Discrimination Spillovers, Glass Ceilings and Pay Gaps*

Gianni De Fraja†
University of Nottingham and CEPR

József Sákovics‡
Universitat de les Illes Balears

August 5, 2026

Abstract

Where identity-based discrimination varies across the levels of a job ladder, an identity-blind, profit-maximising firm that merely prices it generates pay gaps and glass ceilings for equally able workers who may suffer no discrimination themselves, even under equal pay for equal work. Two spillovers operate. Whether a group reaches the top turns only on how its treatment *changes* as it approaches the top, not on how bad conditions *are* there or below. Pay responds to discrimination at a worker's own level and, through informational rents, below it, so a gap at the top may originate entirely at the bottom.

JEL codes: 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 Bhaskar Dutta, Carmen Matutes, Daniele Checchi, Guido Friebel, Jack Fanning, Jonathan Thomas, Paul Seabright, Thomas De Fraja, and participants at seminars in Nottingham, UCD, UIB, Leicester, and Padua 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

Discrimination does not stay where it occurs. This paper shows that when members of a group[1] are badly treated in the jobs they hold at one level of an organisation, the consequences surface in the pay and job prospects of other members of that group elsewhere in the hierarchy, including those in jobs where the group is treated no differently from anybody else. The mechanism is neither prejudice on the employer's part nor mistaken beliefs about ability. It is the ordinary business of designing a pay structure that induces people to sort themselves into the jobs they suit, carried out by a firm indifferent to identity and interested only in profit.

We focus on how discrimination intensity varies with the job level (Adamovic and Leibbrandt, 2023; Dunger, 2025) not just on its overall level within an organisation. A group may face open hostility in the roles it occupies at the lowest positions and none three grades up; another may be unmolested at the bottom and find that the last step to the executive floor brings a scrutiny its colleagues escape. We take this variation as the primitive and ask how a profit-maximising employer reacts to it.

A large empirical literature, concerned mainly with gender, documents pay and achievement gaps that observable characteristics other than group identity do not explain (Gunderson, 1989; Wood et al., 1993; Bertrand, 2018). In her Nobel Lecture, Claudia Goldin emphasises the growth of "greedy" occupations, which reward long and inflexible hours even where those hours raise nobody's productivity (Goldin, 2024, p. 1532). To the extent that women are less willing to supply such hours, entirely unbiased hiring and promotion policies will nonetheless produce a gender imbalance. The mechanism complements explanations resting on differences in bargaining, in preferences for competition, and in expected attachment to the labour force (Goldin, 2014, pp. 1093–94).

Our analysis generalises the idea. Once groups differ in what it costs them to work in unpleasant conditions, and those conditions differ across the levels of a hierarchy, discrimination reaches beyond the jobs it directly affects. Even where the ablest members of a group face no barrier to senior positions, the distortion it causes in job matching lower down propagates upward and surfaces at the top, and pay gaps between equally productive workers can arise where every worker receives equal pay for equal work. To erase a pay gap at the top, an employer must heal the rot in the foundations.

We adapt the canonical principal-agent screening model, in which an employer designs compensation packages to induce workers to select, according to their ability, among the levels of a job ladder (Laffont and Martimort, 2002).[2] Higher levels are more productive but demand greater effort, stress or responsibility, and so carry a higher utility cost.

[1] By a group we mean workers sharing observable characteristics irrelevant to the job: gender, ethnicity, sexual orientation, religion, social origin, etc.

[2] As we explain later, this hiring procedure is not to be taken literally: it is a short-cut allowed by the Revelation Principle (Myerson, 1979).

The framework suits hierarchies because its equilibrium reproduces their central stylised facts: every worker but those at the very bottom earns a rent, that is, utility strictly above the outside option; and abler workers self-select into higher-level jobs, which are more demanding yet leave them both better paid and better off.

We modify this model in two ways. First, workers bear, besides the disutility of their effort that produces output, a further cost unrelated to performance, which varies with the job level and differs across groups. Section 2 discusses this term; in brief it captures the discriminatory conditions a workplace may impose: bias, conscious or otherwise, in appointment and promotion, institutionalised work culture that favours another group, and the incivility or harassment of colleagues.

Second, we allow the firm to respond by adjusting compensation, so that workers of different groups may receive different benefits at the same job level. Since explicit pay differences by group are generally unlawful, the assumption warrants a defence; it rests on the difference between pay and compensation, and Section 4.3 sets it out, together with the scales an outside observer would see.

These ingredients, together with constant returns to scale,[3] allow us to analyse each group's compensation schedule separately. We then find that a firm free to differentiate compensation across groups may optimally decline to do so. In Example 3 of Section 6, all but identical pay schedules conceal both a glass ceiling and a substantial welfare difference between otherwise identical workers. Two things follow. Adherence to equal pay for equal work is compatible with discrimination inside the organisation; and the profile of discrimination that produces this, U-shaped in the job level, is consistent with stylised facts.

Since $\delta_g(\cdot)$, the disutility discrimination imposes on group g as a function of the job level, is the one primitive we add, we should be clear about its status. It is not a parameter we are at liberty to choose, but a property of a particular workplace (how unpleasant that workplace is for members of a given group in the jobs at each of its levels), and so differs from one firm to the next. It is also, in principle, measurable: the audit and field-experimental literatures, and the survey evidence on harassment and selective incivility, measure quantities that belong to it. What this paper supplies is the mapping from that object to what an observer of the firm sees: who holds which job, the shape of each group's pay scale, and the distance between them. The useful question is which feature of $\delta_g(\cdot)$ governs which outcome, and the answer is sharp.

Job allocation responds to a spillover of infinitesimal range. The type of worker found at a given level depends on discrimination only through the slope of $\delta_g(\cdot)$ there. It is a spillover, since a slope speaks of neighbouring jobs rather than this one, but its reach is nil. A group treated abominably in the corner office will hold it anyway, provided the

[3] Constant returns rule out any technological constraint on employment: the firm hires every worker whose contribution to profit is positive, whatever the number and types of those already employed.

treatment is no worse there than one rung below; a group treated well throughout will be shut out if the last step is where its treatment deteriorates. Firms and regulators auditing how badly a group is treated at aggregated grades are measuring the wrong object. Moreover, at the very top the screening distortion vanishes, so the ablest member of each group is assigned *efficiently* given the conditions she faces, and yet the two may reach different heights. The glass ceiling is no artefact of the firm's ignorance, and better measurement of employees' cost types will not remove it.

Pay responds to a spillover that reaches every level below a worker's own, and it runs from bottom up. Because workers privately know their own suitability, the firm cannot simply post them where it wants them: it must leave them to sort themselves, and sorting has to be bought with an informational rent. Single crossing[4] makes a worker tempted to claim she has a higher cost than her true type. Consequently, the rent conceded to a worker is a function of the welfare of all the workers of her group *beneath* her, which is in turn governed by discrimination at all job levels below hers. The pay of a chief executive[5] is thus a function of how all the employees from her group are treated.

The two combine into a single relation. Along the observed pay scale, pay rises with the job level at exactly the rate at which the total disutility of the job rises for whoever holds it: the extra effort a higher post demands, plus the change in the conditions endured there. The distance between two groups' scales widens where conditions improve faster with the job level for the discriminated group, and narrows where they deteriorate faster. More discrimination at a rung below relaxes the incentive constraints the firm must respect: a worker for whom a higher job means escaping ill-treatment needs less money to accept it, so her scale carries smaller increments. Where instead conditions deteriorate with seniority, efficient assignment becomes expensive; the firm distorts the allocation, leaving overqualified workers in intermediate positions, and near the top that distortion is a glass ceiling.

The mapping is thus sharp, and also restrictive. Every result above is a sign restriction rather than a numerical prediction, settled by which of two groups faces the steeper gradient of discrimination at the level in question, with no functional form assumed anywhere; the comparisons survive even the failure of the smoothness conditions we impose. One restriction involves observables alone, the discrimination function nowhere appearing in it: the gap between two groups' scales widens exactly at the levels where the discriminated group's job-holder is the less able of the two incumbents. The model also forbids things. Make a group's treatment uniformly worse throughout the hierarchy and nobody is reallocated and no glass ceiling appears: only *variation* across levels distorts who does which job.

The consequence for measurement is uncomfortable. Discrimination functions running

[4] A widely assumed technical condition on the effort cost function.

[5] Whether or not the CEO is from a given group is determined by the slope effect discussed above.

parallel, however far apart, produce parallel pay scales and identical job assignments. Section 6 exhibits a calibration in which two groups' scales differ nowhere by as much as half of one per cent of average pay while one of them faces a glass ceiling and is excluded from the bottom quarter of the job ladder. Equal pay for equal work is worth having but settles less than has been supposed; the disparities show up instead in the distribution of jobs, which is why those reporting requirements of EU Directive 2023/970 that concern the composition of pay bands are better aimed than those concerning pay gaps.

While our theory pins down the slope of the pay gap without qualification, it only determines its level relative to an anchor, of which differences in outside options are part. Section 6.4 shows that a difference in reservation utilities of some four per cent of senior pay reverses the sign of the senior pay gap with every discrimination function held fixed, the outside-option channel being much the larger and pulling against the spillover, so that pay data alone cannot separate them.

Related literature

Our mechanism differs from both classical accounts of discrimination.[6] Under statistical discrimination (Phelps, 1972; Aigner and Cain, 1977) workers are rewarded according to group-level beliefs about productivity, which may diverge in equilibrium even between ex-ante identical groups (Coate and Loury, 1993); under taste-based discrimination (Becker, 1957; Guryan and Charles, 2013) employers harbour a distaste for particular groups. Here the firm has neither: what generates the disparities is its response to ill-treatment inflicted by others.

Closest in construction is Hurst et al. (2024), who allow discrimination to vary across tasks; because their tasks are not performed within a single organisation, no externality runs between them. Bohren et al. (2025) study spillovers across decisions affecting the same worker, whereas ours run across workers. Grout et al. (2009) derive a glass ceiling from group differences in productivity, which we exclude by assumption, and Du et al. (2022) from the aggregation of biased promotion decisions; Kawamura and Sákovics (2014) examine a related spillover from equal treatment of mid-productivity workers. Because workers here sort themselves rather than being posted, their choices determine effort and labour supply together, which connects the model to experimental work on the effects of pay differences (Breza et al., 2018) and of explicitly discriminatory ones (Gagnon et al., 2025), and to evidence on effort under discrimination (Glover et al., 2017). The non-pecuniary component of job utility through which discrimination operates is the one emphasised by Akerlof and Kranton (2005).

The model should finally be judged against what is known of internal labour markets generally. The regularities catalogued since Baker et al. (1994), namely rents above

[6] A recent survey on theories of discrimination is Onuchic (2025).

the lowest grade, monotone assignment of ability to rank and pay rising with level, are properties of the equilibrium here rather than assumptions. The tournament account of Lazear and Rosen (1981) and the learning-based account of Gibbons and Waldman (1999, 2006) explain other features that we do not, and we claim no superiority over them; what neither addresses is the transmission of a group-specific, level-specific disutility across the levels of a hierarchy, which is our subject.

Section 2 sets out our modelling of discrimination and Section 3 the model. Section 4 characterises the firm's optimal policy and Section 5 its consequences for job allocation and for pay. Section 6 works through three examples, and asks what the disclosures the law requires would reveal about them. Section 7 concludes. Proofs are in the appendix.

2 Discrimination as job-level-specific disutility

The paper turns on a single modelling device: the disutility a worker derives from a job has a term depending jointly on her group and on the level of the job. This section explains what that term represents, why we keep it separate from the cost of effort, and what shapes it can plausibly take. The formal model follows.

A worker's *group*, denoted $g \in G$, is a set of observable characteristics irrelevant to the job in hand; groups partition the workforce, so that every worker belongs to one and only one. Group membership bears in no way on how well a worker performs, the distribution of ability being the same in every group, so that our profit-maximising firm would have no reason to attend to it, were it not for discrimination. Group identity matters here for one reason only, that other people condition their conduct upon it.

Two sources of disutility

A worker of cost type θ employed at job level ℓ and paid w obtains

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

We interpret ψ as the classical cost of effort: exertion, stress, boredom, foregone leisure. These costs are intrinsic to the task, varying with how demanding the job is and with the worker's aptitude for it, but not with group identity. By contrast, $\delta_g(\cdot)$ captures the disutility generated by the conduct of others towards the jobholder: colleagues, superiors, subordinates, and outsiders such as clients, customers and suppliers. When that conduct is conditioned on the worker's identity, $\delta_g(\cdot)$ is discrimination.

A stylised example fixes the distinction. Consider two sources of disutility for a female CFO of low cost type: the difficulty of negotiating a major contract, and unwanted flirting from a colleague. The first belongs to $\psi(\ell, \theta)$: intrinsic to the role, it would generate similar stress for a male CFO of the same type and would be very large for a (high cost-type) checkout operator of either sex. The second belongs to $\delta_g(\ell)$: its incidence depends on the worker's group, and both its frequency and its intensity may differ between the two levels.

Two features bear emphasis.

Discrimination does not originate with the firm. Our firm is identity-neutral and maximises profit. It neither harbours a distaste for any group nor holds group-contingent beliefs about productivity, yet must take $\delta_g(\cdot)$ as a feature of the environment in which it recruits and assigns, in place before any contract we study is signed, and, as Section 4 shows, price it. Discrimination is in this sense a pure social cost: the disutility it imposes benefits nobody, least of all the employer, whose profit it reduces. Everything the firm does in response accommodates behaviour it does not author.

Remark 1 (Discrimination as a residual)**.** Organisations can invest in reducing discrimination, though the gains must be weighed against the inertia that makes rapid cultural change costly (Hermalin, 2001). Let an investment x_g reduce the disutility of group g from $\delta_g(\ell)$ to $(1-\rho_g(x_g))\delta_g(\ell)$, with ρ_g increasing and $\rho_g(0) = 0$. For at least some groups the profit-maximising x_g^* will satisfy $\rho_g(x_g^*) < 1$: complete elimination is too costly, particularly for conduct of the subtle kind identified by Pikulina and Ferreira (2026), hard to prove and so expensive to monitor. We accordingly read $\delta_g(\cdot)$ as the discrimination remaining once the firm has optimally reduced it, and consider x_g sunk.

Discrimination varies across levels, not across types. We write $\delta_g(\ell)$, with no θ argument. Identity-based treatment conditions on what others observe, group membership, and not on a worker's privately known cost type. There is a second reason, which needs the model to state: it keeps the informational distortion the same for every group, so that a cross-group difference in job allocation is a difference in the value of the matches and not in the severity of the incentive problem. We return to it once Theorem 1 is in hand.

Three components, and what each implies for the shape of $\delta_g(\cdot)$

In the model $\delta_g(\ell)$ is a reduced form. To fix ideas, and to discipline the shapes we later use, we decompose it into three components,

$$\delta_g(\ell) = \delta_g^1(\ell) + \delta_g^2(\ell) + \delta_g^3(\ell), \tag{2}$$

associated respectively with the assignment of positions within a level, with organisational culture, and with the conduct of colleagues and outsiders. We write $\delta_g^{i\prime}(\ell)$ for the derivative of the ith component with respect to the job level, and take each component to be twice differentiable.[7] The decomposition motivates the shapes of $\delta_g(\cdot)$ we use rather

[7] Additivity in (2) is a convenience: any increasing, twice differentiable $F(\delta_g^1(\cdot), \delta_g^2(\cdot), \delta_g^3(\cdot))$ would serve. Either way $\delta_g' = \sum_i F_i \delta_g^{i\prime}$, so its sign is settled by whichever component dominates locally, that

than constraining them: Assumption 2 below bears on the aggregate alone, and no result refers to the components individually.

We take the three in turn, drawing out in each case what it implies for the *slope* $\delta_g'(\cdot)$ rather than for the level of $\delta_g(\cdot)$, since it is the slope that Section 5 shows to govern job assignment.

Position assignment within a level. The first component, $\delta_g^1(\ell)$, arises because a job level is not a single post. At any level, positions differ in location, portfolio, reporting line and travel; workers have idiosyncratic preferences over these bundles, as in the horizontal differentiation literature (Beath and Katsoulacos, 1991). At issue is not which *level* a worker reaches (the mechanism itself determines that) but the terms on which a position *at a given level* is occupied. Appointment and promotion decisions are influenced by the subjective judgements of selectors (Schaerer et al., 2023; Benson et al., 2026), candidates (Exley and Kessler, 2022), and referees (Eberhardt et al., 2023). When a selector's conscious or unconscious bias (Hebl et al., 2018; Bertrand and Duflo, 2017) makes a worker from group g less likely to obtain the bundle they prefer, that worker's expected disutility at the level is higher.[8] Thus $\delta_g^{1\prime}(\ell) > 0$ wherever the share of group g among those making the assignment declines with seniority, or the process becomes less rule-bound as one moves up.

Organisational culture and hours. A second component reflects the "long-hours culture" of the organisation (Goldin, 2014). Let $\delta_g^2(\ell)$ denote the disutility generated by hours that are required but unproductive. The number of hours may be common across groups while the opportunity cost of time is not, so the disutility is group-specific; Checchi et al. (2025) document, in four countries, that women are under-represented among the workers supplying the longest hours. More broadly, culture shapes flexibility, expectations of availability, and other non-contractible aspects of work, in the manner of institutional discrimination (Small and Pager, 2020). As Goldin (2024, pp. 1532–33) notes, such requirements typically intensify with seniority: $\delta_g^{2\prime}(\ell) > 0$ above some level.

Workplace interactions. A third component, $\delta_g^3(\ell)$, reflects the civility or toxicity of day-to-day interaction. A substantial literature documents how harassment, bullying and selective incivility disproportionately target particular groups.[9] Its determinants are

conflict generating the non-monotone profiles of interest.

[8]Formally, suppose that at level ℓ there are n positions to be filled, and that a worker in group g derives disutility $\delta_g^{1j}(\ell)$ from position $j = 1, \dots, n$. Appointees choose sequentially, and with probability $q_g(\ell)$ the manager responsible for a position vetoes the choice of a worker from group g, who must then select another. The resulting expected disutility, $\delta_g^1(\ell)$, is increasing in $q_g(\ell)$.

[9]Cortina's seminal contribution (2008) was followed by work on the toleration of sexual harassment (Folke and Rickne, 2022) and of racist (Kern and Grandey, 2009), sexist (Cortina et al., 2013) and homophobic (Di Marco et al., 2018) remarks; Kabat-Farr et al. (2020) survey selective incivility and observe that it may help account for the scarcity of women in the upper reaches of organisations.

inherited, being fixed by past hiring and by the conduct of clients and other outsiders, over whom the firm has less control still. This component need not be monotone. It is often high at the foot of the hierarchy, where a group has historically been numerous but low in status, low in between, and high again near the top, where the group is a small and highly visible minority and sniping by disappointed rivals is concentrated.[10]

The three components rarely point the same way: interaction costs are typically worst at the point of entry and thin out thereafter, while assignment frictions and hours requirements bite hardest near the top, so which dominates changes as one climbs. What matters for the analysis is not the decomposition (2) but the profile that results, and only three profiles are worth distinguishing. Section 6 works each of them through in full.

3 The model

3.1 Workers, groups, and jobs

We study the equilibrium of a one-shot interaction between a profit-maximising firm and a pool of workers.

There is an interval of job levels $L \subseteq \mathbb{R}_+$, fixed by the firm's technology: no post outside L exists to be filled. We take L wide enough that this constraint never binds at the optimum, so that the assignment we characterise is interior. Output is contractible, sold at a unit price, and normalised to equal the job level, so that $\ell \in L$ denotes both the level of a job and the output of a worker holding it. The firm's output is the sum of individual outputs: there are no production externalities across workers.[11] A worker employed at level ℓ and paid w contributes $\ell - w$ to profit, and the firm's profit is the sum of these contributions. Constant returns to scale rule out technological constraints on employment (the firm hires any worker whose contribution to profit is positive, whatever its existing workforce), which is what allows us to treat each group's problem separately.

There is a continuum of workers. They differ, first, in a privately observed *cost type* $\theta \in [0, 1]$, which indexes the utility cost of effort: a higher θ means a higher cost of performing any given job.[12] They differ, second, in their observable group $g \in G$, in the sense of Section 2, and therefore in the disutility $\delta_g(\ell)$ that discrimination imposes on them at each job level. The notation permits any number of groups, but little is lost by thinking of two, differing only in the δ_g they face.

[10]Pascual-Fuster et al. (2025) report evidence of exactly this for female board members of Spanish firms; on the divergence of preferences at the very top see Adams and Funk (2012), and on the impostor phenomenon Price et al. (2024).

[11]A worker's output is independent of the type and number of other workers employed. This lets us study the design of job scales in isolation, and contrasts with models of team production (Costrell and Loury, 2004; Herkenhoff et al., 2024), in which co-workers' characteristics affect individual outcomes.

[12]Where we speak of *ability*, we mean the inverse: a worker of lower θ performs the same task at lower cost and is in that sense more able.

What is private here is the cost of holding a job, not the output produced in it. The distinction matters, because employers demonstrably do learn about productive ability after labour market entry (Gibbons and Katz, 1991; Altonji and Pierret, 2001; Schönberg, 2007; Kahn, 2013), and appraisal exists to that end. But an appraisal revealing how well a worker performs reveals little about what performing costs her: how she values sustained travel, the hours a senior post demands, the exposure that comes with visibility, the relocation a regional directorship requires. It is that cost, not productivity, that θ indexes.

The cost type is distributed on $[0, 1]$ according to $\Phi(\theta)$, with density $\varphi(\theta) = \Phi'(\theta) > 0$, and this distribution is the same for every group.[13] As is standard (Laffont and Martimort, 2002), we impose monotonicity of the inverse hazard rate: $h(\theta) = \Phi(\theta)/\varphi(\theta)$ satisfies $h'(\theta) > 0$.[14] Note that $h(0) = 0$.

A worker of group g who is not employed obtains a reservation utility u_g.[15] In the baseline we set $u_g = u_0 = u$ for every group, so that all cross-group differences in outcomes are attributable to discrimination suffered at the firm alone; Section 5.4 examines what changes when reservation utilities differ. When employed, workers have quasi-linear utility, as set out in (1): compensation, less the cost of effort, less the disutility generated by discrimination.

3.2 The cost of effort

We impose the following on the effort cost function.

Assumption 1 (Effort cost). *ψ is thrice continuously differentiable and satisfies, for every $\ell \in L$ and every $\theta \in [0, 1]$: (i) $\psi_\ell(\ell, \theta) > 0$ and $\psi_\theta(\ell, \theta) > 0$; (ii) $\psi_{\ell\theta}(\ell, \theta) > 0$; (iii) $\psi_{\ell\ell}(\ell, \theta) > 0$ and $\psi_{\theta\theta}(\ell, \theta) \geq 0$; (iv) $\psi_{\ell\ell\theta}(\ell, \theta) \geq 0$ and $\psi_{\ell\theta\theta}(\ell, \theta) \geq 0$.*

Part (i) says that effort is more costly in more demanding jobs and for less able workers. Part (ii) is the strict single-crossing condition: a given increase in job level raises the cost of effort by more for a worker of higher θ. This is the substantive assumption of the four, and it alone delivers the monotonicity of the assignment, thus implementability. Part (iii) makes the cost of effort convex in each argument separately. Part (iv) is a regularity

[13] Allowing Φ to differ across groups has allocative consequences of its own (De Fraja, 2005) that would confound the role of discrimination. A common Φ also removes pay differences driven by heterogeneous preferences over commuting, flexibility and related margins (Wiswall and Zafar, 2018; Petrongolo and Ronchi, 2020; Cook et al., 2021). The declining importance of physical strength in production suggests that cross-group differences in the distribution of θ are in any case smaller now than in the past.

[14] Satisfied by essentially every distribution used in applied and theoretical work (Bagnoli and Bergstrom, 2005).

[15] The reservation utility may in general depend on type; provided it declines more slowly in θ than equilibrium utility in employment, participation binds at the margin alone and taking it constant is without loss. The proviso earns its keep: an outside option falling as fast as utility in employment would let participation bind at interior types, and the problem would change class (Jullien, 2000).

requirement, guaranteeing global concavity of the firm's problem and so removing any need for random mechanisms (Laffont and Tirole, 1993, p. 66). Parts (iii) and (iv) are implied by the familiar additive specification $\psi(\ell,\theta)=\psi(\ell+\theta)$, which we do not impose; Remark 3 records where each part is used.

3.3 Discrimination

We restrict δ_g, the disutility discrimination imposes at each job level,[16] only as follows.

Assumption 2 (Discrimination). *For every $g\in G$, the function $\delta_g : L \to \mathbb{R}_+$ is non-negative and twice continuously differentiable, and satisfies, for every $\ell \in L$ and every $\theta\in[0,1]$: (i) $\delta_g'(\ell) > -\psi_\ell(\ell,\theta)$; and (ii) $\delta_g''(\ell) > -\psi_{\ell\ell}(\ell,0)$.*

Neither part restricts the sign of δ_g' or δ_g'': discrimination may be increasing or decreasing, convex or concave, and may take any of the profiles of Section 6. Both parts are one-sided, bounding only how fast δ_g may fall, and how fast it may bend downwards.

Part (i) says that the total cost of holding a job, $\psi+\delta_g$, increases in the job level: improvements in the working environment at the higher levels may offset, but not outweigh, the greater effort a senior job demands. It is *not* needed for incentive compatibility, which relies on Assumption 1(ii) alone; it delivers the familiar property that compensation rises with the level of the job, and is used in the comparisons of Section 5.

Part (ii) is a regularity condition. Define the virtual surplus generated by a type-θ worker at level ℓ as

$$S_g(\ell,\theta)=\ell-\psi(\ell,\theta)-\delta_g(\ell)-h(\theta)\psi_\theta(\ell,\theta). \tag{3}$$

Then $-\partial^2 S_g/\partial\ell^2=\psi_{\ell\ell}+h(\theta)\psi_{\ell\ell\theta}+\delta_g''$, and Assumption 2(ii) is exactly the requirement that this be positive for all θ, so that S_g is strictly concave in ℓ. Since $h(0)=0$ and $\psi_{\ell\ell\theta}\geq 0$, the binding case is $\theta=0$, which is why the assumption can be stated in the single-variable form given above. Its role is to ensure that the first-order condition characterises a global maximum and that the assignment schedule is differentiable; Remark 3 records which results depend on it.[17]

3.4 The firm's problem

The firm faces a mechanism design problem. Before workers act, it commits to a procedure that assigns each of them a job level and a wage, or leaves them unemployed, and because group identity is observable it may commit to a different procedure for each group, as in Stiglitz (1973, p. 294ff.). As discussed in the introduction, it may be able to do so even

[16] We write $\delta_g(\cdot)$ when the function itself is the subject and $\delta_g(\ell)$ for its value at a level, and drop the argument where the surrounding text supplies it. Displayed conditions always carry it.

[17] Bunching, by contrast, never arises here: with $h'>0$ and $\psi_{\ell\theta}>0$ the monotonicity constraint never binds.

where the law forbids explicit differential treatment on productivity-irrelevant grounds, for instance by offering all employees a range of in-kind options whose take-up, and whose value, differs across groups. We place no restriction on the form the procedure takes: a posted scale, an appointments committee, or an informal understanding will all serve. The restrictions are informational. Workers privately know their own cost of holding a job with given attributes; the firm does not; and a worker may always decline what she is offered and take her outside option.

The revelation principle (Myerson, 1979) is what permits this generality. Since output carries no externalities across workers, each faces a decision problem of her own, and the principle applies in its elementary single-agent form, free of the equilibrium-selection difficulties that attend it when several agents interact: whatever procedure the firm adopts, the equilibrium outcome it produces is replicated by a direct mechanism in which the firm commits to schedules $\ell_g(\theta)$ and $w_g(\theta)$ and each worker reports her type, subject to a participation constraint and an incentive-compatibility constraint. Nothing in what follows, therefore, requires that any worker ever be shown a list of posts. We work with the direct mechanism from here on, returning in Section 4.3 to the pay scale an outside observer would see.

Suppose the firm employs the types $[0, \overline{\theta}_g]$ from group g. Its profit from that group is then

$$\pi_g(\ell_g(\cdot), w_g(\cdot), \overline{\theta}_g) = \int_0^{\overline{\theta}_g} [\ell_g(\theta) - w_g(\theta)]\varphi(\theta)\mathrm{d}\theta, \tag{4}$$

and its total profit is the sum of (4) across groups, weighted by their relative sizes. Those weights play no role: because returns to scale are constant, and because the firm may condition its mechanism on the observable g, the problem separates, and the firm maximises (4) group by group. This is what licenses the notation of the rest of the paper, in which a single group is analysed at a time.

4 The optimal policy

The firm chooses the schedules $\ell_g(\theta)$ and $w_g(\theta)$ that maximise (4), subject to the requirement that no worker wishes to misreport their type and that every employed worker is willing to be employed. We characterise the solution in this section and, in Section 4.3, translate it back into the pay scales an outside observer would see.

4.1 Incentive compatibility and the optimal assignment

Given schedules $\ell_g(\theta)$ and $w_g(\theta)$, the utility of a worker of group g who reports truthfully is

$$U_g(\theta) = w_g(\theta) - \psi(\ell_g(\theta), \theta) - \delta_g(\ell_g(\theta)). \tag{5}$$

Lemma 1. *The schedules $\ell_g(\cdot)$ and $w_g(\cdot)$ are incentive compatible if and only if $\ell_g(\cdot)$ is non-increasing and, for every pair of types θ and θ',*

$$U_g(\theta) = U_g(\theta') + \int_{\theta}^{\theta'} \psi_\theta(\ell_g(x), x)\mathrm{d}x. \tag{6}$$

By (6) and Assumption 1(i), utility is strictly decreasing in the cost type. Since the reservation utility does not vary with θ, it follows that the set of workers from group g willing to be employed is an interval $[0, \overline{\theta}_g]$, and that the participation constraint reduces to the single requirement that the worst employed type be indifferent between employment and their outside option:

$$U_g(\overline{\theta}_g) = u_g. \tag{7}$$

Every abler worker then earns a rent, that is, utility strictly above u_g.

Throughout we assume that the ablest member of each group generates a positive surplus, $\ell^*_g(0) - \psi(\ell^*_g(0), 0) - \delta_g(\ell^*_g(0)) > u_g$; otherwise the firm employs nobody from group g and there is nothing to characterise.

Theorem 1. *For each $g \in G$, the job level schedule $\ell^*_g(\theta)$ that maximises* (4) *subject to incentive compatibility, as characterised in Lemma 1, and to* (7) *satisfies*

$$1 = \psi_\ell(\ell^*_g(\theta), \theta) + h(\theta)\psi_{\ell\theta}(\ell^*_g(\theta), \theta) + \delta'_g(\ell^*_g(\theta)). \tag{8}$$

The types employed are $[0, \overline{\theta}_g]$, where $\overline{\theta}_g$ is the unique solution in $(0, 1]$ of

$$\ell^*_g(\overline{\theta}_g) = u_g + \psi(\ell^*_g(\overline{\theta}_g), \overline{\theta}_g) + h(\overline{\theta}_g)\psi_\theta(\ell^*_g(\overline{\theta}_g), \overline{\theta}_g) + \delta_g(\ell^*_g(\overline{\theta}_g)), \tag{9}$$

if that equation admits one, and $\overline{\theta}_g = 1$ otherwise, in which case the whole group is employed.

Condition (8) is the familiar trade-off of the screening literature, with one addition. Were the firm able to observe θ, it would equate the marginal product of the job level, unity, to its marginal cost to the worker, $\psi_\ell + \delta'_g$. Private information adds the term $h(\theta)\psi_{\ell\theta}$, the marginal rent that must be conceded to every abler worker in order to place this one at this level, and the assignment is distorted downwards accordingly.

Remark 2 (A common informational handicap)**.** This is the second reason, promised in Section 2, for making discrimination depend on the job level alone. The wedge between the two conditions, $h(\theta)\psi_{\ell\theta}$, contains no δ: it belongs to the effort technology and to the distribution of types, both of which are common to every group, so every group labours under the same informational handicap. Had we written $\delta_g(\ell, \theta)$ the wedge would instead be $h(\theta)(\psi_{\ell\theta} + \delta_{g,\ell\theta})$ and would differ across groups: discrimination would then both shift the surplus a match generates and alter the severity of the incentive problem, and the distortions we document could not be attributed to one or the other. As it is, a cross-group difference in assignment is a difference in match values, evaluated at a common handicap.

Condition (9) determines how far down the hierarchy the firm recruits. The marginal worker is employed up to the point at which her output covers her reservation utility, her effort cost, the discrimination she suffers, and the rent that her employment obliges the firm to concede to everyone above her.

One feature of (8) deserves to be singled out at once. Read the condition not as determining the level assigned to a type, but as determining the type assigned to a level: the worker found at level ℓ is the one whose type solves

$$1 = \psi_\ell(\ell, \theta) + h(\theta)\psi_{\ell\theta}(\ell, \theta) + \delta_g'(\ell).$$

Discrimination enters only as $\delta_g'(\ell)$: as a derivative, and evaluated at ℓ itself. Neither its severity there nor its behaviour anywhere else plays any part, so a group may be treated abominably at some level and still supply exactly the same worker to it as a group treated well, provided conditions change no more sharply for the one than for the other between that level and those adjacent to it. We return to this at the end of Section 4.2, and it is the engine of the glass-ceiling results of Section 5.

Corollary 1. *There is no pooling: $\ell_g^*(\theta)$ is continuous and strictly decreasing on $[0, \overline{\theta}_g]$, and for every $g \in G$ and every $\theta \in [0, \overline{\theta}_g]$*

$$\frac{\mathrm{d}\ell_g^*(\theta)}{\mathrm{d}\theta} = -\frac{H(\theta)}{\eta_g(\theta)} < 0, \tag{10}$$

where

$$H(\theta) = (1 + h'(\theta))\psi_{\ell\theta} + h(\theta)\psi_{\ell\theta\theta} > 0, \tag{11}$$

$$\eta_g(\theta) = \psi_{\ell\ell} + h(\theta)\psi_{\ell\ell\theta} + \delta_g''(\ell_g^*(\theta)) > 0, \tag{12}$$

with all derivatives of ψ evaluated at $(\ell_g^*(\theta), \theta)$. The wage schedule implementing $\ell_g^*(\cdot)$, which by Lemma 1 and (7) is unique, satisfies, for every $\theta \in [0, \overline{\theta}_g]$,

$$\frac{\mathrm{d}w_g^*(\theta)}{\mathrm{d}\theta} = \frac{\mathrm{d}\ell_g^*(\theta)}{\mathrm{d}\theta}[\psi_\ell(\ell_g^*(\theta), \theta) + \delta_g'(\ell_g^*(\theta))]. \tag{13}$$

Read along the optimum, (13) fixes the slope of pay: as a worker of marginally lower type is moved up the hierarchy, pay rises by exactly the increase in the total disutility of the job, effort and discrimination together; the increase is positive by Assumption 2(i), so abler workers, who hold the higher-level jobs, are better paid.

The functions H and η_g recur throughout what follows. In the notation of (3), $\eta_g = -\partial^2 S_g/\partial \ell^2$ is the curvature of virtual surplus in the job level, which Assumption 2(ii) requires to be positive, and $H = -\partial^2 S_g/\partial \ell \partial \theta$ is its submodularity, which makes the assignment monotone.

Only one of the two carries a group subscript. Discrimination enters η_g, but H is built from ψ and the distribution of types alone, common to every group by assumption, and is therefore the same function for all $g \in G$. The assignment schedule is accordingly decreasing whatever the shape of δ_g: this is immediate from (10), whose sign is settled by H, and Remark 3 records that it survives even the failure of Assumption 2(ii). Discrimination alters how steeply the schedule falls but cannot reverse it. And any difference between two groups in that slope is a difference in η_g, thus by (12) in δ''_g alone, which is what makes the cross-group comparisons of Section 5 tractable: the common factor H divides out.

Remark 3 (Where the assumptions are used)**.** The assumptions do quite unequal amounts of work.

Assumption	What it delivers
1(i)	U_g strictly decreasing, by (6)
1(ii), single crossing	all of Lemma 1, and thus the rent term of (14)
1(iii), $\psi_{\ell\ell} > 0$	η_g
1(iii), $\psi_{\theta\theta} \geq 0$	uniqueness of $\overline{\theta}_g$, and nothing else
1(iv)	η_g and H respectively
2(i)	$\mathrm{d}w^*_g/\mathrm{d}\theta < 0$ in (13), and the sign in Corollary 3
2(ii)	the second-order condition (A5), and Corollary 1

Two entries deserve comment. Assumption 2 plays no part in Lemma 1, where incentive compatibility rests on the single-crossing condition 1(ii) alone, and part (i) plays none in deriving the optimal policy either: were it to fail over some range, Theorem 1 would stand unaltered, but pay would fall with rank there, conditions improving with the level by more than effort requirements worsen. And if Assumption 2(ii) fails, what is lost is continuity rather than structure. Monotonicity survives: by (3), $\partial^2 S_g/\partial \ell \partial \theta = -H(\theta) < 0$, in which δ_g does not appear, so S_g is submodular in (ℓ, θ) whatever the shape of δ_g, and $\arg\max_\ell S_g(\ell, \theta)$ is decreasing in θ in the strong set order. What is lost is continuity: ℓ^*_g jumps at isolated levels, leaving intervals of the hierarchy unfilled (Mussa and Rosen, 1978; Maskin and Riley, 1984), and the pointwise comparisons of Section 5 become comparisons in the strong set order.

4.2 Compensation

Setting $\theta' = \overline{\theta}_g$ in (6), using (7) as the boundary condition, and substituting into (5) delivers the pay schedule.

Corollary 2. *The optimal compensation schedule for group $g \in G$ is*

$$w_g^*(\theta, \overline{\theta}_g) = \underbrace{u_g}_{(i)} + \underbrace{\psi(\ell_g^*(\theta), \theta)}_{(ii)} + \underbrace{\delta_g(\ell_g^*(\theta))}_{(iii)} + \underbrace{\int_{\theta}^{\overline{\theta}_g} \psi_\theta(\ell_g^*(x), x)\mathrm{d}x}_{(iv)} . \tag{14}$$

Pay decomposes into four parts: (i) compensation for giving up the reservation utility; (ii) compensation for the effort the chosen job level requires; (iii) compensation for the discrimination suffered at that level; and (iv) the information rent, the payment required to deter the worker from claiming to be less able than she is.

The first three parts are local, depending on the worker's own type and on discrimination at her own job level only. The fourth carries the paper's message. The rent conceded to a worker of type θ is an integral over the types *worse* than her, larger the higher the job levels those workers occupy since $\psi_{\ell\theta} > 0$; and their job levels are governed, by (8), by discrimination at the levels *they* occupy, which lie strictly below hers. The pay of a worker at the top of the hierarchy is therefore a function of how workers are treated at the bottom, and the dependence runs one way: the interval of levels bearing on a wage is bounded above by the job it is paid for, so no worker's pay, and no worker's job, is affected in any degree by how those above her are treated. Single crossing makes the tempting misreport always the same: a worker is tempted to claim she is *worse* than she is, since that secures her a less demanding job, and never that she is better. Informational rents exist to defeat that one temptation, and are therefore computed against the jobs *below* the one she holds. Deviations point down, so rents flow up, which is why Section 5 has a rot-at-the-foundation result and no converse about rot in the roof.

One implication is worth recording, since a screening model containing a free function invites the suspicion that it can produce whatever is asked of it. Add a constant to δ_g at every level at once, making a group's treatment uniformly worse throughout the hierarchy, and δ_g' is unaffected: by (8) the entire assignment schedule is therefore unchanged, and nobody is reallocated. What does change is employment, which contracts, and pay, which by (14) rises by less than the constant added. Only *variation* in discrimination across levels distorts who does which job, a sharp and, in principle, refutable prediction, and not one a model with unrestricted degrees of freedom could deliver.

4.3 From the direct mechanism to observed pay scales

The direct mechanism is an analytical device; what an econometrician or a tribunal observes is a pay scale, a schedule relating compensation to job level. Because $\ell_g^*(\cdot)$ is strictly decreasing by Corollary 1, it may be inverted. Write $\Theta_g^*(\ell)$ for its inverse: the type of worker who, in equilibrium, holds a job at level ℓ. Then

$$\omega_g^*(\ell) = w_g^*(\Theta_g^*(\ell), \overline{\theta}_g), \qquad \ell \in [\underline{\ell}_g, \ell_g^*(0)], \tag{15}$$

is the profit-maximising pay scale offered to group g, where $\underline{\ell}_g = \ell_g^*(\overline{\theta}_g)$ denotes the lowest level the group fills. It has a simple slope.

Corollary 3. *For every ℓ in the interior of $[\underline{\ell}_g, \ell_g^*(0)]$,*

$$\frac{\mathrm{d}\omega_g^*(\ell)}{\mathrm{d}\ell} = \psi_\ell(\ell, \Theta_g^*(\ell)) + \delta_g'(\ell) > 0.$$

The pay scale rises, at each level, at exactly the rate at which the total disutility of the job rises for the worker who holds it. Discrimination is thus a *compensating differential in reverse* (Rosen, 1986): where conditions deteriorate with the job level the scale must be steeper to draw anyone to the higher posts, and where they improve the firm may flatten it. Two firms with identical technology and identical workers will therefore exhibit different returns to seniority if their workplaces treat the same group differently across their levels.

Two remarks on implementation. First, the scales ω_g^* need not be posted as such, and where the law forbids explicit group-contingent pay they cannot be. The firm may instead post a single basic salary scale, given by the lower envelope of the ω_g^* curves, together with a level-contingent menu of in-kind benefits from which workers in different groups select differently, so that the total value received at each level is $\omega_g^*(\ell)$ for a member of group g.[18]

Second, the intervals over which the two scales are defined differ across groups, both endpoints depending on δ_g; but the firm may extend each scale over the union of the two, offering outside a group's range contracts no member of it would accept. The scales are then defined over an identical range of job levels, while the ranges actually *filled* differ.

An outside observer comparing what the firm offers the two groups therefore finds nothing, at either end. At the top, the discriminated group may reach less high, by Proposition 2 below; this is how a glass ceiling looks from the outside: the senior posts advertised, open on identical terms to everybody, no vacancy rationed, and nobody from the discriminated group applying. At the bottom, by (9), the firm may not recruit as deeply into the discriminated group, so its lowest filled job is the higher of the two and the group is absent from the bottom grades altogether. That is emphatically not a benefit: fewer of its members hold a job at all. Exclusion at the foot of the ladder and exclusion at its head are concealed by the same device.

[18] Some workers in group g will prefer the benefits intended for group 0, and conversely. Handling this properly would require a multidimensional screening problem (McAfee and McMillan, 1988); we assume instead that such cases are few enough for the firm to disregard them in designing the menu. The question scarcely arises in the case that concerns us most, the two scales being almost indistinguishable there, so that there is little for a worker to prefer.

The defence of group-contingent compensation promised in the introduction rests on this same distinction between pay and compensation. A firm may offer perks and in-kind benefits whose valuation differs across groups: flexible working, childcare, subsidised leisure, conference travel, tickets to cultural and sporting events.[19] In addition, the firm may operate distinct pay scales across career tracks at similar levels, rewarding differentially the groups with differing preferences over those tracks.[20] Finally, the firm may shape culture and promotion criteria so as to tilt allocation and advancement towards particular groups while remaining hard to characterise as explicitly discriminatory (Pikulina and Ferreira, 2026).

What, then, would the firm do if equal pay for equal work could be made to bind, a single schedule for everybody? Within the model, nothing: the requirement binds on pay, the firm complies through the composition of compensation, and the value received at each level is $\omega_g^*(\ell)$ regardless. To have content it would have to bind on total compensation, non-pecuniary terms included, which would require a regulator to observe group-specific valuations of exactly the kind δ_g is and that nobody measures. Were the exercise nevertheless attempted, a single schedule would have to screen a population two-dimensional in (θ, g), and the problem changes class: exclusion and bunching are generic in multidimensional screening (Armstrong, 1996; Rochet and Choné, 1998), so the objects this paper compares, the type holding each level and what each scale pays there, cease to be well defined level by level. The smooth problem solved here is the benchmark any such exercise would have to start from.

5 The consequences of discrimination

We now compare two groups. Group g is discriminated against; group 0 is the reference group. They face discrimination functions δ_g and δ_0, and in everything that follows they are alike in every other respect: the same distribution of cost types, the same effort cost function, and the same reservation utility $u_g = u_0 = u$. Any difference in what happens to them is therefore attributable to δ_g versus δ_0 and to nothing else.

[19] In-kind benefits can be substantial. Following the pandemic, workers in several countries appear to converge on a valuation of around seven per cent of salary for the option of working from home two or three days a week (Barrero et al., 2021; Nagler et al., 2022; De Fraja et al., 2025), the last documenting statistically significant differences in valuation between women and men.

[20] A major UK case illustrates the possibility and its limits. In *Asda Stores Ltd v Brierley* [2021] UKSC 10, the Supreme Court held on 26 March 2021 that shop-floor employees of Wal-Mart's UK subsidiary, who are predominantly female, may use the better-paid warehouse employees as comparators for the purposes of an equal pay claim, permitting their claims for the lower pay they had received to proceed. Cases of this kind help explain the scepticism with which some regard anti-discrimination legislation (Posner, 1989). Stylised facts and some evidence (De Fraja et al., 2019) indicate that professorial pay tends to be higher in economics, finance and computer science than in 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 its having been alleged.

The last of these equalities is a device and not a description. Groups treated badly inside firms are, as a rule, treated badly outside them too, so the empirical case for $u_g = u_0$ is weak; we impose it because, were reservation utilities allowed to differ as well, every disparity we report below could be attributed to the disparity in outside options rather than to the mechanism we wish to isolate. It also costs remarkably little, for a reason set out in Section 5.4: the reservation utility does not appear in the assignment condition (8) at all, so it cannot affect who does which job, and it enters pay only through the extensive margin. Every result about derivatives below is untouched by it; only the levels require the assumption.

Throughout, we compare the two groups at job levels that both of them fill, and we write $\mathcal{L} = [\underline{\ell}_g, \ell_g^*(0)] \cap [\underline{\ell}_0, \ell_0^*(0)]$ for the set of such levels, assumed non-empty.

5.1 Job assignment and the glass ceiling

Which of the two groups supplies the worker for a given post, and how a given worker is placed relative to her counterpart in the other group, turn out to depend on the two discrimination functions in only one way.

Proposition 1. *(i) For every $\ell \in \mathcal{L}$,*

$$\operatorname{sign}[\Theta_g^*(\ell) - \Theta_0^*(\ell)] = -\operatorname{sign}[\delta_g'(\ell) - \delta_0'(\ell)]. \tag{16}$$

(ii) For every type θ employed by both groups, $\delta_g'(\ell) - \delta_0'(\ell)$ has a constant sign on the open interval between $\ell_0^*(\theta)$ and $\ell_g^*(\theta)$, and $\ell_g^*(\theta) - \ell_0^*(\theta)$ carries the opposite sign.

Suppose conditions deteriorate faster with the job level for group g than for group 0 at some level, $\delta_g'(\ell) > \delta_0'(\ell)$. A job past that point is then more costly for a member of group g to accept and the firm must concede more to fill it with her, so the cheaper course is to assign her a lower job. Every type in group g is accordingly placed lower than her counterpart in group 0, and, read the other way, the group-g worker found at that level is the abler of the two. Where conditions improve faster with the job level for group g the reverse holds: its members are placed *higher* than equally able members of the reference group, being cheaper to attract to the higher jobs.

Figure 1 makes the point visible, and makes visible at the same time how local it is. The reference group suffers no discrimination, so $\delta_0' \equiv 0$ and (16) reduces to $\operatorname{sign}[\Theta_g^*(\ell) - \Theta_0^*(\ell)] = -\operatorname{sign}\delta_g'(\ell)$. Let δ_g be a sinusoid. Then the two assignment schedules cross at precisely those levels at which δ_g is flat,[21] and nowhere else, whatever the amplitude of the oscillation: at the peaks and troughs the two groups supply workers of identical ability to the same post, and between them the ranking reverses. Panel (c) anticipates

[21] What is doing the work is not flatness but *parallelism*: the schedules meet where $\delta_g' = \delta_0'$, and here the reference group's function happens to be horizontal, so the two coincide only where δ_g is horizontal too.

Proposition 3 and Corollary 5: the pay gap is stationary at exactly those levels, and, the figure being drawn with the parametric family of Section 6, it returns at every trough to the same value, which the calibration makes zero.

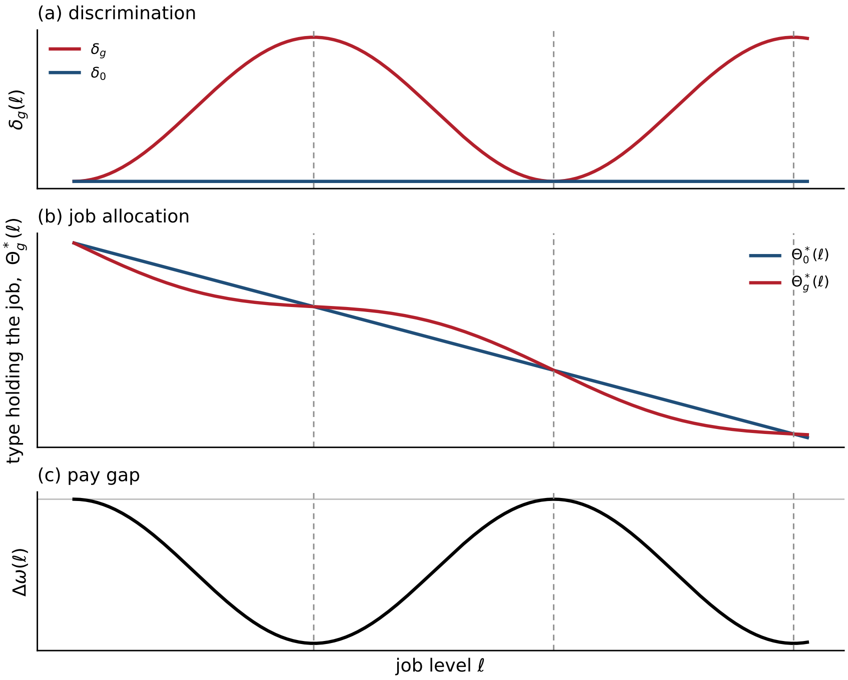

Figure 1: Sinusoidal discrimination. The dashed verticals mark the levels at which δ_g is flat. The assignment schedules cross at these levels and at no other, and the pay gap is stationary there.

The figure also disposes of an intuition that is hard to shake, namely that a group facing more discrimination should be found lower down. Here $\delta_g > \delta_0 = 0$ over the whole range, yet group g holds the better jobs over half of it and the worse jobs over the other half, the alternation being governed by the derivative of a function whose sign never changes.

The most consequential application of Proposition 1 is at the top of the ladder. Say that group g faces a *glass ceiling* relative to group 0 if the ablest member of group g is assigned a lower job than the ablest member of group 0, that is if $\ell^*_g(0) < \ell^*_0(0)$.

The phrase ordinarily connotes rivalry, a contested slot filled at another's expense, and there is none of that here: under constant returns nothing at the top is scarce, and the firm would as happily employ both groups' ablest at one level as either alone. What remains once rivalry is removed is the component the definition isolates, the level past which, given the gradient of its treatment, placing even a group's ablest member stops being worth the firm's while: a ceiling that unlimited room at the top does nothing to lift, just as, by part (i) of the proposition below, complete information does nothing to lift it.

Proposition 2. *(i) The ablest worker of group g is assigned the level that maximises $\ell - \psi(\ell, 0) - \delta_g(\ell)$, so that $\ell_g^*(0)$ satisfies*

$$1 = \psi_\ell(\ell_g^*(0), 0) + \delta_g'(\ell_g^*(0)). \tag{17}$$

Her assignment therefore bears no screening distortion, though it is not free of discrimination, which enters (17) *through δ_g'.*

(ii) $\ell_g^*(0) = \ell_0^*(0)$ if and only if $\delta_g'(\ell_0^*(0)) = \delta_0'(\ell_0^*(0))$; and when the two tops differ, $\operatorname{sign}[\ell_0^*(0) - \ell_g^*(0)] = \operatorname{sign}[\delta_g'(\ell) - \delta_0'(\ell)]$ *for every ℓ strictly between them. Group g therefore faces a glass ceiling if and only if the open interval $(\ell_g^*(0), \ell_0^*(0))$ is non-empty and $\delta_g' > \delta_0'$ on it.*

Part (i) is worth dwelling on. At the very top the screening distortion disappears, because $h(0) = 0$: there is nobody abler than the ablest worker to whom rent must be conceded, so the firm places her at the level that is efficient given the conditions she faces. *The glass ceiling is therefore not an informational distortion at all.* It is the firm's efficient response to a gradient in discrimination, and would persist unchanged if the firm could observe every worker's type. That has an uncomfortable implication for policy: measures improving the firm's information about its workers, however desirable on other grounds, will do nothing whatever about the glass ceiling.

The locality point takes its sharpest form with any number of groups: the highest job filled in the firm, $\ell_M = \max_{g \in G} \ell_g^*(0)$, is held by the ablest member of a group in $\arg\min_{g \in G} \delta_g'(\ell_M)$, the amount of discrimination suffered there being irrelevant. Firms and regulators who audit the *level* of discrimination at senior grades are measuring the wrong object.

No comparable statement can be made at the other end. The lowest job a group fills is $\ell_g^*(\overline{\theta}_g)$, and while the schedule $\ell_g^*(\cdot)$ is free of reservation utilities by Proposition 5(i), the marginal type is not: a group is absent from the bottom grades both when it is treated badly there and when its outside option is good, and the two act in the same direction. Exclusion at the foot of the ladder is therefore real, as Section 6 shows, but not identified from job data alone in the way the ceiling is.

5.2 The pay gap

We turn to compensation. Write

$$\Delta\omega(\ell) = \omega_0^*(\ell) - \omega_g^*(\ell)$$

for the pay gap at job level ℓ: the difference between what a member of the reference group and a member of the discriminated group are paid for doing the same job. Note that, by Proposition 1, these workers need not be of the same cost type.

The remaining analysis rests on one formula. Recall from Corollary 3 that the pay scale rises at rate $\psi_\ell + \delta_g'$, and note that the assignment condition (8), read at level ℓ, says precisely that this quantity equals $1 - h(\Theta_g^*(\ell))\psi_{\ell\theta}(\ell, \Theta_g^*(\ell))$. Writing

$$m(\ell, \theta) = h(\theta)\psi_{\ell\theta}(\ell, \theta)$$

for the marginal informational rent at level ℓ of a worker of type θ, we therefore have the following.

Proposition 3. *For every $g \in G$ and every ℓ in the interior of group g's range,*

$$\frac{\mathrm{d}\omega_g^*(\ell)}{\mathrm{d}\ell} = 1 - m(\ell, \Theta_g^*(\ell)) \leq 1, \tag{18}$$

and consequently, for every $\ell \in \mathcal{L}$,

$$\Delta\omega'(\ell) = m(\ell, \Theta_g^*(\ell)) - m(\ell, \Theta_0^*(\ell)), \tag{19}$$

so that

$$\operatorname{sign} \Delta\omega'(\ell) = \operatorname{sign}[\Theta_g^*(\ell) - \Theta_0^*(\ell)] = -\operatorname{sign}[\delta_g'(\ell) - \delta_0'(\ell)]. \tag{20}$$

Four things follow.

First, pay rises more slowly than output. By (18), the shortfall at each level is exactly the marginal rent m that must be conceded to abler workers in order to place this one there. It follows from that comparison, and from it alone, that profit per worker increases with rank, at rate $m(\ell, \Theta_g^*(\ell))$; since $h(0) = 0$ the rate falls to zero at the top. The firm's ablest employee is thus the one it profits from most, and simultaneously the one whose assignment it would gain nothing by raising.

Second, and substantively, the gap widens at exactly those levels where conditions improve faster with the job level for the discriminated group, and narrows where they deteriorate faster. This is (20). Where the higher jobs bring better treatment for group g, the firm needs less money to draw its members to them; improvement in the workplace substitutes for pay; and the substitution is made at their expense.

The first equality in (20) stands on its own, and it is the one an empiricist can use. It links two observables and mentions δ nowhere: the pay gap between the scales widens, level by level, exactly where the group-g worker holding the job is the less able of the two incumbents, and narrows exactly where she is the abler; the scale of the group whose incumbent is the abler rises faster, by (18), because less rent is being conceded at that level. Every other comparison in the paper conditions on a discrimination function that nobody measures; this one does not. Pay scales are observed directly, and the relative ability of the two groups' incumbents at a level is the object that matched employer–employee data, with qualifications and experience as proxies, exist to estimate. The theory thus imposes a cross-restriction on two schedules estimable from personnel records alone: a firm in which the pay gap widens across a range of levels at which the discriminated group's incumbents are also the relatively abler is a firm this model cannot describe.

Two quantities are involved here, and they can move in opposite directions. The increment separating adjacent levels of the scale is $\psi_\ell + \delta'_g$ by Corollary 3, so where the higher job brings relief the scale is flatter than the cost of effort alone would make it, and less is needed at each step to secure self-selection. The rent accruing to a worker already in post is accumulated over all the workers beneath her, by term (iv) of (14), and is larger the higher those workers are placed, $\psi_{\ell\theta}$ being positive. Where $\delta'_g < \delta'_0$ below her, Proposition 1 places them higher than their counterparts in group 0, so her rent is the larger even as the increments of her scale are the smaller; the reverse inequality reverses both. It is the accumulated rent, not the schedule of increments, that carries conditions at the foot of the hierarchy to the top. Proposition 4 below is about it.

Third, discrimination functions that run parallel are invisible in pay. The corollary is the reason any of this is hard to detect.

Corollary 4. *If $\delta'_g(\ell) = \delta'_0(\ell)$ for every ℓ in an interval $I \subseteq \mathcal{L}$, then $\Delta\omega$ is constant on I, and the two groups supply workers of identical ability to every level in I.*

Two discrimination functions that run parallel, however far apart, produce pay scales that also run parallel, and job assignments that are identical.[22]

It does not follow that nothing is visible. A non-zero constant $\Delta\omega$ is still a difference in pay, and the same difference at every level in I: the most conspicuous form a disparity can take, and in most jurisdictions the form the law prohibits outright. What is invisible is the structure. There is no glass ceiling within I, the group-g and group-0 workers holding any given post there are equally able, and there is no level at which the gap opens or closes and which might be identified as the source of the trouble. The disparity presents as a flat differential with no visible origin, and by Proposition 4 below its origin is indeed elsewhere, at the levels below I where the two functions are not parallel. The case in which the constant is close to zero, and nothing at all is visible, is not a knife-edge: Section 6 exhibits a calibration in which the two scales nowhere differ by as much as half of one per cent of average pay while a glass ceiling and a substantial difference in treatment are present throughout.

[22] Figure 1 is the extreme case of this. There the two functions are parallel at isolated levels only (the turning points of the sinusoid), so the interval I degenerates to a point, and what Corollary 4 delivers is not a constant gap over a range but a stationary point of the gap at each of them, which is panel (c).

Fourth, the sign in (20) *comes with a magnitude, and the magnitude is a deficit.*

Corollary 5. *For every ℓ in the interior of $\mathcal{L}$ there is a type $\widetilde{\theta}(\ell)$ between $\Theta_0^*(\ell)$ and $\Theta_g^*(\ell)$ such that*

$$\Delta\omega'(\ell) = -\kappa(\ell)[\delta_g'(\ell) - \delta_0'(\ell)], \qquad \kappa(\ell) = \frac{m_\theta(\ell, \widetilde{\theta})}{\psi_{\ell\theta}(\ell, \widetilde{\theta}) + m_\theta(\ell, \widetilde{\theta})} \in (0, 1), \tag{21}$$

so that the pay gap moves at every level in the direction opposite to the discrimination gap and strictly more slowly: $|\Delta\omega'(\ell)| < |\delta_g'(\ell) - \delta_0'(\ell)|$ wherever the latter is not zero. If, moreover, $\psi(\ell, \theta) = \alpha(\ell + \theta)^2 + \beta(\ell + \theta)$ and cost types are uniform, then $\kappa \equiv \frac{1}{2}$, and

$$\Delta\omega(\ell) + \tfrac{1}{2}[\delta_g(\ell) - \delta_0(\ell)] \tag{22}$$

is constant on $\mathcal{L}$.

Compensation for discrimination is systematically incomplete. Between one rung and the next, whatever conditions do, pay does less: the firm makes up in money a fraction $\kappa < 1$ of any deterioration, because it makes up the remainder in kind, by Proposition 1 passing the job to an abler worker, who is cheaper to keep incentive-compatible there. The split between the two margins of adjustment is governed by (21): money bears the share that the marginal rent m contributes to the cost of self-selection, reassignment the share contributed by the cost of effort itself. In the parametric family of Section 6 the two contributions are equal, and the split is exactly half-and-half at every level: the pay gap traces the discrimination gap in mirror image at half scale, so that (22) is a conserved quantity of the pay policy—the gap returns to the same value at every level at which $\delta_g - \delta_0$ returns to the same value, however violently either moves in between.

5.3 Where a senior pay gap comes from

We can now state the paper's central result. Proposition 3 determines how the gap evolves but not where it starts, and the two questions turn out to have quite different answers. We take the evolution first, because it is where the spillover lives.

Proposition 4. *Let $\widehat{\ell}$ be a level filled by both groups, with $\widehat{\ell} < \ell_0^*(0)$, suppose $\delta_g(\ell) = \delta_0(\ell)$ for every $\ell \geq \widehat{\ell}$, and write $\widehat{\theta} = \Theta_0^*(\widehat{\ell})$. Then*

(i) neither group faces a glass ceiling, and the two groups supply workers of identical ability to every job level above $\widehat{\ell}$;

(ii) the pay gap is constant above $\widehat{\ell}$, and equal to

$$\Delta\omega = \int_{\widehat{\theta}}^{\overline{\theta}_0} \psi_\theta(\ell_0^*(x), x)\mathrm{d}x - \int_{\widehat{\theta}}^{\overline{\theta}_g} \psi_\theta(\ell_g^*(x), x)\mathrm{d}x. \tag{23}$$

This is the message of the paper stated as a theorem rather than illustrated by an example. Above $\widehat{\ell}$ there is nothing for an audit to find: the same δ, the same assignments, the same people in the same jobs, and no glass ceiling. And there is a pay gap at every one of those levels, of a magnitude given by (23), in which not a single term refers to any job level above $\widehat{\ell}$. The rot is in the foundations and the subsidence appears in the penthouse.

Expression (23) also shows why the sign of the gap is not settled by the theory alone. Suppose group g is the more discriminated one below $\widehat{\ell}$, with $\delta_g \geq \delta_0$ there and $\delta_g - \delta_0$ non-increasing, so that $\delta_g' \leq \delta_0'$ throughout: the case, for instance, in which δ_g declines steadily to meet a flat δ_0 at $\widehat{\ell}$. (Convergence at $\widehat{\ell}$ on its own delivers the ranking of slopes only at $\widehat{\ell}$ itself, not below it.) Two forces then pull in opposite directions. The first is *placement*. By Proposition 1, $\ell_g^* \geq \ell_0^*$ below $\widehat{\ell}$: conditions improve steeply with the job level for group g, so its members are cheap to attract to higher jobs and the firm places them higher than equally able members of group 0. Those below any given worker therefore hold better jobs, and since $\psi_{\ell\theta} > 0$ the rents conceded to stop her imitating them are larger, making the second integral in (23) the larger and pushing $\Delta\omega$ negative: group g is paid *more* at each senior level. The second is the *extensive margin*. By (9), $\delta_g \geq \delta_0$ at the bottom implies $\overline{\theta}_g \leq \overline{\theta}_0$, which shortens the range of integration in that integral and pushes $\Delta\omega$ positive.

Which force dominates is a quantitative matter, and Section 6 shows that the first can dominate under an unremarkable calibration. It is easy to assume the answer, so it is worth being plain. Where the first dominates, a group treated worse at the foot of the hierarchy is paid *above* the reference group at every senior level, and its senior members are not merely better paid but strictly better off, holding the same jobs under the same conditions and drawing more for them. Its junior members, who are the ones actually ill-treated, are worse off than their counterparts, and at the margin its members are not employed at all, since $\overline{\theta}_g < \overline{\theta}_0$. Discrimination concentrated at the bottom of a hierarchy does not depress a group uniformly; it redistributes within it, from those who suffer it to those who do not.

Where the two groups do not reconverge, the level of the gap needs an anchor, and the natural one is that they are treated alike at the foot of the hierarchy. If $\delta_g = \delta_0$ up to some level above $\underline{\ell}_0$, then (9) is the same equation for both groups, so they share the marginal type and the lowest job filled and are paid the same there: the extensive-margin force is absent and $\Delta\omega(\underline{\ell}) = 0$. The gap above is then the accumulated placement effect alone, and by (20) a group whose conditions improve with the job level no faster than the reference group's, $\delta_g' \leq \delta_0'$, is paid no more than it anywhere, and strictly less above any level at which the two functions have already begun to diverge. This is the compensating differential in reverse of Corollary 3 at work: an improvement in conditions that arrives with the higher job is an improvement the firm need not pay for.

5.4 Reservation utilities

We have assumed $u_g = u_0$. It remains to say what changes without it, and the answer is: less than one might expect.

Proposition 5. *Let $\delta_g(\cdot)$ and $\delta_0(\cdot)$ be given and let $u_g \neq u_0$.*

(i) The assignment schedules $\ell^*_g(\cdot)$ and $\ell^*_0(\cdot)$, and thus Propositions 1 and 2, are unaffected.

(ii) $\Delta\omega'(\ell)$ is unaffected, so Proposition 3 and Corollaries 4 and 5 continue to hold verbatim.

(iii) $\overline{\theta}_g$, where interior, is strictly decreasing in u_g. In particular, if $\delta_g = \delta_0$ in a neighbourhood of the lowest filled levels, then $\overline{\theta}_0 \gtreqless \overline{\theta}_g$ according as $u_0 \lesseqgtr u_g$.

Reservation utilities therefore act on the extensive margin and nowhere else, determining how far down each group the firm recruits without touching the assignment of workers to jobs or the rate at which the two scales diverge. They do move the *level* of the pay gap, and by enough to reverse its sign; Section 6.4 sets out the two channels through which they do so and measures them. The consequence for us is that the spillover of Proposition 4 survives intact but is superimposed on a level effect of the conventional kind, from which pay data alone cannot separate it. That is our reason for equalising reservation utilities in the baseline: a reason of exposition rather than of realism.

6 Three examples

Which of the three components of Section 2 dominates changes across the levels of a hierarchy, and δ_g takes its shape accordingly. Three shapes are worth distinguishing, and this section works one example of each through in full: discrimination that fades with the job level, discrimination that mounts, and discrimination at both ends of the ladder. The three examples share a calibration and differ only in δ_g, so the comparison across them isolates the profile.

Throughout, $\psi(\ell, \theta) = \alpha(\ell + \theta)^2 + \beta(\ell + \theta)$ with $\alpha = 0.4$ and $\beta = 0.1$, cost types are uniform on $[0, 1]$ so that $h(\theta) = \theta$, and $u = 0$.[23] The reference group suffers no discrimination, $\delta_0 \equiv 0$. With these values the firm employs the types $[0, 0.267]$ of an undiscriminated group and fills the job levels $[0.590, 1.125]$, and we place the interesting behaviour of δ_g inside that range, with a transition at $\ell_m = 0.87$. The technology admits the job levels $L = [0.55, 1.2]$, which contains the filled range with room at either end;

[23] Nothing depends on these forms; small perturbations leave every qualitative feature intact. A single script reproduces the four figures and every statistic reported below; it is included in the replication package.

on L the specifications below satisfy Assumption 2 throughout. Reservation utilities are equal throughout, $u_g = u_0 = 0$: the examples are designed to isolate the consequences of δ_g, and Section 6.4 shows what changes when outside options differ. Example 1 takes $\delta_g(\ell) = v((\ell_m - \ell)/W)^3$ for $\ell < \ell_m$ and zero above; Example 2 its mirror image, zero below ℓ_m and $v((\ell - \ell_m)/W)^3$ above; and Example 3 the sum of the two. In each, $v = 0.03$ and $W = 0.28$. The cube rather than the square makes δ_g twice continuously differentiable at ℓ_m, where δ_g, δ'_g and δ''_g all vanish, so that Assumption 2 holds as stated. Each figure has the same four panels, the left-hand column indexed by the job level and the right-hand column by the cost type: (a) the discrimination functions; (b) the assignment schedules $\ell^*_g(\theta)$; (c) the pay gap $\Delta\omega(\ell)$, as a percentage of group 0's average pay; and (d) the welfare difference $U_g(\theta) - U_0(\theta)$, as a percentage of the same average pay. Utility is quasi-linear, so the welfare difference is a compensating variation, measured in the same units as pay, and the two lower panels are directly comparable. In (d) the curve is solid over the types the firm employs from both groups, and dashed beyond, where it employs from only one.

6.1 Discrimination that fades

In the first example δ_g declines to zero at ℓ_m and is zero above it: the third component dominating, with hostility and incivility concentrated in the jobs a group holds at the foot of the hierarchy and thinning out above them. Above ℓ_m the two groups are treated identically.

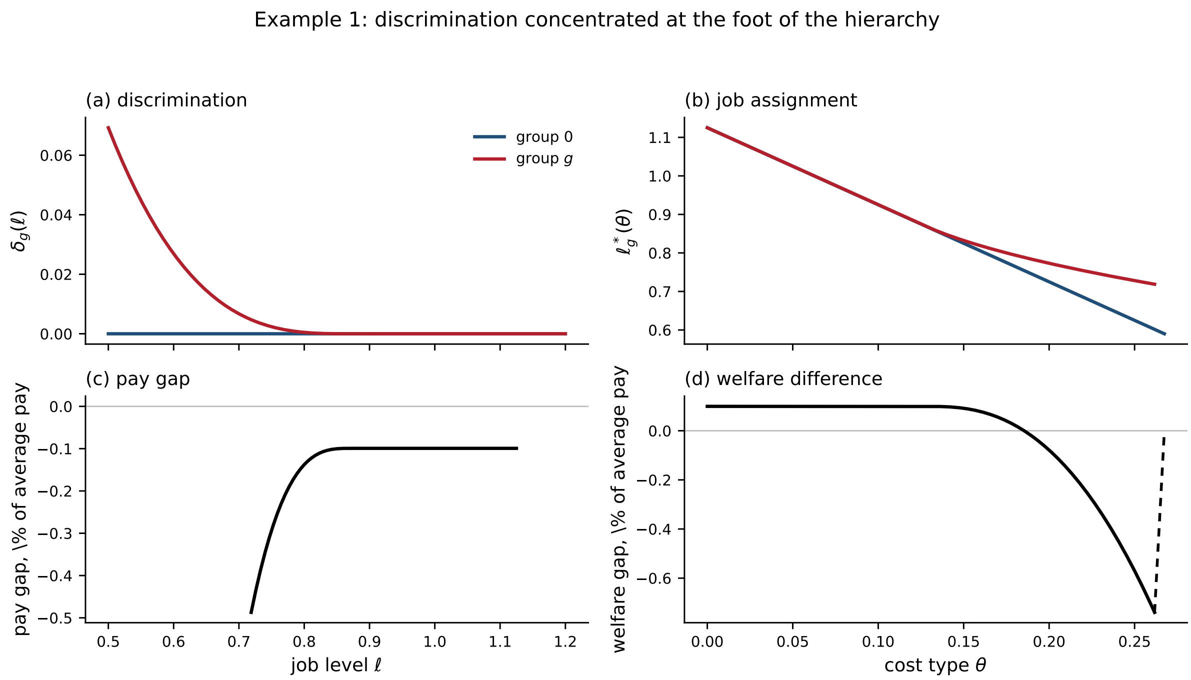

Figure 2: Discrimination concentrated at the foot of the hierarchy.

The predictions of Proposition 4 are visible in Figure 2. There is no glass ceiling:

both groups reach $\ell_g^*(0) = 1.125$. Above ℓ_m the assignment schedules coincide exactly, so the two groups supply workers of identical ability to every senior post. And the pay gap, panel (c), is constant above ℓ_m, at -0.10 per cent: a permanent difference in pay at every level of the hierarchy above ℓ_m, at which no discrimination of any kind is present.

The sign repays attention: it is group g, discriminated against lower down, that is paid *more* at each senior level. This is the placement force of Proposition 4 dominating the extensive-margin force. Below ℓ_m conditions improve steeply with the job level for group g, so its members are cheap to attract to higher jobs and are placed above equally able members of group 0, as panel (b) shows, the red schedule lying above the blue at low levels. The rents required to stop senior workers imitating them are correspondingly larger. The countervailing force is present but weaker: the firm recruits group g only down to $\overline{\theta}_g = 0.262$ against $\overline{\theta}_0 = 0.267$, and excludes it altogether from job levels below 0.719.

The size of that gap depends on how deep the discrimination below it runs, and monotonically so. Raising the peak of δ_g over the filled range from 0 to 0.075, holding everything else fixed, moves the senior pay gap steadily from 0 to -0.14 per cent of average pay. A family of groups differing only in how badly they are treated at the foot of the hierarchy therefore fans out at the top, in the order given by (23), though every one of them is treated identically there.

This is not to say that group g does better, and panel (d) shows why. Its ablest members are better off by 0.10 per cent of average pay, but the difference falls with the cost type, changes sign at $\theta \approx 0.185$, and reaches -0.74 per cent at $\overline{\theta}_g$; the dashed arm beyond records the types group 0 employs and group g does not. So the premium at senior levels compensates a disadvantage borne by juniors rather than by those who receive it, and while the balance aggregated over the group is negative, no senior member of it is worse off than her counterpart. That a group can be paid a premium at every senior grade and still be the group discriminated against is, we think, the least intuitive implication of the model, and the one most likely to be missed by an observer working from pay data.

6.2 Discrimination that mounts

The second example reverses the profile: δ_g is zero up to ℓ_m and rises above it, as the first two components generate: assignment frictions biting harder where appointments are least rule-bound, and a long-hours culture intensifying with seniority. The two groups are now treated identically at the bottom and diverge at the top.

Now there is a glass ceiling. The ablest member of group g reaches only $\ell_g^*(0) = 1.016$ against 1.125 for group 0, and by Proposition 2(i) this is not an artefact of the firm's ignorance: both assignments are efficient given the conditions each group faces. Below ℓ_m

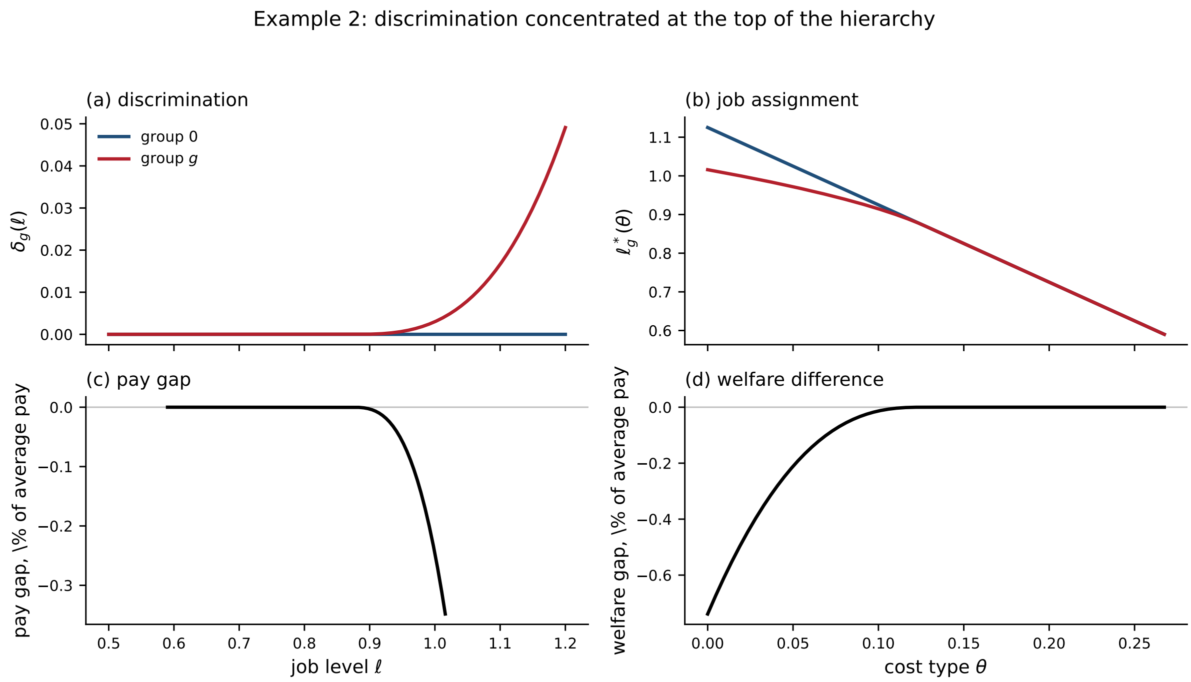

Figure 3: Discrimination concentrated at the top of the hierarchy.

nothing whatever distinguishes the two groups, in assignment or in pay; above it, group g is placed lower and, by Proposition 3, is paid *more* at any level it does reach, by up to 0.35 per cent at the highest common level. The firm must buy its consent to conditions group 0 does not face.

Measured level by level, group g is paid at or above group 0 everywhere; on average across its members it is paid 96.7 per cent of what group 0 is paid, and in the top quarter of the pay distribution 93.5 per cent. Job allocation reconciles the two: the pay gap is entirely a gap in the jobs held, not in the rate for the job. A firm in this position complies with equal pay for equal work exactly and reports a substantial raw pay gap, both facts being consequences of the same discrimination. This configuration (a disutility rising with job level, producing a wage premium at the levels reached together with exclusion from those above) is the one a reader is most likely to construct when first meeting the model.

6.3 Discrimination at both ends

The third example combines the two: δ_g falls to zero at ℓ_m and rises again above it. This U-shape is on the evidence the most plausible of the three. Bircan et al. (2024) find that workers in low-level jobs face discrimination that declines with the job level, which delivers the left arm, and the considerations of Section 2 imply a reversal higher up. Its inversion (a group insulated at the bottom, in roles conventionally associated with it, and again at the top, where status protects, but exposed at the supervisory grades in between) is obtained by reflecting the first panel, and reverses every comparison below.[24]

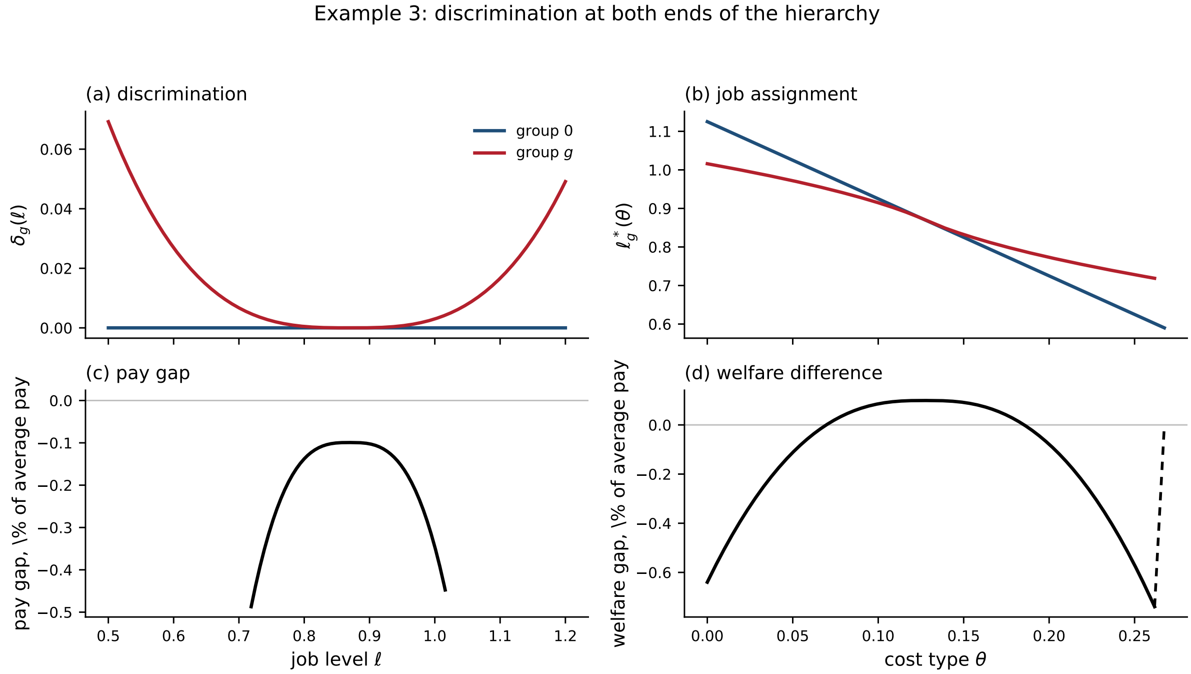

Figure 4: Discrimination at both ends of the hierarchy.

The two lower panels are the point of the section. Panel (c) shows a pay gap that nowhere exceeds half of one per cent, and which is negative over most of the range: measured level by level, group g is paid slightly *more* than group 0. Panel (d) shows its members up to 0.74 per cent worse off. The two statements are in the same units and they disagree in sign, which is the whole difficulty with reading discrimination off pay. Panels (a) and (b) show what the near-coincidence of the scales conceals: a glass ceiling of the same magnitude as in Figure 3, and exclusion of group g from the bottom quarter of the job ladder.

A firm in this position may, without any impropriety, publish its pay scales, open every senior post on identical terms to all comers, and then observe, truthfully, that suitable applicants from the discriminated group have not come forward.

6.4 Varying the outside option

The three examples hold reservation utilities equal, for the reasons given at the start of Section 5. It is worth seeing what the assumption is doing, and Example 1 is the convenient vehicle, since it is the one whose senior pay gap is smallest and therefore most easily overturned. We hold δ_g and δ_0 exactly as in Section 6.1 and vary u_g alone, with

[24] The hump is the profile that puts most pressure on Assumption 2(ii), being the only one of the three requiring $\delta_g'' < 0$, near the peak. The three profiles used here are convex throughout and so satisfy it with room to spare; a hump satisfies it provided its curvature at the peak stays below $\psi_{\ell\ell} = 2\alpha = 0.8$.

$u_0 = 0$ throughout. Table 1 reports the result, measured at $\ell = 1$, a level above ℓ_m at which the two groups face identical discrimination.

Table 1: Example 1 with unequal outside options.

u_g	$\overline{\theta}_g$	lowest job	$\Delta\omega$	direct	induced	mean pay, $g/0$
-0.06	0.292	0.696	$+0.0327$	$+0.060$	-0.027	96.2%
-0.03	0.277	0.707	$+0.0160$	$+0.030$	-0.014	100.2%
0.00	0.262	0.719	-0.0006	0.000	-0.001	104.3%
$+0.03$	0.247	0.731	-0.0172	-0.030	$+0.013$	108.5%
$+0.06$	0.231	0.744	-0.0336	-0.060	$+0.026$	112.7%

Three observations. First, and most important for the argument of Section 5, the assignment schedules are identical in every row of the table, not approximately unchanged: the outside option does not appear in (8), so the curves in panel (b) of Figure 2 are the same curves throughout. Whatever else outside options do, they do not allocate anybody to a different job, and the apparent movement in the third column is entirely that of $\overline{\theta}_g$ along a fixed schedule. The slope of the pay gap is likewise unchanged, as Proposition 5(ii) requires; only its level moves, so that every qualitative statement of Section 5.2 survives unequal outside options and only the anchoring does not.

Second, the level moves through two channels of very unequal size, shown separately in the table. The direct channel is immediate: by (14) every wage in group g contains u_g additively, so a difference in outside options passes one-for-one into $\Delta\omega$. The induced channel is the extensive-margin effect of Proposition 5(iii): a group with a worse outside option is recruited more deeply, lengthening the tail of low-level workers and so raising the rents of everyone above them. The two pull in opposite directions and the direct channel dominates throughout, contributing $+0.060$ against the induced channel's -0.027 in the first row.

Third, the sign of the senior pay gap turns over within the table. At $u_g = 0$ the discriminated group is paid a shade more at every level above ℓ_m, for the reasons given in Section 6.1; at $u_g = -0.03$, an outside-option difference of some five per cent of average pay, it is paid less. That is the empirically plausible direction, and in those rows the discriminated group is paid less at every senior level; but the table shows the pattern to be produced by the outside option rather than by the spillover, which pulls the other way in every row: at $u_g = -0.03$ the outside-option channel contributes $+0.017$ against the spillover's -0.0006. Only the sum is observed, so identifying the spillover requires knowing something about δ directly.

6.5 What the reporting requirements would show

It is worth asking what an observer armed with the disclosures that the law actually requires would make of Figure 4. The EU Pay Transparency Directive 2023/970 goes further than most: employers must report not only the mean and median pay gap between groups, but mean pay by job level, and the proportion of each group in each quartile pay band. Table 2 gives those statistics for the three examples.

Table 2: What the disclosures would show.

	Example 1 (fades)	Example 2 (mounts)	Example 3 (both)
Glass ceiling	no	yes	yes
Mean pay of group g, per cent of group 0	104.3	96.7	101.0
Mean pay ratio within quartile pay band			
lowest	112.1	100.0	112.1
second	99.1	100.0	99.1
third	100.0	100.9	100.9
highest	100.0	93.5	93.5
Share of group g in each quartile pay band			
lowest	0.356	0.500	0.356
second	0.621	0.500	0.621
third	0.501	0.589	0.592
highest	0.501	0.411	0.410
Kolmogorov–Smirnov statistic	0.247	0.200	0.247
Workers per group for significance at 1%	87	133	87

Three things stand out. First, the headline number is uninformative and can point the wrong way: in Example 1 the discriminated group is paid 4.3 per cent *more* on average, so an employer reporting that figure would appear exemplary, and in Example 3 it is paid 1.0 per cent more with a glass ceiling.

Second, the quartile statistics are the informative ones, and informative about the right thing. In Examples 2 and 3 the share of group g in the top pay band falls to 0.41 and its mean pay within that band to 93.5 per cent. What these detect is not unequal pay but unequal *allocation*, exactly the object the theory says matters and exactly the object that "equal pay for equal work" does not govern; the Directive's requirement to report the composition of pay bands is on this reading better targeted than its requirement to report pay gaps.

Third, the disparities are statistically detectable in organisations of very moderate size. The Kolmogorov–Smirnov statistic comparing the two pay distributions is 0.247 in Examples 1 and 3, significant at the one per cent level in a firm with as few as 87 employees per group. The disparities in Figure 4 are invisible to inspection and comfortably visible to a test: what is required is not more data but the right comparison, between the distributions of pay and of jobs rather than between rates of pay at a level.

This returns us to the elephant in the room, job allocation. Equal pay has been argued for since Fawcett (1918) and Edgeworth (1922), hard fought for (Cohen, 2012; Meehan and Kahn, 2016; Engstrom, 2018), and is still imperfectly implemented even in

Scandinavia (Meyersson Milgrom et al., 2001). The examples of this section suggest that its achievement, though worth having, settles less than has been supposed: a firm may satisfy it to within half a per cent while allocating two groups to systematically different jobs, and it is in the allocation that the consequences of discrimination are recorded.

7 Conclusion

To paraphrase Justice Potter Stewart, discrimination is easy to recognise on encountering it and hard to formalise. Our purpose has not been to explain why it arises, which we take as given, but to work out what a firm that would rather not bear its cost does in response, and what that response does to workers the discrimination never touched.

Discrimination influences outcomes through spillovers of two very different ranges. Allocation responds to a spillover of infinitesimal range: who holds a post is settled by how conditions change in its immediate neighbourhood, not by how bad they are there, and a uniform worsening of a group's treatment reallocates nobody. Pay responds to a spillover whose range is every level below a worker's own, and one that runs strictly upward: rents accumulate from the foot of the hierarchy, so what is paid at the top is fixed by what is endured at the bottom, while nobody's pay is affected by the treatment of those above her. The asymmetry is not a modelling choice: workers are tempted to understate their ability and never to overstate it, so deviations point down and rents flow up. Discrimination confined to the lowest grades will therefore be invisible at the highest, nothing observable there differing between the groups, and consequential all the same.

Three implications seem worth drawing out.

The first concerns what should be measured. Compliance with equal pay for equal work is compatible with a substantial and entirely genuine difference in the treatment of two groups, as Example 3 shows; but the disparities show up in the *distribution* of jobs and of pay rather than in the rate for the job, and at moderate sample sizes. Regulators asking for the composition of pay bands, as EU Directive 2023/970 does, are asking for the more informative object; those asking for average pay gaps may be told, truthfully, that the discriminated group earns more.

The second concerns what should be fixed, and where. Because rents accumulate from below, a pay gap at the top can be closed only by acting on conditions further down; raising senior pay does not touch its cause. And the two disparities stand in opposite relations to what the firm knows: the glass ceiling would survive complete information, since the ablest worker of each group is already placed where such information would place her, while the pay gap, a difference in informational rents and nothing else, would vanish under it. No single remedy addresses both.

The third concerns the limits of a pay-based diagnosis. The theory determines the slope of the pay gap without qualification but its level only relative to an anchor, and a difference in reservation utilities of a few per cent suffices to reverse the sign of the senior gap with every discrimination function held fixed. Establishing that a spillover of the kind described here is operating therefore requires evidence about δ itself, of the sort the audit and field experimental literatures produce, rather than about wages.

Three limitations are worth naming. We do not model the firm's investment in reducing discrimination, beyond observing in Section 2 that the profit-maximising investment falls short of elimination, so that the δ_g we analyse is a residual. We do not model internal labour markets: the firm here hires from outside at every level, and the assignment of workers to levels is made once; combining the career dynamics that Gibbons and Waldman (1999) address with a level-specific disutility is a natural next step, the more inviting because today's allocation shapes the workplace tomorrow's entrants inherit, so that the spillovers documented here could compound. And we do not allow discrimination to depend on the worker's private type, which would let it act on the screening technology itself and confound the mechanism we wish to isolate.

Nor does competition dissolve any of this. In Becker (1957) competition erodes discrimination because the discriminating employer pays for a taste; here the employer harbours none, so there is nothing for rivals to compete away. Stiglitz (1973) asks how equally productive workers can receive different wages in equilibrium when hiring the cheaper one is profitable. Extend our framework to identical firms, each taking as given the discrimination its workers face, and there is no scope to poach, since a worker who moves meets the same environment at her new employer. Competition would press firms to reduce discrimination, but only to the point at which the marginal cost of doing so is met, and does nothing to remove the need to provide incentives for truthful revelation of type (Epstein and Peters, 1999). An earlier version of this paper (De Fraja and Sákovics, 2025) confirms this in an oligopsonistic setting in the spirit of Bhaskar et al. (2002). The conclusion sits comfortably with evidence finding competition to have only a modest effect on discrimination (Winter-Ebmer, 1995; Cooke et al., 2019; Siddique et al., 2023).

The message for an employer is finally rather simple, and it is not the one such an employer is likely to expect. Auditing pay at senior grades may well reveal a problem; what it cannot reveal is that senior grades did not create it. To close a gap at the top, mend the foundations.

Appendix: proofs

Throughout the appendix, a dot over a function of θ denotes its derivative with respect to θ.

Proof of Lemma 1. A worker of type θ who reports $\widetilde{\theta}$ is assigned to level $\ell_g(\widetilde{\theta})$ and paid $w_g(\widetilde{\theta})$, and so obtains

$$V(\widetilde{\theta},\theta) = w_g(\widetilde{\theta}) - \psi(\ell_g(\widetilde{\theta}),\theta) - \delta_g(\ell_g(\widetilde{\theta})). \tag{A1}$$

Incentive compatibility is the requirement that, for every θ, the report $\widetilde{\theta} = \theta$ maximise (A1), so that $U_g(\theta) = V(\theta,\theta) = \max_{\widetilde{\theta}} V(\widetilde{\theta},\theta)$.

Necessity. For monotonicity, write $Z(\theta) = w_g(\theta) - \delta_g(\ell_g(\theta))$. Incentive compatibility requires that type θ_1 prefer $\ell_g(\theta_1)$ to $\ell_g(\theta_2)$ and conversely, that is

$$\begin{aligned} Z(\theta_1) - \psi(\ell_g(\theta_1),\theta_1) &\geq Z(\theta_2) - \psi(\ell_g(\theta_2),\theta_1),\\ Z(\theta_2) - \psi(\ell_g(\theta_2),\theta_2) &\geq Z(\theta_1) - \psi(\ell_g(\theta_1),\theta_2). \end{aligned}$$

Adding and rearranging,

$$\int_{\theta_1}^{\theta_2} [\psi_\theta(\ell_g(\theta_1),x) - \psi_\theta(\ell_g(\theta_2),x)]\mathrm{d}x \geq 0,$$

that is,

$$\int_{\ell_g(\theta_2)}^{\ell_g(\theta_1)} \int_{\theta_1}^{\theta_2} \psi_{\ell\theta}(y,x)\mathrm{d}x\,\mathrm{d}y \geq 0. \tag{A2}$$

Since $\psi_{\ell\theta} > 0$ by Assumption 1(ii), (A2) holds for every $\theta_1 \neq \theta_2$ only if $\ell_g(\theta_1) \geq \ell_g(\theta_2)$ whenever $\theta_2 > \theta_1$: the schedule $\ell_g(\cdot)$ is non-increasing.

For (6), note first that the range of the monotone $\ell_g(\cdot)$ is bounded, so that $K = \sup \psi_\theta(y,x)$, the supremum taken over y in that range and $x \in [0,1]$, is finite, and, for every report, $|V(\widetilde{\theta},\theta) - V(\widetilde{\theta},\theta')| = |\int_{\theta'}^{\theta} \psi_\theta(\ell_g(\widetilde{\theta}),x)\mathrm{d}x| \leq K|\theta-\theta'|$: the functions $V(\widetilde{\theta},\cdot)$ share the Lipschitz constant K. Their pointwise maximum $U_g(\cdot)$ is then Lipschitz with the same constant, hence absolutely continuous: differentiable almost everywhere and equal to the integral of its derivative. Wherever $\dot{U}_g(\theta)$ exists, its value is pinned down by comparison with the payoff of the single report $\widetilde{\theta} = \theta$: for $h > 0$, $U_g(\theta+h) \geq V(\theta,\theta+h)$ and $U_g(\theta-h) \geq V(\theta,\theta-h)$, with equality at $h = 0$, so

$$\begin{aligned} \frac{V(\theta,\theta+h) - V(\theta,\theta)}{h} &\leq \frac{U_g(\theta+h) - U_g(\theta)}{h},\\ \frac{U_g(\theta) - U_g(\theta-h)}{h} &\leq \frac{V(\theta,\theta) - V(\theta,\theta-h)}{h}, \end{aligned}$$

and as $h \to 0$ both outer quotients converge to $-\psi_\theta(\ell_g(\theta),\theta)$, which is therefore the value of $\dot{U}_g(\theta)$. Integrating between θ and θ' gives (6).

Sufficiency. Let $\ell_g(\cdot)$ be non-increasing and (6) hold. For any type θ and any report

$\widetilde{\theta}$, by (A1) and (6),

$$U_g(\theta) - V(\widetilde{\theta}, \theta) = [U_g(\theta) - U_g(\widetilde{\theta})] + [V(\widetilde{\theta}, \widetilde{\theta}) - V(\widetilde{\theta}, \theta)]$$
$$= \int_{\widetilde{\theta}}^{\theta} [\psi_\theta(\ell_g(\widetilde{\theta}), x) - \psi_\theta(\ell_g(x), x)]\mathrm{d}x = \int_{\widetilde{\theta}}^{\theta} \int_{\ell_g(x)}^{\ell_g(\widetilde{\theta})} \psi_{\ell\theta}(y, x)\mathrm{d}y\,\mathrm{d}x.$$

If $\widetilde{\theta} < \theta$, then $\ell_g(x) \leq \ell_g(\widetilde{\theta})$ for every x between them, so the inner integral is non-negative by Assumption 1(ii) while the outer integral runs forwards; if $\widetilde{\theta} > \theta$, the inner integral is non-positive while the outer integral runs backwards. Either way $U_g(\theta) \geq V(\widetilde{\theta}, \theta)$: no misreport is profitable. □

Proof of Theorem 1. Constant returns to scale imply that (4) may be maximised separately for each group, so we suppress the subscript g throughout. From (5),

$$w(\theta) = U(\theta) + \psi(\ell(\theta), \theta) + \delta(\ell(\theta)), \tag{A3}$$

and substituting (A3) into (4) the firm's objective becomes

$$\int_0^{\overline{\theta}} [\ell(\theta) - \psi(\ell(\theta), \theta) - U(\theta) - \delta(\ell(\theta))]\varphi(\theta)\mathrm{d}\theta.$$

By Lemma 1, the constraints are that $\ell(\cdot)$ be non-increasing and that (6) hold. We ignore the former, which Corollary 1 verifies at the solution, and impose (7) at $\overline{\theta}$ alone, with equality: U is strictly decreasing by (6), so participation at $\overline{\theta}$ implies participation below it, and profit falls one for one with $U(\overline{\theta})$. By (6) with $\theta' = \overline{\theta}$ and $U(\overline{\theta}) = u_g$, and exchanging the order of integration,

$$\int_0^{\overline{\theta}} U(\theta)\varphi(\theta)\mathrm{d}\theta = u_g\Phi(\overline{\theta}) + \int_0^{\overline{\theta}} \psi_\theta(\ell(x), x)\Phi(x)\mathrm{d}x,$$

since disutility borne at x is conceded, as rent, to the $\Phi(x)$ abler types. With $h = \Phi/\varphi$ and (3), the firm's problem is therefore to choose $\ell(\cdot)$ and $\overline{\theta} \in (0, 1]$ to maximise

$$J(\ell(\cdot), \overline{\theta}) = \int_0^{\overline{\theta}} S(\ell(\theta), \theta)\varphi(\theta)\mathrm{d}\theta - u_g\Phi(\overline{\theta}). \tag{A4}$$

Profit is virtual surplus, and (A4) has no interaction across types: it is maximised pointwise. The first-order condition $S_\ell(\ell, \theta) = 0$ is (8), and identifies a global maximum, ruling out with it any gain from randomising, because

$$\frac{\partial^2 S(\ell, \theta)}{\partial \ell^2} = -[\psi_{\ell\ell} + h(\theta)\psi_{\ell\ell\theta} + \delta''(\ell)] = -\eta(\theta) < 0 \tag{A5}$$

by Assumption 2(ii). Differentiating (A4) with respect to $\overline{\theta}$ and dividing by $\varphi(\overline{\theta}) > 0$ gives, at an interior optimum,

$$0 = \ell^*(\overline{\theta}) - u_g - \psi(\ell^*(\overline{\theta}), \overline{\theta}) - \delta(\ell^*(\overline{\theta})) - h(\overline{\theta})\psi_\theta(\ell^*(\overline{\theta}), \overline{\theta}), \tag{A6}$$

which is (9). To see that its solution is unique, write $\Gamma(\theta) = \max_\ell S(\ell, \theta) - u_g$, so that

$\partial J/\partial\overline{\theta} = \varphi\Gamma$. By the envelope theorem,

$$\Gamma'(\theta) = S_\theta(\ell_g^*(\theta), \theta) = -(1 + h'(\theta))\psi_\theta - h(\theta)\psi_{\theta\theta} < 0$$

by Assumption 1(i), the monotone hazard rate assumption, and Assumption 1(iii).[25] Thus Γ is strictly decreasing, and $\Gamma(0) > 0$ is the positive-surplus condition imposed before the statement of the theorem, without which the firm employs nobody from group g. If $\Gamma(1) \leq 0$, then Γ has a unique zero in $(0, 1]$, which is $\overline{\theta}_g$; if instead $\Gamma(1) > 0$, then $\partial J/\partial\overline{\theta} > 0$ throughout, the constraint $\overline{\theta} \leq 1$ binds, and $\overline{\theta}_g = 1$: the whole group is employed. □

Proof of Corollary 1. Totally differentiating (8) with respect to θ gives

$$[\psi_{\ell\ell} + h(\theta)\psi_{\ell\ell\theta} + \delta_g'']\mathrm{d}\ell + [\psi_{\ell\theta} + h'(\theta)\psi_{\ell\theta} + h(\theta)\psi_{\ell\theta\theta}]\mathrm{d}\theta = 0,$$

which is (10) with H and η_g as defined in (11) and (12). That $H(\theta) > 0$ follows from $\psi_{\ell\theta} > 0$, $h' > 0$ and $\psi_{\ell\theta\theta} \geq 0$; that $\eta_g(\theta) > 0$ is Assumption 2(ii). Thus $\mathrm{d}\ell_g^*/\mathrm{d}\theta < 0$, so $\ell_g^*(\cdot)$ is strictly decreasing and ignoring monotonicity in the proof of Theorem 1 was legitimate. Continuity of ℓ_g^* follows from (8), whose right-hand side is continuous in (ℓ, θ) and strictly increasing in ℓ by Assumption 2(ii); indeed, that side being continuously differentiable with ℓ-derivative $\eta_g > 0$, the implicit function theorem makes $\ell_g^*(\cdot)$ continuously differentiable.

For the wage, (6) with $\theta' = \overline{\theta}_g$ and (7) give

$$U_g(\theta) = u_g + \int_\theta^{\overline{\theta}_g} \psi_\theta(\ell_g^*(x), x)\mathrm{d}x;$$

the integrand is continuous, $\ell_g^*(\cdot)$ being so, and therefore $U_g(\cdot)$ is continuously differentiable, with

$$\dot{U}_g(\theta) = -\psi_\theta(\ell_g^*(\theta), \theta). \tag{A7}$$

By (5), $w_g^*(\theta) = U_g(\theta) + \psi(\ell_g^*(\theta), \theta) + \delta_g(\ell_g^*(\theta))$ is then continuously differentiable too, and differentiating,

$$\dot{U}_g(\theta) = \dot{w}_g^*(\theta) - [\psi_\ell(\ell_g^*(\theta), \theta) + \delta_g'(\ell_g^*(\theta))]\dot{\ell}_g^*(\theta) - \psi_\theta(\ell_g^*(\theta), \theta), \tag{A8}$$

which combined with (A7) gives (13). Uniqueness of the wage is immediate: (6) determines $U_g(\cdot)$ up to a constant, (7) fixes the constant, and (5) then determines $w_g(\cdot)$. □

Proof of Corollary 2. Setting $\theta' = \overline{\theta}_g$ in (6),

$$U_g(\theta) = U_g(\overline{\theta}_g) + \int_\theta^{\overline{\theta}_g} \psi_\theta(\ell_g^*(x), x)\mathrm{d}x.$$

Substituting $U_g(\overline{\theta}_g) = u_g$ from (7) and then substituting the result into (5) gives (14). □

[25] Only the weaker requirement $(1 + h'(\theta))\psi_\theta + h(\theta)\psi_{\theta\theta} > 0$ is needed, which permits ψ to be concave in θ provided it is not too strongly so.

Proof of Corollary 3. Differentiating (15) and using $\Theta_g^{*\prime}(\ell) = 1/\dot{\ell}_g^*(\Theta_g^*(\ell))$, which exists and is finite by Corollary 1,

$$\frac{\mathrm{d}\omega_g^*(\ell)}{\mathrm{d}\ell} = \dot{w}_g^*(\Theta_g^*(\ell))\Theta_g^{*\prime}(\ell) = \frac{\dot{\ell}_g^*[\psi_\ell + \delta_g']}{\dot{\ell}_g^*} = \psi_\ell(\ell, \Theta_g^*(\ell)) + \delta_g'(\ell),$$

where the second equality uses (13). The sign follows from Assumption 2(i). □

Proof of Proposition 1. Fix $\ell \in \mathcal{L}$ and define $z_\ell(\theta) = \psi_\ell(\ell,\theta) + h(\theta)\psi_{\ell\theta}(\ell,\theta)$. By (8), the types found at level ℓ in the two groups satisfy

$$z_\ell(\Theta_0^*(\ell)) + \delta_0'(\ell) = 1 = z_\ell(\Theta_g^*(\ell)) + \delta_g'(\ell),$$

so that $z_\ell(\Theta_0^*(\ell)) - z_\ell(\Theta_g^*(\ell)) = \delta_g'(\ell) - \delta_0'(\ell)$. By the mean value theorem there is $\widetilde{\theta}$ between $\Theta_0^*(\ell)$ and $\Theta_g^*(\ell)$ with

$$z_\ell'(\widetilde{\theta})[\Theta_0^*(\ell) - \Theta_g^*(\ell)] = \delta_g'(\ell) - \delta_0'(\ell).$$

Now $z_\ell'(\theta) = \psi_{\ell\theta} + h'(\theta)\psi_{\ell\theta} + h(\theta)\psi_{\ell\theta\theta} = H(\theta) > 0$ by (11), which gives (16). For (ii), suppose $\ell_g^*(\theta) > \ell_0^*(\theta)$ and take any ℓ strictly between them. By (A5), $S_g(\cdot,\theta)$ is strictly concave in the job level with maximiser $\ell_g^*(\theta) > \ell$, so $\partial S_g(\ell,\theta)/\partial\ell > 0$; likewise $S_0(\cdot,\theta)$ is strictly concave with maximiser $\ell_0^*(\theta) < \ell$, so $\partial S_0(\ell,\theta)/\partial\ell < 0$. By (3), $\partial S_g/\partial\ell - \partial S_0/\partial\ell = \delta_0'(\ell) - \delta_g'(\ell)$, so subtracting the two inequalities gives $\delta_g'(\ell) < \delta_0'(\ell)$ throughout the interval: the constant sign claimed, opposite to that of $\ell_g^*(\theta) - \ell_0^*(\theta)$. The case $\ell_g^*(\theta) < \ell_0^*(\theta)$ is symmetric, and if the two coincide the interval is empty.

The comparison is more robust than the assumptions under which it is stated. From (3), $S_g(\ell,\theta) - S_0(\ell,\theta) = -[\delta_g(\ell) - \delta_0(\ell)]$, which does not involve θ. If $\delta_g - \delta_0$ is increasing on an interval, the difference of the two virtual surpluses is decreasing in ℓ there, so S_g is submodular relative to S_0 and $\arg\max_\ell S_g \leq \arg\max_\ell S_0$ in the strong set order by Topkis's theorem. Neither differentiability nor concavity is needed, so the conclusion survives the failure of Assumption 2(ii), with the pointwise comparison replaced by the set-order one; the same holds of Proposition 2. □

Proof of Proposition 2. (i) Setting $\theta = 0$ in (8) and using $h(0) = \Phi(0)/\varphi(0) = 0$ gives (17), which is the first-order condition for the maximisation of $\ell - \psi(\ell,0) - \delta_g(\ell)$; the second-order condition is $\psi_{\ell\ell} + \delta_g'' > 0$, which is Assumption 2(ii) at $\theta = 0$.

(ii) For the equality claim, note that by (17) the two tops satisfy first-order conditions differing only in δ_g' and δ_0'. If $\ell_g^*(0) = \ell_0^*(0)$ the two conditions are evaluated at the same level, so δ_g' and δ_0' agree there. Conversely, if they agree at $\ell_0^*(0)$ then that level satisfies group g's first-order condition, and since $\ell - \psi(\ell,0) - \delta_g(\ell)$ is strictly concave by Assumption 2(ii) it is group g's unique maximiser, so the tops coincide. When the tops differ, the sign claim is Proposition 1 applied at $\theta = 0$.

The multi-group statement follows: if some g had $\delta_g'(\ell_M) < \delta_{g^*}'(\ell_M)$, where $\ell_M = \ell_{g^*}^*(0)$ is the highest job filled, then by (17) the derivative of $\ell - \psi(\ell,0) - \delta_g(\ell)$ would be strictly positive at ℓ_M, and that function being strictly concave its maximiser $\ell_g^*(0)$ would exceed ℓ_M, a contradiction; so $g^* \in \arg\min_{g\in G} \delta_g'(\ell_M)$. $\square$

Proof of Proposition 3. By Corollary 3, $\mathrm{d}\omega_g^*/\mathrm{d}\ell = \psi_\ell(\ell, \Theta_g^*(\ell)) + \delta_g'(\ell)$. Evaluating (8) at $\theta = \Theta_g^*(\ell)$, so that $\ell_g^*(\theta) = \ell$, gives

$$\psi_\ell(\ell, \Theta_g^*(\ell)) + \delta_g'(\ell) = 1 - h(\Theta_g^*(\ell))\psi_{\ell\theta}(\ell, \Theta_g^*(\ell)) = 1 - m(\ell, \Theta_g^*(\ell)),$$

which is (18); the bound follows from $m \geq 0$, since $h \geq 0$ and $\psi_{\ell\theta} > 0$. Differencing (18) across the two groups gives (19). For (20), note that

$$\frac{\partial m(\ell,\theta)}{\partial\theta} = h'(\theta)\psi_{\ell\theta}(\ell,\theta) + h(\theta)\psi_{\ell\theta\theta}(\ell,\theta) > 0,$$

by $h' > 0$, $\psi_{\ell\theta} > 0$, $h \geq 0$ and $\psi_{\ell\theta\theta} \geq 0$. Thus $\operatorname{sign}\Delta\omega'(\ell) = \operatorname{sign}[\Theta_g^*(\ell) - \Theta_0^*(\ell)]$, the first equality of (20); the second follows from Proposition 1. $\square$

Proof of Corollary 4. By Proposition 1, $\delta_g' = \delta_0'$ on I implies $\Theta_g^* = \Theta_0^*$ there, whence $\Delta\omega' = 0$ on I by (19). $\square$

Proof of Corollary 5. Evaluating (8) at $\theta = \Theta_g^*(\ell)$, as in the proof of Proposition 3, gives $\psi_\ell(\ell, \Theta_g^*(\ell)) + m(\ell, \Theta_g^*(\ell)) = 1 - \delta_g'(\ell)$ for each group; subtracting the two,

$$[\psi_\ell + m](\ell, \Theta_g^*(\ell)) - [\psi_\ell + m](\ell, \Theta_0^*(\ell)) = \delta_0'(\ell) - \delta_g'(\ell). \tag{A9}$$

Both $\theta \mapsto \psi_\ell(\ell,\theta)$ and $\theta \mapsto m(\ell,\theta)$ are strictly increasing, the first because $\psi_{\ell\theta} > 0$, the second because $m_\theta = h'\psi_{\ell\theta} + h\psi_{\ell\theta\theta} > 0$, as in the proof of Proposition 3. If $\delta_g'(\ell) = \delta_0'(\ell)$ the two types coincide and both sides of (21) are zero; otherwise, by the Cauchy mean value theorem there is $\widetilde{\theta}$ strictly between $\Theta_0^*(\ell)$ and $\Theta_g^*(\ell)$ with

$$m(\ell, \Theta_g^*(\ell)) - m(\ell, \Theta_0^*(\ell)) = \frac{m_\theta(\ell, \widetilde{\theta})}{\psi_{\ell\theta}(\ell, \widetilde{\theta}) + m_\theta(\ell, \widetilde{\theta})} \left\{[\psi_\ell + m](\ell, \Theta_g^*(\ell)) - [\psi_\ell + m](\ell, \Theta_0^*(\ell))\right\},$$

and (21) follows from (19) and (A9); that $\kappa \in (0,1)$ follows from the strict positivity of $\psi_{\ell\theta}$ and m_θ. In the parametric case $\psi_{\ell\theta} = 2\alpha$, $\psi_{\ell\theta\theta} = 0$ and $h(\theta) = \theta$, so $m_\theta = 2\alpha$ and $\kappa(\ell) = \frac{1}{2}$ for every ℓ and $\widetilde{\theta}$; integrating $\Delta\omega' = -\frac{1}{2}(\delta_g' - \delta_0')$ over $\mathcal{L}$ gives the constancy of (22). $\square$

Proof of Proposition 4. (i) Since $\delta_g = \delta_0$ above $\widehat{\ell}$, we have $\delta_g' = \delta_0'$ there, so $\Theta_g^* = \Theta_0^*$ above $\widehat{\ell}$ by Proposition 1, and $\ell_g^*(0) = \ell_0^*(0)$ by Proposition 2(ii), since δ_g' and δ_0' agree at $\ell_0^*(0) > \widehat{\ell}$.

(ii) Constancy is Corollary 4. For the value, take any $\ell \geq \widehat{\ell}$ and write $\theta = \Theta_0^*(\ell) = \Theta_g^*(\ell)$. Since $\delta_g(\ell) = \delta_0(\ell)$ and the two groups place the same type at ℓ, the first three terms of (14) coincide, so by (14)

$$\Delta\omega = \int_\theta^{\bar{\theta}_0} \psi_\theta(\ell_0^*(x), x)\mathrm{d}x - \int_\theta^{\bar{\theta}_g} \psi_\theta(\ell_g^*(x), x)\mathrm{d}x.$$

By (i), $\ell_0^*(x) = \ell_g^*(x)$ for every $x \leq \widehat{\theta}$, so the parts of the two integrals over $[\theta, \widehat{\theta}]$ cancel, leaving (23). That the resulting expression is independent of ℓ confirms the constancy directly. $\square$

Proof of Proposition 5. (i) u_g does not appear in (8). (ii) By (19), $\Delta\omega'$ depends on the primitives only through Θ_g^* and Θ_0^*, which by (i) do not depend on reservation utilities. (iii) In the notation of the proof of Theorem 1, an increase in u_g shifts Γ down at every θ, and Γ is strictly decreasing, so its zero falls. If in addition $\delta_g = \delta_0$ near the bottom then Γ differs between the two groups only by the constant $u_0 - u_g$, and the stated comparison follows. $\square$

References

Mladen Adamovic and Andreas Leibbrandt. Is there a glass ceiling for ethnic minorities to enter leadership positions? evidence from a field experiment with over 12,000 job applications. *The Leadership Quarterly*, 34(2):101655, 2023.

Renée B. Adams and Patricia Funk. Beyond the glass ceiling: does gender matter? *Management Science*, 58(2):219–235, 2012.

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

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

Joseph G. Altonji and Charles R. Pierret. Employer learning and statistical discrimination. *The Quarterly Journal of Economics*, 116(1):313–350, 2001.

Mark Armstrong. Multiproduct nonlinear pricing. *Econometrica*, 64(1):51–75, 1996.

Mark Bagnoli and Ted Bergstrom. Log-concave probability and its applications. *Economic Theory*, 26:445–469, 2005.

George Baker, Michael Gibbs, and Bengt Holmstrom. The internal economics of the firm: evidence from personnel data. *The Quarterly Journal of Economics*, 109(4):881–919, 1994.

Jose Maria Barrero, Nicholas Bloom, and Steven J. Davis. Why working from home will stick. Working Paper 28731, National Bureau of Economic Research, 2021.

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

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

Alan Benson, Danielle Li, and Kelly Shue. 'potential' and the gender promotion gap. *American Economic Review*, 116(2):375–417, 2026.

Marianne Bertrand. Coase lecture–the glass ceiling. *Economica*, 85(338):205–231, 2018.

Marianne Bertrand and Esther Duflo. Field experiments on discrimination. In Esther Duflo and Abhijit Banerjee, editors, *Handbook of Field Experiments*, pages 309–393. Elsevier, 2017.

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

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

J. Aislinn Bohren, Peter Hull, and Alex Imas. Systemic discrimination: theory and measurement. *The Quarterly Journal of Economics*, 140(3):1743–1799, 2025.

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

Daniele Checchi, Daniel Kreisman, and Cecilia García-Peñalosa. The evolution of hours worked and the gender wage gap: Theory and evidence from four countries. Discussion Paper 18265, IZA – Institute of Labor Economics, 2025.

Stephen Coate and Glenn C. Loury. Will affirmative-action policies eliminate negative stereotypes? *American Economic Review*, 83(5):1220–1240, 1993.

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

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

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

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

Lilia M. Cortina, Dana Kabat-Farr, Emily A. Leskinen, Marisela Huerta, and Vicki J. Magley.

Selective incivility as modern discrimination in organizations: evidence and impact. *Journal of Management*, 39(6):1579–1605, 2013.

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

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

Gianni De Fraja and József Sákovics. Discrimination spillovers, glass ceilings and pay gaps. Discussion Paper 11, Nottingham Interdisciplinary Centre for Economic and Political Research, 2025.

Gianni De Fraja, Giovanni Facchini, and John Gathergood. Academic salaries and public evaluation of university research: evidence from the uk research excellence framework. *Economic Policy*, 34(99):523–583, 2019.

Gianni De Fraja, Jesse Matheson, Paul Mizen, James Rockey, Shivani Taneja, and Gregory Thwaites. Remote work and compensation inequality. Discussion Paper 20055, CEPR, 2025.

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

Yuhao Du, Jessica Nordell, and Kenneth Joseph. Insidious nonetheless: how small effects and hierarchical norms create and maintain gender disparities in organizations. *Socius*, 8, 2022.

Susann Dunger. It's the hierarchy, stupid: varying perceptions of organizational culture between demographic groups. *International Studies of Management & Organization*, 55(1):83–111, 2025.

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

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

David Freeman Engstrom. "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, 2018.

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

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

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

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

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

Nickolas Gagnon, Kristof Bosmans, and Arno Riedl. The effect of gender discrimination on labor supply. *Journal of Political Economy*, 133(3):1047–1081, 2025.

Robert Gibbons and Lawrence F. Katz. Layoffs and lemons. *Journal of Labor Economics*, 9(4): 351–380, 1991.

Robert Gibbons and Michael Waldman. A theory of wage and promotion dynamics inside firms. *The Quarterly Journal of Economics*, 114(4):1321–1358, 1999.

Robert Gibbons and Michael Waldman. Enriching a theory of wage and promotion dynamics inside firms. *Journal of Labor Economics*, 24(1):59–107, 2006.

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

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

Claudia Goldin. Nobel lecture: an evolving economic force. *American Economic Review*, 114 (6):1515–1539, 2024.

Paul Grout, In-Uck Park, and Silvia Sonderegger. An economic theory of glass ceiling. Discussion Paper 09/227, CMPO, Bristol, 2009.

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

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

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

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

Benjamin E. Hermalin. Economics and corporate culture. In Cary Cooper, P. Christopher Earley, and Susan Cartwright, editors, *The International Handbook of Organizational Culture and Climate*. John Wiley & Sons, Chichester, 2001.

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

Bruno Jullien. Participation constraints in adverse selection models. *Journal of Economic Theory*, 93(1):1–47, 2000.

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

Lisa B. Kahn. Asymmetric information between employers. *American Economic Journal: Applied Economics*, 5(4):165–205, 2013.

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

Julie H. Kern and Alicia A. Grandey. Customer incivility as a social stressor: the role of race and racial identity for service employees. *Journal of Occupational Health Psychology*, 14(1): 46–57, 2009.

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

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

Edward P. Lazear and Sherwin Rosen. Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, 89(5):841–864, 1981.

Daniel Leonard and Ngo Van Long. *Optimal Control Theory and Static Optimization in Economics*. Cambridge University Press, Cambridge, 1992.

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

R. Preston McAfee and John McMillan. Multidimensional incentive compatibility and mechanism design. *Journal of Economic Theory*, 46(2):335–354, 1988.

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

Eva M. Meyersson Milgrom, Trond Petersen, and Vemund Snartland. Equal pay for equal work? evidence from sweden and a comparison with norway and the us. *Scandinavian Journal of Economics*, 103(4):559–583, 2001.

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

Roger B. Myerson. Incentive compatibility and the bargaining problem. *Econometrica*, 47(1): 61–73, 1979.

Markus Nagler, Johannes Rincke, and Erwin Winkler. How much do workers actually value working from home? Working Paper 10073, CESifo, Munich, 2022.

Paula Onuchic. Recent contributions to theories of discrimination. *arXiv preprint 2205.05994*, 2025.

Bartolomé Pascual-Fuster, Ryan Federo, Rafel Crespí-Cladera, and Patricia Gabaldón. The second glass ceiling: The dark side of women recategorization in corporate boards. *Gender, Work & Organization*, 32(3):1018–1043, 2025.

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

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

Elena S. Pikulina and Daniel Ferreira. Subtle discrimination. *Journal of Finance*, 2026. forthcoming.

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

Paul C. Price, Brandi Holcomb, and Makayla B. Payne. Gender differences in impostor phenomenon: a meta-analytic review. *Current Research in Behavioral Sciences*, 7:100155, 2024.

Jean-Charles Rochet and Philippe Choné. Ironing, sweeping, and multidimensional screening. *Econometrica*, 66(4):783–826, 1998.

Sherwin Rosen. The theory of equalizing differences. In Orley C. Ashenfelter and Richard Layard, editors, *Handbook of Labor Economics*, volume 1, pages 641–692. Elsevier, 1986.

Michael Schaerer, Christilene Du Plessis, My Hoang Bao Nguyen, et al. 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, 2023.

Uta Schönberg. Testing for asymmetric employer learning. *Journal of Labor Economics*, 25(4): 651–691, 2007.

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

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

Joseph E. Stiglitz. Approaches to the economics of discrimination. *American Economic Review*, 63(2):287–295, 1973.

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

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

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