)$ is weak and insufficient to induce the isolated firm to take step in this direction, as for example used to draw Figure 3 in Section 2.6, it will be at least as weak in competitive conditions, suggesting that competitive pressure may well be ineffective to lower discrimination.

4 Conclusion

To paraphrase Justice Potter Stewart, it is easy to recognise discrimination when we encounter it, but it is difficult to formalise it in a rigorous manner. Perhaps this difficulty is at the root of the relative dearth of theoretical analyses of this pervasive phenomenon. The aim of this paper is to propose a formal, though admitted partial and unidimensional, definition of discrimination and provide a sound theoretical foundation for the analysis of its effects in labour markets. We build on Becker's intuition (1957) of irrational preferences, preferences, that is, that are unjustified by underlying productivity parameters. Our key contribution is to show that, even when the irrational dislikes which lead to harmful or distressful behaviour towards members of some groups is present only in a circumscribed portion of an organisation's hierarchy, they can spread their insidious tentacles to the entire hierarchical scale, all the way to the top, causing the discriminatory distortions of efficient allocation known as the pay-gap and the glass ceiling.

The paper ends by addressing the key question "posed by the economics of discrimination [...]: under what circumstances is it possible for groups with identical economic characteristics to receive different wages in a market equilibrium? If people of the same productivity receive different wages, then there are profits to be made by hiring the low-wage individual" (Stiglitz, 1973, p. 287). In the set-up of our paper, organisations' internal pressure to eliminate the spillover effects of discrimination is weak, and likewise, so is competitive pressure from other organisations competing for workers.

References

Aigner, Dennis J and Glen G Cain (1977) "Statistical theories of discrimination in labor markets," *Industrial and Labor Relations Review*, 30 (2), 175–187.

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

Bagnoli, Mark and Ted Bergstrom (2005) "Log-Concave Probability and Its Applications," *Economic Theory*, 26, 445–469.

Beath, John and Yannis Katsoulacos (1991) *The Economic Theory of Product Differentiation*: Cambridge University Press.

Becker, Gary S (1957) *The Economics of Discrimination*: University of Chicago Press.

Bertrand, Marianne (2018) "Coase Lecture: The glass ceiling," *Economica*, 85 (338), 205–231.

Bertrand, Marianne and Esther Duflo (2017) "Field experiments on discrimination," in Duflo, Esther and Abhijit Banerjee eds. *Handbook of Field Experiments*, 309–393: Elsevier.

Bhaskar, Venkataraman, Alan Manning, and Ted To (2002) "Oligopsony and monopsonistic competition in labor markets," *Journal of Economic Perspectives*, 16 (2), 155–174.

Bircan, Cagatay, Guido Friebel, and Tristan Stahl (2024) "Gender promotion gaps in knowledge work: The role of task assignment in teams," Working Paper 291, EBRD.

Brady, Dorothy S (1947) "Equal pay for women workers," *Annals of the American Academy of Political and Social Science*, 251 (1), 53–60.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani (2018) "The morale effects of pay inequality," *Quarterly Journal of Economics*, 133 (2), 611–663.

Coate, Stephen and Glenn C Loury (1993) "Will affirmative-action policies eliminate negative stereotypes?" *American Economic Review*, 1220–1240.

Cohen, Sheila (2012) "Equal pay–or what? Economics, politics and the 1968 Ford sewing machinists strike," *Labor History*, 53 (1), 51–68.

Cook, Cody, Rebecca Diamond, Jonathan V Hall, John A List, and Paul Oyer (2021) "The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers," *Review of Economic Studies*, 88 (5), 2210–2238.

Cooke, Dudley, Ana P Fernandes, and Priscila Ferreira (2019) "Product market competition and gender discrimination," *Journal of Economic Behavior & Organization*, 157, 496–522.

Cortina, Lilia M (2008) "Unseen injustice: Incivility as modern discrimination in organizations," *Academy of Management Review*, 33 (1), 55–75.

Cortina, Lilia M, Dana Kabat-Farr, Emily A Leskinen, Marisela Huerta, and Vicki J Magley (2013) "Selective incivility as modern discrimination in organizations: Evidence and impact," *Journal of Management*, 39 (6), 1579–1605.

Costrell, Robert M and Glenn C Loury (2004) "Distribution of ability and earnings in a hierarchical job assignment model," *Journal of Political Economy*, 112 (6), 1322–1363.

De Fraja, Gianni (2005) "Reverse discrimination and efficiency in education," *International Economic Review*, 46 (3), 1009–1031.

De Fraja, Gianni, Giovanni Facchini, and John Gathergood (2019) "Academic salaries and public evaluation of university research: Evidence from the UK Research Excellence Framework," *Economic Policy*, 34 (99), 523–583.

De Paola, Maria, Michela Ponzo, and Vincenzo Scoppa (2018) "Are men given priority for top jobs? Investigating the glass ceiling in Italian academia," *Journal of Human Capital*, 12 (3), 475–503.

Di Marco, Donatella, Helge Hoel, Alicia Arenas, and Lourdes Munduate (2018) "Workplace incivility as modern sexual prejudice," *Journal of Interpersonal Violence*, 33 (12), 1978–2004.

Eberhardt, Markus, Giovanni Facchini, and Valeria Rueda (2023) "Gender differences in reference letters: Evidence from the economics job market," *Economic Journal*, 133 (655), 2676–2708.

Edgeworth, Francis Ysidro (1922) "Equal pay to men and women for equal work," *Economic Journal*, 32 (128), 431–457.

Elisburg, Donald (1978) "Equal pay in the United States: The development and implementation of the Equal Pay Act of 1963," *Labor Law Journal*, 29 (4), 195.

Engstrom, David Freeman (2018) “" Not merely there to help the men": Equal pay laws, collective rights, and the making of the modern class action,” *Stanford Law Review*, 70, 1.

Epstein, Larry G and Michael Peters (1999) “A revelation principle for competing mechanisms,” *Journal of Economic Theory*, 88 (1), 119–160.

Exley, Christine L and Judd B Kessler (2022) “The gender gap in self-promotion,” *Quarterly Journal of Economics*, 137 (3), 1345–1381.

Fagan, Colette and Nina Teasdale (2021) “Women professors across STEMM and non-STEMM disciplines: navigating gendered spaces and playing the academic game,” *Work, Employment and Society*, 35 (4), 774–792.

Fawcett, Millicent G (1918) “Equal pay for equal work,” *Economic Journal*, 28 (109), 1–6.

Folke, Olle and Johanna Rickne (2022) “Sexual harassment and gender inequality in the labor market,” *Quarterly Journal of Economics*, 137 (4), 2163–2212.

Gagnon, Nickolas, Kristof Bosmans, and Arno Riedl (2025) “The effect of gender discrimination on labor supply,” *Journal of Political Economy*, forthcoming.

Glover, Dylan, Amanda Pallais, and William Pariente (2017) “Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores,” *Quarterly Journal of Economics*, 132 (3), 1219–1260.

Goldin, Claudia (2014) “A grand gender convergence: Its last chapter,” *American Economic Review*, 104 (4), 1091–1119.

——— (2024) “Nobel Lecture: An evolving economic force,” *American Economic Review*, 114 (6), 1515–1539.

Grout, Paul A, In-Uck Park, and Silvia Sonderegger (2009) “An economic theory of glass ceiling,” Working Paper 09/227, Centre for Market and Public Organisation, Bristol.

Gunderson, Morley (1989) “Male-female wage differentials and policy responses,” *Journal of Economic Literature*, 27 (1), 46–72.

Guryan, Jonathan and Kerwin Kofi Charles (2013) “Taste-based or statistical discrimination: The economics of discrimination returns to its roots,” *The Economic Journal*, 123 (572), F417–F432.

Hebl, Mikki, Christine Nittrouer, Abigail Corrington, and Juan Madera (2018) “How we describe male and female job applicants differently,” *Harvard Business Review*, 27.

Herkenhoff, Kyle, Jeremy Lise, Guido Menzio, and Gordon M Phillips (2024) “Production and learning in teams,” *Econometrica*, 92 (2), 467–504.

Hermalin, Benjamin E. (2001) “Economics and corporate culture,” in Cooper, Cary, P Christopher Earley, and Susan Cartwright eds. *International Handbook of Organizational Culture and Climate*: John Wiley & Sons, Chichester, England , Available at SSRN 162549.

Hotelling, Harold (1929) “Stability in competition,” *Economic Journal*, 39 (153), 41–57.

Hurst, Erik, Yona Rubinstein, and Kazuatsu Shimizu (2024) "Task-based discrimination," *American Economic Review*, 114 (6), 1723–1768.

Kabat-Farr, Dana, Isis H Settles, and Lilia M Cortina (2020) "Selective incivility: An insidious form of discrimination in organizations," *Equality, Diversity and Inclusion*, 39 (3), 253–260.

Kawamura, Kohei and József Sákovics (2014) "Spillovers of equal treatment in wage offers," *Scottish Journal of Political Economy*, 61 (5), 487–501.

Kern, Julie H and Alicia A Grandey (2009) "Customer incivility as a social stressor: The role of race and racial identity for service employees.," *Journal of Occupational Health Psychology*, 14 (1), 46.

Laffont, Jean-Jacques and David Martimort (2002) *The Theory of Incentives: The Principal-Agent Model*: Princeton University Press.

Laffont, Jean-Jacques and Jean Tirole (1993) *A Theory of Incentives in Procurement and Regulation*, Cambridge, Massachusetts: MIT Press.

Leonard, Daniel and Ngo van Long (1992) *Optimal Control Theory and Static Optimization in Economics*, Cambridge, UK: Cambridge University Press.

Maskin, Eric and John Riley (1984) "Monopoly with incomplete information," *RAND Journal of Economics*, 15 (2), 171–196.

Meehan, Elizabeth and Peggy Kahn (2016) *Equal Value/Comparable Worth in the UK and the USA*: Springer.

Meyersson Milgrom, Eva M, Trond Petersen, and Vemund Snartland (2001) "Equal pay for equal work? Evidence from Sweden and a comparison with Norway and the US," *Scandinavian Journal of Economics*, 103 (4), 559–583.

Mussa, Michael and Sherwin Rosen (1978) "Monopoly and product quality," *Journal of Economic Theory*, 18 (2), 301–317.

Myerson, Roger B (1979) "Incentive compatibility and the bargaining problem," *Econometrica*, 61–73.

Onuchic, Paula (2022) "Recent contributions to theories of discrimination," *arXiv preprint*, 2205.05994.

Pascual-Fuster, Bartolomé, Ryan Federo, Rafel Crespí-Cladera, and Patricia Gabaldón (2024) "The second glass ceiling: The dark side of women recategorization in corporate boards," *Gender, Work & Organization*.

Petrongolo, Barbara and Maddalena Ronchi (2020) "Gender gaps and the structure of local labor markets," *Labour Economics*, 64, 101819.

Phelps, Edmund S (1972) "The statistical theory of racism and sexism," *American Economic Review*, 62 (4), 659–661.

Posner, Richard A (1989) "An economic analysis of sex discrimination laws," *University of Chicago Law Review*, 56 (4), 1311–1335.

Salop, Steven C (1979) "Monopolistic competition with outside goods," *Bell Journal of Economics*, 10 (1), 141–156, `http://www.jstor.org/stable/3003323`.

Schaerer, Michael, Christilene Du Plessis, My Hoang Bao Nguyen et al. (2023) "On the trajectory of discrimination: A meta-analysis and forecasting survey capturing 44 years of field experiments on gender and hiring decisions," *Organizational Behavior and Human Decision Processes*, 179, 104280.

Siddique, Abu, Michael Vlassopoulos, and Yves Zenou (2023) "Market competition and discrimination," *European Economic Review*, 152, 104361.

Small, Mario L and Devah Pager (2020) "Sociological perspectives on racial discrimination," *Journal of Economic Perspectives*, 34 (2), 49–67.

Stiglitz, Joseph E (1973) "Approaches to the economics of discrimination," *American Economic Review*, 287–295.

Winter-Ebmer, Rudolf (1995) "Sex discrimination and competition in product and labour markets," *Applied Economics*, 27 (9), 849–857.

Wiswall, Matthew and Basit Zafar (2018) "Preference for the workplace, investment in human capital, and gender," *Quarterly Journal of Economics*, 133 (1), 457–507.

Wood, Robert G, Mary E Corcoran, and Paul N Courant (1993) "Pay differences among the highly paid: The male-female earnings gap in lawyers' salaries," *Journal of Labor Economics*, 11 (3), 417–441.

Appendix: Proof of the results

Proof of Lemma 1. Begin by noting that, from (3):

$$\dot{U}_g(\theta) = \dot{w}_g(\theta) - \psi'\left(\ell_g(\theta) + \theta\right)\left(\dot{\ell}_g(\theta) + 1\right) - \delta_g'\left(\ell_g(\theta)\right)\dot{\ell}_g(\theta). \tag{A1}$$

Next note that if type θ tries to pretend to be of type $\tilde{\theta}$, they will apply to jobs at job level $\ell_g(\tilde{\theta})$, and we can thus write the utility of misreporting $\tilde{\theta}$ instead of θ as

$$\varphi(\tilde{\theta}, \theta) = w_g(\tilde{\theta}) - \psi\left(\ell_g(\tilde{\theta}) + \theta\right) - \delta_g\left(\ell_g(\tilde{\theta})\right). \tag{A2}$$

If this is maximised by choice of $\tilde{\theta}$, we have

$$\frac{\partial \varphi(\tilde{\theta}, \theta)}{\partial \tilde{\theta}} = \dot{w}_g(\tilde{\theta}) - \psi'\left(\ell_g(\tilde{\theta}) + \theta\right)\dot{\ell}_g(\tilde{\theta}) - \delta_g'\left(\ell_g(\tilde{\theta})\right)\dot{\ell}_g(\tilde{\theta}) = 0. \tag{A3}$$

Incentive compatibility requires (A2) to be maximised at $\tilde{\theta} = \theta$, and so

$$\dot{U}_g(\theta) + \psi'\left(\ell_g(\theta) + \theta\right) = 0. \tag{A4}$$

The above is the first order condition for $\tilde{\theta} = \theta$ to be a local maximum. Now simply substitute (A1) into (A4) to obtain (5). Incentive compatibility also requires that type θ_1 prefers job $\ell_g(\theta_1)$ to job $\ell_g(\theta_2)$, and conversely that type θ_2 prefers job $\ell_g(\theta_2)$ to job $\ell_g(\theta_1)$:

$$Z(\theta_1) - \psi\left(\ell_g(\theta_1) + \theta_1\right) \geq Z(\theta_2) - \psi\left(\ell_g(\theta_2) + \theta_1\right),$$
$$Z(\theta_2) - \psi\left(\ell_g(\theta_2) + \theta_2\right) \geq Z(\theta_1) - \psi\left(\ell_g(\theta_1) + \theta_2\right),$$

where $Z(\theta) = w_g(\theta) - \delta_g\left(\ell_g(\theta)\right)$. Adding up

$$\psi\left(\ell_g(\theta_1) + \theta_2\right) - \psi\left(\ell_g(\theta_1) + \theta_1\right) \geq \psi\left(\ell_g(\theta_2) + \theta_2\right) - \psi\left(\ell_g(\theta_2) + \theta_1\right),$$

or

$$\int_{\theta_1}^{\theta_2} \psi' \left(\ell_g(\theta_1) + x\right) \mathrm{d}x - \int_{\theta_1}^{\theta_2} \psi' \left(\ell_g(\theta_2) + x\right) \mathrm{d}x \geq 0, \tag{A5}$$

which can be written as

$$\int_{\ell_g(\theta_2)}^{\ell_g(\theta_1)} \int_{\theta_1}^{\theta_2} \psi'' (y + x) \,\mathrm{d}x\mathrm{d}y \geq 0.$$

This must hold for any $\theta_1 \neq \theta_2$ in $[0, 1]$, and so, since the the integrand is positive, it must be that $\dot{\ell}_g(\theta) \leq 0$.

We complete the proof by noting that if $\dot{\ell}_g(\theta) \leq 0$, then the second order conditions are sufficient (Laffont and Tirole, 1993, p.121). Assume for contradiction that type θ strictly prefers to apply to job level $\ell_g\left(\tilde{\theta}\right)$, that is that:

$$\varphi\left(\tilde{\theta}, \theta\right) > \varphi(\theta, \theta),$$

which is equivalent to

$$\int_{\theta}^{\tilde{\theta}} \varphi_1(x, \theta) \,\mathrm{d}x > 0.$$

This, using the first order condition for the optimality of the choice of $\tilde{\theta}$ in (A3), can be written as $\int_{\theta}^{\tilde{\theta}} (\varphi_1(x, \theta) - \varphi_1(x, x))\, dx > 0$, that is as

$$\int_{\theta}^{\tilde{\theta}} \int_{x}^{\theta} \varphi_{12}(y, x) \,\mathrm{d}y\mathrm{d}x > 0. \tag{A6}$$

Again from (A3),

$$\begin{aligned}
\frac{\partial \varphi_1(y, x)}{\partial x} &= \frac{\partial \left(\dot{w}_g(y) - \psi'\left(\ell_g(y) + x\right)\dot{\ell}_g(y) - \delta_g'\left(\ell_g(y)\right)\dot{\ell}_g(y)\right)}{\partial x} \\
&= -\psi''\left(\ell_g(y) + x\right)\dot{\ell}_g(y) \geq 0.
\end{aligned}$$

It is either $\tilde{\theta} > \theta$, in which case $x \geq \theta$ for all $x \in \left[\theta, \tilde{\theta}\right]$ and (A6) is violated. Or $\tilde{\theta} < \theta$, $x \leq \theta$ for all $x \in \left[\tilde{\theta}, \theta\right]$, contradicting again (A6). $\square$

Proof of Theorem 1: The assumption of constant returns to scale implies that the optimisation problem can be solved independently for each group. Thus, to shorten notation in this proof, we omit the subscript g from the functions $w_g(\theta)$, $\ell_g(\theta)$, $U_g(\theta)$, and $\delta_g(\theta)$. Towards finding a solution, we set the problem as an optimal control one, with $U(\theta)$ as the state variable and $\ell(\theta)$ as the control variable. Once the optimal values of these two variables are obtained, salary is derived from (3) as

$$w(\theta) = U(\theta) + \psi(\ell(\theta) + \theta) + \delta(\ell(\theta)). \tag{A7}$$

Substituting (A7) in (7), we can write the firm's profit as

$$\int_0^{\bar{\theta}_g} \Big(\ell(\theta) - \psi(\ell(\theta) + \theta) - U(\theta) - \delta(\ell(\theta)) \Big) \phi(\theta)\, \mathrm{d}\theta. \tag{A8}$$

This is the firm's value function; note that this is a free terminal time problem, as $\bar{\theta}_g$ is chosen by the firm. Towards finding a solution, we begin by using (A7) again to write the incentive compatibility constraint (5) as the law of motion for this problem,

$$\dot{U}_g(\theta) = -\psi'\left(\ell_g(\theta) + \theta\right), \tag{A9}$$

and then constructing its Hamiltonian:

$$\mathcal{H}(U(\theta), \ell(\theta)) = \Big[\ell(\theta) - \psi(\ell(\theta) + \theta) - U(\theta) - \delta(\ell(\theta)) \Big] \phi(\theta) - \mu(\theta)\, \psi'(\ell(\theta) + \theta).$$

In the above, $\mu(\theta)$ is the Pontryagin multiplier for constraint (A9). The first-order conditions for this problem are (Leonard and van Long, 1992, p. 222):

$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial \ell(\theta)} &= 0 = \Big(1 - \psi'(\ell(\theta) + \theta) - \delta'(\ell(\theta))\Big) \phi(\theta) - \mu(\theta)\, \psi''(\ell(\theta) + \theta), \\ \dot{\mu}(\theta) &= -\frac{\partial \mathcal{H}}{\partial U(\theta)} = \phi(\theta), \qquad \mu(0) = 0, \;\; \mu(\theta) \text{ free}. \end{aligned}$$

Solving the differential equation for the multiplier, we obtain $\mu(\theta) = \Phi(\theta)$, and so the first-order condition for $\ell(\theta)$ becomes

$$1 - \psi'(\ell(\theta) + \theta) = \frac{\Phi(\theta)}{\phi(\theta)} \psi''(\ell(\theta) + \theta) + \delta'(\ell(\theta)),$$

which is (8).

Next, we show that ignoring the constraint (4), which we have done so far, is in fact legitimate, as it turns out to be satisfied at the solution we have found. Indeed, we can show a stronger result, the strict monotonicity of the relationship between ℓ and θ, as stated in Corollary 1 in the main text.

Proof of Corollary 1. To obtain this result, totally differentiate (8) to write

$$\begin{aligned}&\left(-\psi''\left(\ell\left(\theta\right)+\theta\right)-h\left(\theta\right)\psi'''\left(\ell\left(\theta\right)+\theta\right)-\delta''\left(\ell\left(\theta\right)\right)\right)d\ell+\\&\left(-\psi''\left(\ell\left(\theta\right)+\theta\right)-\frac{dh\left(\theta\right)}{d\theta}\psi''\left(\ell\left(\theta\right)+\theta\right)-h\left(\theta\right)\psi'''\left(\ell\left(\theta\right)+\theta\right)\right)d\theta=0.\end{aligned}$$

and so, at the optimum,

$$\frac{\mathrm{d}\ell}{\mathrm{d}\theta}=\frac{H\left(\theta\right)}{-\eta\left(\theta\right)}<0,$$

where:

$$H\left(\theta\right)=\left(1+h'\left(\theta\right)\right)\psi''\left(\ell^*\left(\theta\right)+\theta\right)+h\left(\theta\right)\psi'''\left(\ell^*\left(\theta\right)+\theta\right)>0, \tag{A10}$$

$$\eta\left(\theta\right)=\psi''\left(\ell^*\left(\theta\right)+\theta\right)+h\left(\theta\right)\psi'''\left(\ell^*\left(\theta\right)+\theta\right)+\delta''\left(\ell^*\left(\theta\right)\right)>0. \tag{A11}$$

In (A11), the inequality follows from Assumption 1.(ii). □

To continue with the proof, we note that when the firm is free to choose the "terminal time", in our case the marginal worker's type $\bar{\theta}_g$, it will choose it in such a way that the Hamiltonian is 0 at that value (see Leonard and van Long, 1992, p 255): this balances the loss of revenue due to not employing workers of type $\bar{\theta}_g+\varepsilon$, with the additional salary that must be paid to workers in $\left[0,\bar{\theta}_g\right]$, to incentivise them to self select. That is, at the optimum, when ℓ is given by $\ell^*\left(\theta\right)$, $\bar{\theta}_g$ is the value of θ such that:

$$0=\ell^*\left(\theta\right)-u-\psi\left(\ell^*\left(\theta\right)+\theta\right)-\delta\left(\ell^*\left(\theta\right)\right)-\frac{\Phi\left(\theta\right)}{\phi\left(\theta\right)}\psi'\left(\ell^*\left(\theta\right)+\theta\right), \tag{A12}$$

where we have used the fact that the utility of the marginal worker is optimally equal to her outside option, u. From (A12), (9) can be directly obtained. Finally, to show that $\bar{\theta}_g$ is unique, we differentiate the RHS of (A12) with respect to $\bar{\theta}_g$:

$$\frac{\mathrm{d}\ell_g^*(\bar{\theta}_g)}{\mathrm{d}\bar{\theta}_g}\Big(1-\psi'(\ell_g^*(\bar{\theta}_g)+\bar{\theta}_g)-h(\bar{\theta}_g)\psi''(\ell_g^*(\bar{\theta}_g)+\bar{\theta}_g)-\delta_g'(\ell_g^*(\bar{\theta}_g)\Big)$$
$$-\psi'(\ell_g^*(\bar{\theta}_g)+\bar{\theta}_g)-h(\bar{\theta}_g)\psi''(\ell_g^*(\bar{\theta}_g)+\bar{\theta}_g)-h'(\bar{\theta}_g)\psi'(\ell_g^*(\bar{\theta}_g)+\bar{\theta}_g).$$

This is negative, because the term in brackets in the first line is 0 by (8), and all the terms in the second line are negative. □

Proof of Proposition 1: The constancy of the equilibrium salary scale in α_g and u_g follows from the fact that (8) does not contain either parameter.

Totally differentiate (8) after substitution of (2):

$$\begin{aligned}&\left(\psi''\left(\ell_g^*(\theta)+\theta\right)+h(\theta)\,\psi'''\left(\ell_g^*(\theta)+\theta\right)+\beta_g f_g''\left(\ell_g^*(\theta)\right)\right)\mathrm{d}\ell_g^*(\theta)+\\&\left(\psi''\left(\ell_g^*(\theta)+\theta\right)+h'(\theta)\,\psi''\left(\ell_g^*(\theta)+\theta\right)+h(\theta)\,\psi'''\left(\ell_g^*(\theta)+\theta\right)\right)\mathrm{d}\theta+\\&f_g'\left(\ell_g^*(\theta)\right)\mathrm{d}\beta_g=0.\end{aligned} \tag{A13}$$

This can be written, using (A10) and (A11) as:

$$\eta(\theta)\,\mathrm{d}\ell_g^*(\theta)+H(\theta)\,\mathrm{d}\theta+f_g'\left(\ell_g^*(\theta)\right)\mathrm{d}\beta_g=0$$

and so we have

$$\frac{\mathrm{d}\ell_g^*(\theta)}{\mathrm{d}\beta_g}=\frac{-f_g'\left(\ell_g^*(\theta)\right)}{\eta(\theta)}. \tag{A14}$$

Since $\eta(\theta)>0$ the second result follows. □

Proof of Proposition 2. Substituting (2) into (9) and totally differentiating:

$$\begin{aligned}\mathrm{d}u_g=&\left(1-\psi'\left(\ell_g^*+\bar{\theta}_g\right)-h\left(\bar{\theta}_g\right)\psi''\left(\ell_g^*+\bar{\theta}_g\right)+\beta_g f'\left(\ell^m-\ell_g^*\right)\right)\left(\mathrm{d}\ell_g^*+\frac{\mathrm{d}\ell_g^*}{\mathrm{d}\beta_g}\mathrm{d}\beta_g\right)\\&\left(-\psi'\left(\ell_g^*+\bar{\theta}_g\right)-h\left(\bar{\theta}_g\right)\psi''\left(\ell_g^*+\bar{\theta}_g\right)-h'\left(\bar{\theta}_g\right)\psi'\left(\ell_g^*+\bar{\theta}_g\right)\right)\mathrm{d}\bar{\theta}_g\\&-d\alpha_g-f\left(\ell_g^*\left(\bar{\theta}_g\right)\right)\mathrm{d}\beta_g.\end{aligned}$$

The first line is 0 by (8), and so

$$\frac{\mathrm{d}\bar{\theta}_g}{\mathrm{d}u_g} = \frac{\mathrm{d}\bar{\theta}_g}{\mathrm{d}\alpha_g} = \frac{-1}{h\left(\bar{\theta}_g\right)\psi''\left(\ell_g^* + \bar{\theta}_g\right) + \left(1 + h'\left(\bar{\theta}_g\right)\right)\psi'\left(\ell_g^* + \bar{\theta}_g\right)} < 0, \tag{A15}$$

$$\frac{\mathrm{d}\bar{\theta}_g}{\mathrm{d}\beta_g} = \frac{-f\left(\ell_g^*\left(\bar{\theta}_g\right)\right)}{h\left(\bar{\theta}_g\right)\psi''\left(\ell_g^* + \bar{\theta}_g\right) + \left(1 + h'\left(\bar{\theta}_g\right)\right)\psi'\left(\ell_g^* + \bar{\theta}_g\right)} < 0, \tag{A16}$$

□

Proof of Proposition 3: By Proposition 2 and $\psi'(x) > 0$, the information rent component of utility decreases in both u_g and α_g.

To show that $\frac{\mathrm{d}w_g(\theta)}{\mathrm{d}u_g} = \frac{\mathrm{d}U_g(\theta)}{\mathrm{d}u_g} \in (0,1)$, differentiate (11) with respect to u_g to obtain

$$\frac{\mathrm{d}w_g(\theta)}{\mathrm{d}u_g} = 1 + \frac{\mathrm{d}\bar{\theta}_g}{\mathrm{d}u_g}\psi'\left(\ell_g^*\left(\bar{\theta}_g\right) + \bar{\theta}_g\right). \tag{A17}$$

$\frac{\mathrm{d}\bar{\theta}_g}{\mathrm{d}u_g}$ is given in (A15), and so (A17) becomes

$$\begin{aligned}\frac{\mathrm{d}w_g(\theta)}{\mathrm{d}u_g} &= 1 - \frac{\psi'\left(\ell_g^*\left(\bar{\theta}_g\right) + \bar{\theta}_g\right)}{\left(1 + h'\left(\bar{\theta}_g\right)\right)\psi'\left(\ell_g^*\left(\bar{\theta}_g\right) + \bar{\theta}_g\right) + h\left(\bar{\theta}_g\right)\psi''\left(\ell_g^*\left(\bar{\theta}_g\right) + \bar{\theta}_g\right)} \\ &= \frac{h'\left(\bar{\theta}_g\right)\psi'\left(\ell_g^*\left(\bar{\theta}_g\right) + \bar{\theta}_g\right) + h\left(\bar{\theta}_g\right)\psi''\left(\ell_g^*\left(\bar{\theta}_g\right) + \bar{\theta}_g\right)}{\left(1 + h'\left(\bar{\theta}_g\right)\right)\psi'\left(\ell_g^*\left(\bar{\theta}_g\right) + \bar{\theta}_g\right) + h\left(\bar{\theta}_g\right)\psi''\left(\ell_g^*\left(\bar{\theta}_g\right) + \bar{\theta}_g\right)},\end{aligned}$$

which is positive and less than 1. Next, note that employed workers are fully compensated for an increase in α_g, and so $\frac{\mathrm{d}w_g(\theta)}{\mathrm{d}u_g} = \frac{\mathrm{d}\alpha_g}{\mathrm{d}u_g}$, and $\frac{\mathrm{d}U_g(\theta)}{\mathrm{d}\alpha_g} = \frac{\mathrm{d}\bar{\theta}_g}{\mathrm{d}u_g}\psi'\left(\ell_g^*\left(\bar{\theta}_g\right) + \bar{\theta}_g\right) < 0$. □

Proof of Proposition 4. Consider a worker of group g, if they stick to the pay scale designed for group g they will have utility

$$U(\theta) = w_g^*(\theta) - \psi\left(\ell_g^*(\theta) + \theta\right) - \delta_g\left(\ell_g^*(\theta)\right), \tag{A18}$$

which is

$$U(\theta) = u_g + \int_\theta^{\bar{\theta}_g} \psi'\left(\ell_g^*(x) + x\right)\mathrm{d}x. \tag{A19}$$

Suppose instead they choose the job level assigned to worker of type θ_m, the best type to pretend to be in group 0, that is

$$\theta_m = \arg\max_{\theta} \left\{ w_0^* (\theta) - \psi \left(\ell_0^* (\theta) + \theta \right) - \delta_g \left(\ell_0^* (\theta) \right) \right\}. \tag{A20}$$

In (A20), $w_0^* (\theta)$ is given by

$$w_g^* (\theta) = u_0 + \psi \left(\ell_0^* (\theta) + \theta \right) + \delta_0 \left(\ell_0^* (\theta) \right) + \int_{\theta}^{\theta_0} \psi' \left(\ell_0^* (x) + x \right) \mathrm{d}x,$$

and so the maximum utility that a worker of group g can obtain by picking a contract available to workers in group 0 is:

$$u_0 + \psi \left(\ell_0^* (\theta_m) + \theta_m \right) + \delta_0 \left(\ell_0^* (\theta_m) \right) + \int_{\theta}^{\theta_0} \psi' \left(\ell_0^* (x) + x \right) \mathrm{d}x - \psi \left(\ell_0^* (\theta_m) + \theta \right) - \delta_g \left(\ell_0^* (\theta_m) \right). \tag{A21}$$

This implies that the gain in utility which this worker can obtain by picking a contract available to workers in group 0 is the difference between (A21) and (A19):

$$\begin{aligned} \Delta U = & \int_{\theta}^{\theta_0} \psi' \left(\ell_0^* (x) + x \right) \mathrm{d}x - \int_{\theta}^{\bar{\theta}_g} \psi' \left(\ell_g^* (x) + x \right) \mathrm{d}x + \\ & \psi \left(\ell_0^* (\theta_m) + \theta_m \right) - \psi \left(\ell_0^* (\theta_m) + \theta \right) - \delta_g \left(\ell_0^* (\theta_m) \right) + \delta_0 \left(\ell_0^* (\theta_m) \right), \end{aligned}$$

using the fact that $u_0 = u_g$. The above is negative if, for some $\theta_1 \in [\min \{\theta_m, \theta\}, \max \{\theta_m, \theta\}]$ and $z(x) \in \left[\min \left\{ \ell_0^* (x), \ell_g^* (x) \right\} + x, \max \left\{ \ell_0^* (x), \ell_g^* (x) \right\} + x\right]$:

$$\begin{aligned} & \delta_g \left(\ell_0^* (\theta_m) \right) - \delta_0 \left(\ell_0^* (\theta_m) \right) > \\ & \psi' \left(\ell_0^* (\theta_1) + \theta_1 \right) (\theta_m - \theta) + \int_{\theta}^{\bar{\theta}_g} \psi'' (z(x)) \left(\ell_g^* (x) - \ell_0^* (\theta_m) \right) \mathrm{d}x + \int_{\bar{\theta}_g}^{\theta_0} \psi' \left(\ell_0^* (x) + x \right) .\mathrm{d}x \end{aligned}$$

On the RHS, the second term is positive, the third is also positive, as $\theta_0 > \bar{\theta}_g$ provided there is discrimination at the lower end of the hierarchy. Thus, if there is "sufficient" discrimination at level $\ell_0^* (\theta_m)$, then the assert holds. □

Proof of Proposition 5. To determine a symmetric Nash equilibrium, define $\left(\varepsilon_w^0, \bar{\theta}_c^0\right)$:

$\mathbb{R} \times [0,1] \to \mathbb{R} \times [0,1]$ as the best response function for the firm located at $t = 0$ on the assumption that all the other firms choose the strategy pair $\{\varepsilon_w, \bar{\theta}_c\}$. A fixed point of this best response function, where $\{\varepsilon_w^0 (\varepsilon_w, \bar{\theta}_c), \bar{\theta}_c^0 (\varepsilon_w, \bar{\theta}_c)\} = \{\varepsilon_w, \bar{\theta}_c\}$, is the sought symmetric Nash equilibrium, since every firm prefers to follow this strategy.

At a symmetric equilibrium where there is actual competition for workers, the individual rationality constraint (6) needs to be satisfied with the condition that each employed worker living in $t = \frac{1}{2N}$ be indifferent between working at the firm in $t = 0$ and at the firm in $t = \frac{1}{N}$, where the marginal worker's utility is $u + \varepsilon_w$. If c is sufficiently small, $\varepsilon_w > 0$ and consequently *all* the workers employed by firm 0 are strictly better off than unemployed. The highest such c is c^C. In this case, all workers receive utility strictly above their reservation utility, and so the marginal type is not determined by the participation constraint, and therefore it is independent of location, and all the villages in $[-\frac{1}{2N}, \frac{1}{2N}]$ have the same marginal worker, $\bar{\theta}_c^0$. In this case, the distribution of the firm's potential employees is proportional to the distribution in each village, and is given by $\frac{1}{N}\Phi(\theta)$. It follows that its hazard rate is $\frac{\frac{1}{N}\Phi(\theta)}{\frac{1}{N}\phi(\theta)} = h(\theta)$. Therefore, the analysis in Section 2 holds unaltered for this case.

Next, we characterise this equilibrium: using symmetry, we restrict attention to workers in $\left[0, \frac{1}{N}\right]$. Define firm 0's profit from employing the workers in a generic village where all the workers who are employed travel to firm 0, as a function of its strategy:

$$\pi\left(\varepsilon_w^0, \bar{\theta}_c^0\right) = \int_0^{\bar{\theta}_c^0} \left(\ell^*(x) - w^*\left(x, \bar{\theta}_c^0\right) - \varepsilon_w^0\right) \phi(x)\, \mathrm{d}x. \tag{A22}$$

In the above, the functions $\ell^*(\theta)$ and $w^*\left(\theta, \bar{\theta}_c^0\right)$ are determined by (8) and (11) with $\bar{\theta}_g = \bar{\theta}_c$. Note that, conditional on the workers choosing firm 0, π is independent of the other firms' strategy. Firm 0's total profit is

$$\Pi\left(\varepsilon_w^0, \bar{\theta}_c^0, \varepsilon_w, \theta_c\right) = 2\int_0^{t_0\left(\varepsilon_w^0, \bar{\theta}_c^0, \varepsilon_w, \bar{\theta}_c\right)} \pi\left(\varepsilon_w^0, \bar{\theta}_c^0\right) \mathrm{d}t = 2t_0\left(\varepsilon_w^0, \bar{\theta}_c^0, \varepsilon_w, \bar{\theta}_c\right) \pi\left(\varepsilon_w^0, \bar{\theta}_c^0\right), \tag{A23}$$

where $t_0\left(\varepsilon_w^0, \bar{\theta}_c^0, \varepsilon_w, \bar{\theta}_c\right) \in \left[0, \frac{1}{N}\right]$ is the village where the employed workers are indifferent between working for firm 0 or for the firm at $t = \frac{1}{N}$. We determine this village in the following lemma, which also has independent interest.

Lemma 2. *Let firms in $t = 0$ and in $t = \frac{1}{N}$ offer job-level allocation $\ell^*(x)$ and pay schedules $w^*\left(x, \bar{\theta}_c^0\right) + \varepsilon_w^0$ and $w^*\left(x, \bar{\theta}_c\right) + \varepsilon_w$. Let*

$$t_0\left(\varepsilon_w^0, \bar{\theta}_c^0, \varepsilon_w, \bar{\theta}_c\right) = \frac{1}{2N} + \frac{\varepsilon_w^0 - \varepsilon_w}{2c} + \frac{\int_{\bar{\theta}_c}^{\bar{\theta}_c^0} \psi'\left(\ell^*(x) + x\right) \mathrm{d}x}{2c}. \tag{A24}$$

Then, workers living in village $t < t_0\left(\varepsilon_w^0, \bar{\theta}_c^0, \varepsilon_w, \bar{\theta}_c\right)$ prefer to work in firm in $t = 0$, those living in village $t > t_0\left(\varepsilon_w^0, \bar{\theta}_c^0, \varepsilon_w, \bar{\theta}_c\right)$ prefer to work in firm in $t = \frac{1}{2N}$, and those living in village $t_0\left(\varepsilon_w^0, \bar{\theta}_c^0, \varepsilon_w, \bar{\theta}_c\right)$ are indifferent between the two firms.

Proof of Lemma 2. The payoff of a type θ worker living in village t and working in firm 0 is

$$u + \int_{\theta}^{\bar{\theta}_c^0} \psi'\left(\ell^*(x) + x\right) \mathrm{d}x + \varepsilon_w^0 - tc. \tag{A25}$$

This is the wage given in (9), reduced by the terms $\psi\left(\ell^*(\theta) + \theta\right)$ and $\delta\left(\ell^*(\theta)\right)$, increased by the "premium" ε_w^0 paid by the firm in $t = 0$, and reduced by the travel cost to firm in $t = 0$ from the worker's location t. Similarly if the same worker travelled to the firm in $t = \frac{1}{2N}$, their utility would be:

$$u + \int_{\theta}^{\bar{\theta}_c} \psi'\left(\ell^*(x) + x\right) \mathrm{d}x + \varepsilon_w - \left(\frac{1}{N} - t\right) c. \tag{A26}$$

Equating (A25) and (A26), we obtain (A24), and this concludes the proof of the lemma. □

The first order conditions for (A23) are:

$$\frac{\partial \Pi\left(\varepsilon_w^0, \bar{\theta}_c^0, \varepsilon_w, \bar{\theta}_c\right)}{\partial \bar{\theta}_c^0} = 2t_0\left(\varepsilon_w^0, \bar{\theta}_c^0, \varepsilon_w, \bar{\theta}_c\right) \frac{\partial \pi\left(\varepsilon_w^0, \bar{\theta}_c^0\right)}{\partial \bar{\theta}_c^0} + 2\pi\left(\varepsilon_w^0, \bar{\theta}_c^0\right) \frac{\partial t_0\left(\varepsilon_w^0, \bar{\theta}_c^0, \varepsilon_w, \bar{\theta}_c\right)}{\partial \bar{\theta}_c^0} = 0, \tag{A27}$$

$$\frac{\partial \Pi\left(\varepsilon_w^0, \bar{\theta}_c^0, \varepsilon_w, \bar{\theta}_c\right)}{\partial \varepsilon_w^0} = 2t_0\left(\varepsilon_w^0, \bar{\theta}_c^0, \varepsilon_w, \bar{\theta}_c\right) \frac{\partial \pi\left(\varepsilon_w^0, \bar{\theta}_c^0\right)}{\partial \varepsilon_w^0} + 2\pi\left(\varepsilon_w^0, \bar{\theta}_c^0\right) \frac{\partial t_0\left(\varepsilon_w^0, \bar{\theta}_c^0, \varepsilon_w, \bar{\theta}_c\right)}{\partial \varepsilon_w^0} = 0. \tag{A28}$$

where, from (A24):

$$\frac{\partial t_0\left(\varepsilon_w^0,\bar{\theta}_c^0,\varepsilon_w,\bar{\theta}_c\right)}{\partial\bar{\theta}_c^0}=\frac{\psi'\left(\ell^*\left(\bar{\theta}_c^0\right)+\bar{\theta}_c^0\right)}{2c}>0, \tag{A29}$$

$$\frac{\partial t_0\left(\varepsilon_w^0,\bar{\theta}_c^0,\varepsilon_w,\bar{\theta}_c\right)}{\partial\varepsilon_w^0}=\frac{1}{2c}>0. \tag{A30}$$

In a symmetric equilibrium $\varepsilon_w^0=\varepsilon_w$ and $\bar{\theta}_c^0=\bar{\theta}_c$, and so (A24) becomes

$$t_0\left(\varepsilon_w,\bar{\theta}_c\right)=\frac{1}{2N}. \tag{A31}$$

To compact notation, let

$$r\left(\bar{\theta}_c\right)=\int_0^{\bar{\theta}_c}\left(\ell^*\left(x\right)-w^*\left(x,\bar{\theta}_c\right)\right)\phi\left(x\right)\mathrm{d}x,$$

with, clearly,

$$r'\left(\bar{\theta}_c\right)=\left(\ell^*\left(\bar{\theta}_c\right)-w^*\left(\bar{\theta}_c,\bar{\theta}_c\right)\right)\phi\left(\bar{\theta}_c\right)-\int_0^{\bar{\theta}_c}\frac{\partial w^*\left(x,\bar{\theta}_c\right)}{\partial\bar{\theta}_c}\phi\left(x\right)\mathrm{d}x. \tag{A32}$$

Write (A22) as

$$\pi\left(\varepsilon_w,\bar{\theta}_c\right)=r\left(\bar{\theta}_c\right)-\varepsilon_w\Phi\left(\bar{\theta}_c\right), \tag{A33}$$

with

$$\frac{\partial\pi\left(\varepsilon_w,\bar{\theta}_c\right)}{\partial\bar{\theta}_c}=r'\left(\bar{\theta}_c\right)-\varepsilon_w\phi\left(\bar{\theta}_c\right).$$

From (11):

$$\frac{\partial w^*\left(\theta,\bar{\theta}_c\right)}{\partial\bar{\theta}_c}=\psi'\left(\ell^*\left(\bar{\theta}_c\right)+\bar{\theta}_c\right). \tag{A34}$$

Substitute (A34) into (A32):

$$r'\left(\bar{\theta}_c\right)=\left(\ell^*\left(\bar{\theta}_c\right)-w^*\left(\bar{\theta}_c,\bar{\theta}_c\right)\right)\phi\left(\bar{\theta}_c\right)-\psi'\left(\ell^*\left(\bar{\theta}_c\right)+\bar{\theta}_c\right)\Phi\left(\bar{\theta}_c\right). \tag{A35}$$

and so at a symmetric equilibrium, from (A27)-(A28):

$$\frac{\partial\Pi\left(\varepsilon_w,\bar{\theta}_c,\varepsilon_w,\bar{\theta}_c\right)}{\partial\bar{\theta}_c}=\frac{r'\left(\bar{\theta}_c\right)-\varepsilon_w\phi\left(\bar{\theta}_c\right)}{N}+\frac{r\left(\bar{\theta}_c\right)-\varepsilon_w\Phi\left(\bar{\theta}_c\right)}{c}\psi'\left(\ell^*\left(\bar{\theta}_c\right)+\bar{\theta}_c\right)=0, \tag{A36}$$

$$\frac{\partial\Pi\left(\varepsilon_w,\bar{\theta}_c,\varepsilon_w,\bar{\theta}_c\right)}{\partial\varepsilon_w}=-\frac{\Phi\left(\bar{\theta}_c\right)}{N}+\frac{r\left(\bar{\theta}_c\right)-\varepsilon_w\Phi\left(\bar{\theta}_c\right)}{c}=0. \tag{A37}$$

The solutions to this system of equations fully characterises the equilibrium. In order to derive the comparative static results, use (A37) to write (A36) as:

$$\frac{r'\left(\bar{\theta}_c\right)-\varepsilon_w\phi\left(\bar{\theta}_c\right)}{N}+\frac{\Phi\left(\bar{\theta}_c\right)}{N}\psi'\left(\ell^*\left(\bar{\theta}_c^0\right)+\bar{\theta}_c^0\right)=0.$$

Multiply through by $\frac{N}{\phi\left(\bar{\theta}_c\right)}$ and use (A35) to obtain:

$$\ell^*\left(\bar{\theta}_c\right)-w^*\left(\bar{\theta}_c,\bar{\theta}_c\right)=\varepsilon_w. \tag{A38}$$

The firms make zero profit on the marginal workers. The slope of this curve is obtained from total differentiation, using the expression for $w^*\left(\bar{\theta}_c,\bar{\theta}_c\right)$ obtained from (11):

$$\left(\frac{\partial\ell^*\left(\bar{\theta}_c\right)}{\partial\bar{\theta}_c}\Big(1-\psi'\left(\ell^*\left(\bar{\theta}_c\right)+\bar{\theta}_c\right)-\delta'\left(\ell^*\left(\bar{\theta}_c\right)\right)\Big)-\psi'\left(\ell^*\left(\bar{\theta}_c\right)+\bar{\theta}_c\right)\right)\mathrm{d}\bar{\theta}_c=\mathrm{d}\varepsilon_w, \tag{A39}$$

where the term multiplying $\frac{\partial\ell^*\left(\bar{\theta}_c\right)}{\partial\bar{\theta}_c}$ is zero by the envelope theorem. Write (A37) as $\left(\varepsilon_w+\frac{c}{N}\right)\Phi\left(\bar{\theta}_c\right)-r\left(\bar{\theta}_c\right)=0$ and totally differentiate to obtain:

$$\left(\left(\varepsilon_w+\frac{c}{N}\right)\phi\left(\bar{\theta}_c\right)-r'\left(\bar{\theta}_c\right)\right)\mathrm{d}\bar{\theta}_c+\Phi\left(\bar{\theta}_c\right)\left(\mathrm{d}\varepsilon_w+\mathrm{d}\left(\frac{c}{N}\right)\right)=0. \tag{A40}$$

Now substitute in the above ε_w from (A38), and $r'\left(\bar{\theta}_c\right)$ from (A35):

$$\left(\frac{c}{N}\frac{1}{h\left(\bar{\theta}_c\right)}+\psi'\left(\ell^*\left(\bar{\theta}_c\right)+\bar{\theta}_c\right)\right)\mathrm{d}\bar{\theta}_c+\mathrm{d}\varepsilon_w+\mathrm{d}\left(\frac{c}{N}\right)=0,$$

and so we have that at the intersection of the loci in the $\left(\bar{\theta}_c,\varepsilon_w\right)$ Cartesian plane which identify the points where (A36) and (A37) are satisfied, which is the simultaneous solution of (A27)-(A28), that is the symmetric Nash equilibrium:

$$-\psi'\left(\ell^*\left(\bar{\theta}_c\right)+\bar{\theta}_c\right)\mathrm{d}\bar{\theta}_c-\mathrm{d}\varepsilon_w=0,$$
$$\left(\frac{c}{N}\frac{1}{h\left(\bar{\theta}_c\right)}+\psi'\left(\ell^*\left(\bar{\theta}_c\right)+\bar{\theta}_c\right)\right)\mathrm{d}\bar{\theta}_c+\mathrm{d}\varepsilon_w+\mathrm{d}\left(\frac{c}{N}\right)=0.$$

These can be written as:

$$\begin{bmatrix} -\psi'\left(\ell^*\left(\bar{\theta}_c\right)+\bar{\theta}_c\right) & -1 \\ \frac{c}{N}\frac{1}{h\left(\bar{\theta}_c\right)}+\psi'\left(\ell^*\left(\bar{\theta}_c\right)+\bar{\theta}_c\right) & 1 \end{bmatrix}\begin{bmatrix} \mathrm{d}\bar{\theta}_c \\ \mathrm{d}\varepsilon_w \end{bmatrix}=\begin{bmatrix} 0 \\ -\mathrm{d}\frac{c}{N} \end{bmatrix}.$$

Solving for the dependent variables:

$$\begin{bmatrix} \mathrm{d}\bar{\theta}_c \\ \mathrm{d}\varepsilon_w \end{bmatrix}=\frac{Nh\left(\bar{\theta}_c\right)}{c}\begin{bmatrix} 1 & 1 \\ \frac{c}{N}\frac{1}{h\left(\bar{\theta}_c\right)}-\psi'\left(\ell^*\left(\bar{\theta}_c\right)+\bar{\theta}_c\right) & -\psi'\left(\ell^*\left(\bar{\theta}_c\right)+\bar{\theta}_c\right) \end{bmatrix}\begin{bmatrix} 0 \\ -\mathrm{d}\frac{c}{N} \end{bmatrix},$$

and so

$$\frac{\mathrm{d}\bar{\theta}_c}{\mathrm{d}\frac{c}{N}}=-\frac{Nh\left(\bar{\theta}_c\right)}{c}, \tag{A41}$$
$$\frac{\mathrm{d}\varepsilon_w}{\mathrm{d}\frac{c}{N}}=\frac{\psi'\left(\ell^*\left(\bar{\theta}_c\right)+\bar{\theta}_c\right)Nh\left(\bar{\theta}_c\right)}{c}. \tag{A42}$$

That is, an increase in competition (lower c, or higher N) increases $\bar{\theta}_c$ and reduces ε_w, establishing the last part of Proposition 5.(i).

When c is small, ε_w must be positive, as otherwise a firm would have the incentive to increase the marginal worker's wage to capture more workers. This follows from the fact that in monopsony the firm makes a positive profit on the marginal worker (compare (9) with (11)). Now, if $\varepsilon_w > 0$ then there exists c sufficiently small so that the utility of the marginal worker employed in village $\frac{1}{2N}$, and thus of all employed workers, is strictly higher than u. Now start increasing c. Denote the utility of a worker of type θ living at location t by $U(\theta, t)$, and note that

$$\frac{\mathrm{d}U(\theta,t)}{\mathrm{d}c}=\frac{\mathrm{d}\varepsilon_w}{\mathrm{d}c}+\frac{\partial w^*\left(\theta,\bar{\theta}_c\right)}{\partial\bar{\theta}_c}\frac{\mathrm{d}\bar{\theta}_c}{\mathrm{d}c}-t.$$

Substituting in from (A42), (11), and (A41), we can see that the first two terms cancel and thus

$$\frac{\mathrm{d}U(\theta,t)}{\mathrm{d}c} = -t.$$

That is, the utility of the marginal worker of the middle village is decreasing at rate $\frac{1}{2N}$ as c increases. Eventually, it hits u. This value of c is c^C, since for $c > c^C$, workers living in $t = \frac{1}{2N}$ whose type is in a right neighbourhood of $\bar{\theta}_c$ prefer not to work. This implies that the marginal worker is no longer the same in every village, proving the first part of the proposition.

By the above, $c > c^C$ implies that the distribution of the workers who are willing to work at the firm located in $t = 0$ is no longer a scaled up version of $\Phi(\theta)$. This matters, because the overall distribution of potential workers affects the job-allocation schedule, via the information rent term, which therefore will differ from (8); importantly for the qualitative features studied in Propositions 1-3, the hazard rate of the distribution of θ must be monotonic. The next result shows that the monotonicity of the hazard migrates to case where $c > c^C$.

Lemma 3. *If $h'(\theta) > 0$, then the hazard rate of the distribution of the workers who are employed by the firm is monotonically increasing.*

Proof of Lemma 3. We focus on a group for whose workers $\delta(\ell) = 0$ for every ℓ. This is to save notation, the extension to different groups is obvious. The density of the population of the workers who go to firm in $t = 0$ from the right is

$$\lambda(\theta) = \int_0^{\frac{1}{2N}} \phi(\theta)\, \zeta(\theta; t)\, \mathrm{d}t, \tag{A43}$$

where $\zeta(\theta; t)$ is the indicator function of the set of workers in $t \in \left[0, \frac{1}{2N}\right)$ who travel to work to firm in $t = 0$:

$$\zeta(\theta; t) = \begin{cases} 1 & \text{if} \quad \theta \leq \bar{\theta}_c(t) \\ 0 & \text{if} \quad \theta > \bar{\theta}_c(t) \end{cases},$$

with the obvious adjustment for locations where no one is employed, and so there is no marginal worker. Density (A43) can be written as

$$\lambda(\theta) = \int_0^{\hat{t}(\theta)} \phi(\theta)\,\mathrm{d}t, \tag{A44}$$

where $\hat{t}(\theta)$ is the function which maps the location of the village into the marginal worker in the village, which is strictly monotonic wherever is not constant, and hence invertible in the relevant range. Intuitively $\hat{t}(\theta)$ is the distance from 0 such that, given unit transport cost c, the type θ worker is the marginal worker, or more precisely, the highest θ worker who would not receive utility greater than u if they worked at the firm in $t = 0$. Formally, if $t = \hat{t}(\theta)$, and $\theta_c(t) = y$, then $\theta\left(\hat{t}(\theta)\right) = y$. Integrate (A44):

$$\lambda(\theta) = [t\phi(\theta)]_0^{\hat{t}(\theta)} = \hat{t}(\theta)\,\phi(\theta), \tag{A45}$$

and so the distribution of $\lambda(\theta)$, $\Lambda(\theta)$, is simply

$$\Lambda(\theta) = \int_0^{\theta} \lambda(x)\,\mathrm{d}x = \int_0^{\theta} \hat{t}(x)\,\phi(x)\,\mathrm{d}x.$$

Integration by parts gives $\Lambda(\theta) = \int_0^{\theta} \hat{t}(x)\,\phi(x)\,\mathrm{d}x = \hat{t}(\theta)\,\Phi(\theta) - \hat{t}(0)\,\Phi(0) - \int_0^{\theta} \hat{t}'(x)\,\phi(x)\,\mathrm{d}x$, or, using $\Phi(0) = 0$:

$$\Lambda(\theta) = \hat{t}(\theta)\,\Phi(\theta) - \int_0^{\theta} \hat{t}'(x)\,\phi(x)\,\mathrm{d}x. \tag{A46}$$

The hazard rate is computed from (A45) and (A46)

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\Lambda(\theta)}{\lambda(\theta)}\right) &= \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\hat{t}(\theta)\,\Phi(\theta) - \int_0^{\theta} \hat{t}'(x)\,\phi(x)\,\mathrm{d}x}{\hat{t}(\theta)\,\phi(\theta)}\right) = \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\Phi(\theta)}{\phi(\theta)} - \frac{\int_0^{\theta} \hat{t}'(x)\,\phi(x)\,\mathrm{d}x}{\hat{t}(\theta)\,\phi(\theta)}\right)\\ &= h'(\theta) - \frac{\hat{t}'(\theta)\,\hat{t}(\theta)\,\phi(\theta)^2 - \left(\hat{t}'(\theta)\,\phi(\theta) + \hat{t}(\theta)\,\phi'(\theta)\right)\int_0^{\theta} \hat{t}'(x)\,\phi(x)\,\mathrm{d}x}{\hat{t}(\theta)^2\,\phi(\theta)^2}.\end{aligned}$$

The hazard rate of $\Lambda(\theta)$ is non-negative if the numerator of the fraction on the RHS is is non-negative. This is the case if

$$\frac{\left(\hat{t}(\theta)\,\phi(\theta) - \int_0^{\theta} \hat{t}'(x)\,\phi(x)\,\mathrm{d}x\right)\hat{t}'(\theta)\,\phi(\theta)}{\hat{t}(\theta)\int_0^{\theta} \hat{t}'(x)\,\phi(x)\,\mathrm{d}x} > \phi'(\theta). \tag{A47}$$

A sufficient condition for (A47) to hold, in view of $-\frac{\hat{t}'(\theta)\phi(\theta)}{\hat{t}(\theta)} > 0$, is:

$$\frac{\hat{t}'(\theta)\,\phi(\theta)^2}{\int_0^\theta \hat{t}'(x)\,\phi(x)\,\mathrm{d}x} > \phi'(\theta)\,. \tag{A48}$$

Note that the only part of the marginal worker's utility that depends on his type is the information rent (that the same type living at $t = 0$ would earn). Consequently,

$$\hat{t}' = -\frac{\psi'(\ell^*(\theta) + \theta)}{c}. \tag{A49}$$

Thus, we can rewrite (A48) as

$$\frac{\psi'(\ell^*(\theta) + \theta)}{\int_0^\theta \psi'(\ell^*(x) + x)\phi(x)\,\mathrm{d}x} > \frac{\phi'(\theta)}{\phi(\theta)^2}. \tag{A50}$$

Since $\psi''(\theta) > 0$, $\int_0^\theta \psi'(\ell^*(x) + x)\phi(x)\,\mathrm{d}x > \int_0^\theta \psi'(\ell^*(\theta) + \theta)\phi(x)\,\mathrm{d}x$, (A50) becomes

$$\begin{aligned}\frac{1}{\Phi(\theta)} = \frac{1}{\int_0^\theta \phi(x)\,\mathrm{d}x} &= \frac{\psi'(\ell^*(\theta) + \theta)}{\int_0^\theta \psi'(\ell^*(\theta) + \theta)\phi(x)\,\mathrm{d}x} \\ &> \frac{\psi'(\ell^*(\theta) + \theta)}{\int_0^\theta \psi'(\ell^*(x) + x)\phi(x)\,\mathrm{d}x} > \frac{\phi'(\theta)}{\phi(\theta)^2}.\end{aligned}$$

The above shows that if $\frac{1}{\Phi(\theta)} > \frac{\phi'(\theta)}{\phi(\theta)^2}$, that is if $h'(\theta) > 0$, then $\Lambda(\theta)$ also has monotonic hazard rate, which establishes the lemma. □

Therefore, although the pay scale will in general be different, its qualitative features in particular the distortions caused by discrimination are the same as in Propositions 1-3. This establishes Proposition 5.(ii).

Finally, to prove Proposition 5.(iii), simply note that as c increases at some point the worker of type $\theta = 0$ living in $t = \frac{1}{2N}$ is indifferent between working (at either firm in $t = 0$, or firm $t = \frac{1}{N}$) and remaining unemployed. This is the value c^H, and completes the proof. □