

Bonus Culture:
Competitive Pay, Screening, and Multitasking

Roland Bénabou¹ and Jean Tirole^{2 3}

September 7, 2012

Preliminary and Incomplete

Comments Welcome

¹Princeton University and IAST

²Toulouse School of Economics and IAST

³We thank Thomas Mariotti, David Sraer, Glen Weyl and participants at conferences and seminars at TSE-IAS, CEPR, the NBER Summer Institute and Princeton University for helpful comments. Both authors gratefully acknowledge financial support from the ERC programme grant FP7/2007-2013 No. 249429, “Cognition and Decision-Making: Laws, Norms and Contracts”.

1 Introduction

Recent years have seen a literal explosion of pay, both in levels and in differentials, at the top echelons of many occupations. Large bonuses and salaries are needed, it is typically said, to retain “talent” and “top performers” in finance, corporations, medicine, academia, as well as to incentivize them to perform to the best of their high abilities. Paradoxically, this trend has been accompanied by mounting revelations of poor actual performance, severe moral hazard and even outright fraud in those same sectors. Oftentimes these behaviors impose negative spillovers on the rest of society (e.g., bank bailouts), but even when not, the firms involved themselves ultimately suffer: large trading losses, declines in stock-value, loss of consumer goodwill, regulatory fines and legal liabilities, or even bankruptcy.

This paper proposes a resolution of the puzzle, by showing how competition for the most productive workers can interact with the incentive structure inside firms to undermine work ethics –the extent to which agents “do the right thing” beyond what their material self-interest commands. More generally, the underlying idea is that highly competitive labor markets make it difficult for employers to strike the proper balance between the benefits and costs of high-powered incentives.¹ The result is a “bonus culture” that takes over the workplace, generating distorted decisions and significant efficiency losses, particularly in the long run. To make this point we develop a model that combines multitasking, screening and imperfect competition, thus making a methodological contribution in the process.

Inside each firm, agents perform both a task that is easily measured (sales, output, trading profits, billable medical procedures) and one that is not and therefore involves an element of public-goods provision (intangible investments affecting long-run value, financial or legal risk-taking, cooperation among individuals or divisions). Agents potentially differ in their productivity for the rewardable task and in their intrinsic willingness to provide the unrewarded one –their work ethic. When types are observable, the standard result applies: principals set relatively low-powered incentives that optimally balance worker’s effort allocation; competition then only affects the size of fixed compensation. Things change fundamentally when skill differences are unobservable, leading firms to offer contracts designed to screen different types of workers. A single principal (monopsonist, collusive industry) sets the power of incentives even lower than the social optimum, so as to extract rents from the more productive agents. Labor-market competition, however, introduces a new role for performance pay: because it is differentially attractive to more productive workers, it also serves as a device which firms use to attract (or retain) these types. Focusing first on the limiting case of perfect competition, we show that the degree of incentivization is always above the social optimum, and we identify a simple condition under which the resulting distortion exceeds that occurring under monopsony. Competitive bidding for talent is thus destructive of work ethics, and ultimately welfare-reducing.

¹In a companion paper (Bénabou and Tirole (2012)), we explore a second mechanism for the destructive escalation of pay (now even with a single task), based on the two-dimensional self-selection and associated reputations of workers with heterogenous preferences over ethics and money.

We then develop a Hotelling-like variant of competitive screening to analyze the equilibrium contracts under arbitrary degrees of imperfect competition. As mobility costs (or horizontal differentiation) decline, the monopsonistic underincentivization of low-skill agents gradually decreases, then at some point gives way to a growing overincentivization of high-skill ones. Aggregate welfare is thus hill-shaped with respect to competition, while comprehensive measures of inequality (gaps in utility or total earnings) tend to rise monotonically. This leads us to analyze different policies, such as bonus caps or taxes on total compensation, that can potentially improve efficiency and restore balance in agents' incentives and focus. The extent to which this is achievable depends on how well the government or regulator is able to distinguish the incentive versus fixed parts of compensation packages, as well as on the distortions that may arise as firms try to blur that line or resort to even less efficient screening devices.

Finally, we contrast our main analysis of competition for talent with the polar case where agents have the same productivity in the measurable task but differ in their ethical motivation for unmeasurable one. In this case, competition is shown to be either beneficial (reducing the overincentivization which a monopsonist uses to extract rent, but never causing underincentivization), or neutral—as occurs in a variant of the model where ethical motivation generates positive spillovers inside the firm instead of private benefits for the agent.

Related evidence. Although bankers' bonuses and CEO pay packages attract the most attention, the parallel rise in incentive pay and earnings inequality is a much broader phenomenon, as established by Lemieux et al. (2009). Between the late 1970's and the 1990's, the fraction of jobs explicitly paid based on performance rose from 38% to 45%. Further compounding the direct impact on inequality is the fact that the returns to skills, both observable (education, experience, job tenure) and unobservable, are much higher in such jobs. This last finding also suggests that different compensation structures may play an important sorting role.² Lemieux et al. calculate that the interaction of structural change and differential returns account for 21% of the growth in the variance of male log-wages over the period, and for essentially 100% (or even more) above the 80th percentile.

The source of escalation in incentive pay in our model is increased competition for the best workers, and this also fits well with the evidence on managerial compensation in advanced countries. In a long-term study (1936-2003) of the market for top US executives, Frydman (2007) documents a major shift, starting in the 1970's and sharply accelerating since the late 1980's, from firm-specific skills to more general managerial ones—e.g., from engineering degrees to MBA's. In addition, there has been a concomitant rise in the diversity of sectorial experiences acquired over the course of a typical career. Frydman argues that these decreases in mobility costs have intensified competition for managerial skills and shows that, consistent with view, executives with higher general (multipurpose) human capital received higher compensation and were also the most likely

²Consistent with this view and with our modelling premise that performance incentives affect not only moral hazard (e.g. Bandiera et al. 2007, Shearer 2004) but also selection, Lazear's (2000) study of Safelite Glass Company found that half of the 44% productivity increase reaped when the company replaced the hourly wage system by a piece rate was due to in- and out-selection effects.

to switch companies. Using panel data on the 500 largest firms in Germany over 1977-2009, Fabbri and Marin (2011) show that domestic and (to a lesser extent) global competition for managers has contributed significantly to the rise of executive pay in that country, particularly in the banking sector.

Our theory is based on competition not simply bidding up the level of compensation at the top, but also significantly altering its structure toward high-powered incentives, with a resulting shift in the mix of tasks performed toward more easily quantifiable and short-term-oriented ones. This seems to be precisely what occurred on Wall Street as market-based compensation spreading from the emerging alternative-assets industry to the rest of the financial world:

“Talent quickly migrated from investment banks to hedge funds and private equity. Investment banks, accustomed to attracting the most-talented executives in the world and paying them handsomely, found themselves losing their best people (and their best MBA recruits) to higher-paid and, for many, more interesting jobs... Observing the remarkable compensation in alternative assets, sensing a significant business opportunity, and having to fight for talent with this emergent industry led banks to venture into proprietary activities in unprecedented ways. From 1998 to 2006 principal and proprietary trading at major investment banks grew from below 20% of revenues to 45%. In a 2006 Investment Dealers’ Digest article... one former Morgan Stanley executive said... that extravagant hedge fund compensation –widely envied on Wall Street, according to many bankers– was putting upward pressure on investment banking pay, and that some prop desks were even beginning to give traders "carry." Banks bought hedge funds and private equity funds and launched their own funds, creating new levels of risk within systemically important institutions and new conflicts of interest. By 2007 the transformation of Wall Street was complete. Faced with fierce new rivals for business and talent, investment banks turned into risk takers that compensated their best and brightest with contracts embodying the essence of financial-markets-based compensation.” (Desai 2012, *The Incentive Bubble*).

Similar transformations have occurred in the medical world with the rise of for-profit hospital chains: Gawande (2009) documents the explosion of compensation driven by the overuse of revenue-generating tests and surgeries, with parallel declines in preventive care and coordination on cases between specialists, increases in costs and worse patient outcomes.

Related literature. Our paper relates to and extends several lines of work. The technically closest one is that on competitive screening with exclusive contracts, initiated by Rothschild and Stiglitz (1976). Studying a perfectly competitive (free entry) insurance market with two types of agents, they characterized the separating equilibrium when it exists and identified a condition for this to be the case. Croker and Snow (1985) studied the whole Pareto frontier for the two types and showed how it ranges from suboptimal insurance for the safer type (as in the Rothschild-Stiglitz equilibrium) to over-insurance for the risky type. The subsequent literature (Armstrong

and Vickers 2001, Rochet and Stole 2002) studied private-value models where, in contrast with the present work, principals do not directly care about the agent’s type but are purely concerned with rent extraction. Our paper extends the literature in two directions. First, it examines competitive screening in a multitask environment, and second, it is the first to derive the equilibrium for the whole range of competition intensities between the polar cases of monopoly and perfect competition on which most previous work has focused. This allows us to study how individual payoffs, income inequality and aggregate welfare vary with market frictions, characterizing in particular the socially optimal degree of competitiveness.

From the multitasking literature we borrow and build on the idea that incentivizing easily measurable tasks may jeopardize the provision of effort on less measurable ones (e.g. Holmström and Milgrom 1991, Itoh 1991, Baker et al. 1994, Dewatripont et al. 1999, Fehr and Schmidt 2004). Somewhat surprisingly, the impact of competition on the multitasking problem has not been previously investigated –a fortiori not in combination with adverse selection, which is in fact needed to generate novel results. As in earlier work, employers choose compensation structures aiming to balance incentives, but the desire to extract rents or the need to select the best employees lead them to offer socially distorted compensation schemes. In relatively competitive labor markets, in particular, a firm raising its performance-based pay exerts a negative social externality: it fails to internalize the fact that competitors, in order to retain their own “talent”, will also have to distort their incentive structure and effort allocation, thereby reducing the total surplus generated by their workforce.

While our paper is not specifically about executive pay, this is an important application of the model. The literature on managerial compensation is usually seen as organized along two contrasting lines.³ On one hand is the view that high executive rewards reflect a high demand for rare skills (Rosen 1981) and the efficient workings of a competitive market allocating talent to where it is most productive, for instance to manage larger firms (Gabaix and Laibson 2008). Rising pay at the top is then simply the appropriate price response to market trends favoring the best workers: skilled-biased technical change, improvements in monitoring, growth in the size of firms, entry or decreases in mobility costs. On the other side is the view that the level and structure of managerial compensation reflect instead significant market failures. For instance, indolent or captured boards may grant top executives pay packages far in excess of their marginal product, (Bertrand and Mullainathan 2001, Bebchuk and Fried 2004). Alternatively, managers are given incentive schemes that do maximize profits but impose significant negative externalities on the rest of society by inducing excessive short-termism and risk-taking at the expense of consumers, depositors or taxpayers (public bailouts and environmental cleanups, tax arbitrage, etc.) (e.g. Bolton et al. 2006, Besley and Ghatak 2011). In particular, private returns in the finance industry are often argued to exceed social returns (Baumol 1990, Philippon and Reshef 2012). Our paper takes on board the first view’s premise that pay levels and differentials largely reflect market returns to both talent and measured performance, magnified in recent decades by technical change

³See, e.g., Frydman and Jenter (2010) for a recent survey.

and increased mobility. At the same time, and closer in that to the second view, we show that this very same escalation of performance-based pay can be the source of severe distortions and long-run welfare losses in the sectors where it occurs –even absent any externalities on the rest of society, and a fortiori in their presence.

The idea that labor market pressure will force firms to alter the structure of contracts they offer to workers is shared with a couple of recent papers.⁴ In Marin and Verdier (2009), international trade integration leads new entrants to compete with incumbents for managerial talent required to operate a firm; within each firm principals also find it increasingly optimal to delegate decision-making to middle-management, further raising the demand for skilled labor. The papers’ focus and mechanism are very different from ours, however. Agents receive no monetary incentives but derive private benefits from delegation, and those rents are non-monotonic with respect to competition. Furthermore, equilibrium changes in organizational design and activities performed tend to be efficient responses to relative factor endowments. In Acharya et al. (2011), firms can use two types of incentives: a reward in case of success and the threat of being taken over by a raider in case of failure. The latter deprives the manager of private benefits, so making takeovers hard to resist (“strong governance”) can be used to economize on bonus pay. Managers with high skills (which are observable) are in short supply, however, so in equilibrium they appropriate all the rents they generate. This forces their employers to renounce the threat of takeovers (“weak governance”), whereas firms employing the more abundant low-skill managers can still avail themselves of it. In contrast to our model, competition weakens here certain aspects of incentives (dismissal for failure) while strengthening others (reward for success). Most importantly, firms’ governance choices, and therefore also the competitiveness of the labor market, have no allocative impact: they only redistribute a fixed surplus between managers, shareholders and potential raiders. In Acharya et al. (2012), labor market competition interferes with the process of learning about agents’ abilities. A manager can invest in a safe asset or in a risky one whose return depends on his ability but will be observable (to the firm and others) only if he remains in charge of it, with the same employer, for two periods. There are no bonuses, so incentives are implicit (career concerns) ones. Absent mobility, firms can commit ex-ante to paying everyone the same lifetime wage, thus insuring managers against the risk of being of low ability; this also makes it optimal to quickly find out one’s talent, so as to choose the type of project one is better suited to. With free mobility, managers who stayed in a firm long enough to be revealed as talented would be bid away by competitors; thus, as in Harris and Holmström (1982), such insurance is precluded. Instead, during the early stages of their careers everyone moves to a different firm in each period so as to delay learning, and in each of these short-term jobs all managers inefficiently select the risky investment, as it has a higher unconditional expected return.

Finally, a recent literature has started incorporating considerations of intrinsic motivation into compensation design and labor-market sorting. Besley and Ghatak (2005, 2006) find conditions

⁴There is also an earlier literature examining the (generally ambiguous) effects of *product market* competition on managerial incentives and slack, whether through information revelation (Holmström 1982, Nalebuff and Stiglitz 1983) or demand elasticities and the level of profits (Schmidt 1997, Raith 2003).

under which assortative matching leads agents who derive private benefits from working in mission-oriented sectors to match with such firms, where they receive low pay but exert substantial effort. Focusing on civil-service jobs, Prendergast (2007) identifies conditions under which it is optimal to select employees who are either in empathy with their “clients” (teachers, social workers, firefighters) or somewhat hostile to them (police officers, tax or customs inspectors). When the state has imperfect information about agent’s types, however, it is generally not feasible to induce proper self-selection into jobs. Perhaps most closely related to our work in this literature is the multitask model of Kosfeld and von Siemens (2011), in which potential employees differ in their social preferences rather than productivity: some are purely self-interested, others conditional cooperators. Competition among employers leads to agents’ sorting themselves between “selfish jobs”, which involve high bonuses but no cooperation among coworkers and thus attract only selfish types, and “cooperative jobs” characterized by muted incentives and cooperative behavior, populated by conditional cooperators. Notably, positive profits emerge despite perfect competition. Because the source of heterogeneity is different from the main one emphasized in our paper, it is not surprising that the issue of excessive incentive pay does not arise in theirs.

2 Model

2.1 Agents

Preferences. A unit continuum of agents (workers) engage in two activities, exerting effort a and b respectively:

- Activity A is one in which individual contributions are not (easily) measurable and thus cannot be part of a formal compensation scheme: long-term investments enhancing the firm’s value, avoiding excessive risks and liabilities, cooperation, teamwork, etc. An agent’s contribution to A is then driven entirely by his intrinsic motivation, va , linear in the effort a exerted in this task. In addition to a genuine preference to “do the right thing” (e.g., an aversion to ripping off shareholders or customers, selling harmful products, teaching shoddily, etc.), v can also reflect social and self-image concerns such as fear of stigma or an executive’s concern for his “legacy”.⁵ In some contexts, v can also capture outside incentives not controlled by the firm, such as an executive’s potential legal liability.

- Activity B , by contrast, is measurable and therefore contractible: individual output, sales, short-term revenue, etc. When exerting effort b , a worker’s productivity is $\theta + b$, where θ is a talent parameter, privately known to each agent.

⁵Such preferences leading agents to provide some level of unrewarded effort were also part of Milgrom and Holmström’ original multitasking model (1991, Section 3). They make the model most tractable, but we would obtain similar results with an alternative version in which both tasks are rewarded with money, task A is observed with (more) noise and thus given weaker incentives, as is typically the case for long-term performance, and there is (more) variance in agents’ unknown productivity for task B . For recent analyzes of intrinsic motivation and social norms see, e.g., Bénabou and Tirole (2003, 2006), Besley and Ghatak (2005, 2006). In the specific context of firms, see Akerlof and Kranton (2005) for evidence and Ramalingam and Rauh (2010) for a model of investment in employees’ loyalty and identification.

The total effort cost, $C(a, b)$, is strictly increasing and strictly convex in (a, b) , with $C_{ab} > 0$ unless otherwise noted, meaning that the two activities are substitutes. A particularly convenient specification is the quadratic one, $C(a, b) = a^2/2 + b^2/2 + \gamma ab$, where $0 < \gamma < 1$, as it allows for simple and explicit analytical solutions to the whole model.⁶ These are provided in Appendix A, whereas in the text we shall maintain a general cost function, except where needed to obtain further results.

We assume an affine compensation scheme with incentive power or bonus rate y and fixed wage z , so that total compensation is $z + (\theta + b)y$.⁷ Agents have quasi-linear preferences

$$U(a, b; \theta, y, z) = va + (\theta + b)y + z - C(a, b). \quad (1)$$

Types. To emphasize the roles of heterogeneity in v and θ , respectively, we shall focus on two polar cases. Here and throughout Section 3, agents differ only in their productivities for the B task. Thus $\theta \in \{\theta_L, \theta_H\}$, with $\Delta\theta \equiv \theta_H - \theta_L > 0$, and θ_i having probability q_i . In Section 4, conversely, we shall consider agents who differ only in their intrinsic motivations v for task A .

Effort allocation. When facing compensation scheme (y, z) , the agent chooses efforts $a(y)$ and $b(y)$ so as to solve

$$\max_{(a, b)} \{va + (\theta + b)y + z - C(a, b)\}, \quad (2)$$

leading to the first-order conditions $\partial C/\partial a = v$, $\partial C/\partial b = y$. Our assumptions on the cost function imply that increasing the power of the incentive scheme raises effort in the measured task and decreases it in the unobserved one:

$$\frac{da}{dy} < 0 < \frac{db}{dy}.$$

It will prove convenient to decompose the agent's utility into an "allocative" term, $u(y)$, which depends on the endogenous efforts, and a "redistributive" one, $\theta y + z$, which does not:

$$U(y; \theta, z) \equiv U(a(y), b(y); \theta, y, z) = u(y) + \theta y + z, \quad (3)$$

where

$$u(y) \equiv va(y) + yb(y) - C(a(y), b(y)). \quad (4)$$

Note that $u'(y) = b(y)$ and $\partial U(y; \theta, z)/\partial y = \theta + b(y)$.

Outside opportunities. We assume that any agent can obtain a reservation utility \bar{U} , so that employers must respect the participation constraint:

$$U(y; \theta, z) = u(y) + \theta y + z \geq \bar{U}. \quad (5)$$

⁶ The model also works when the two tasks are complements, $C_{ab} < 0$ (e.g., $-1 < \gamma < 0$) but the results in this case are less interesting, e.g., competition is now, predictably, always more efficient than monopsony.

⁷ Unrestricted nonlinear schemes (as in Laffont and Tirole (1986)) lead to qualitatively very similar results; see Appendix B and the Propositions therein.

The type-independence of the outside option is a polar case that will help highlight the effects of competition *inside* the labor market. Thus, under monopsony every one has reservation utility equal to \bar{U} , whereas with competition reservation utilities become endogenous and type-dependent.⁸

2.2 Firm(s)

A worker of ability θ exerting efforts (a, b) generates a gross revenue $Aa + B(\theta + b)$ for his employer. Employing such an agent under contract (y, z) thus results in a net profit of

$$\Pi(\theta, y, z) = \pi(y) + (B - y)\theta - z, \quad (6)$$

where

$$\pi(y) \equiv Aa(y) + (B - y)b(y). \quad (7)$$

represents the allocative component and $(B - y)\theta - z$ a purely redistributive one.

2.3 Social welfare

In order to better highlight the mechanism at work in the model, we take as our measure of social welfare the sum of workers' and employers' payoffs, thus abstracting from any externalities on the rest of society.⁹ Again, it will prove convenient to decompose it into an allocative part, $w(\theta)$, and a surplus, $B\theta$, that is independent of the compensation scheme (the pure transfer, $(\theta + b)y + z$, nets out):

$$W(\theta, y) \equiv U(a(y), b(y); \theta, y, z) + \Pi(\theta, y, z) = w(y) + B\theta, \quad (8)$$

where

$$w(y) \equiv u(y) + \pi(y) = (A + v)a(y) + Bb(y) - C(a(y), b(y)). \quad (9)$$

Using the envelope theorem for the worker, $u'(y) = b(y)$, we have:

$$w'(y) = Aa'(y) + (B - y)b'(y). \quad (10)$$

We take w to be strictly concave,¹⁰ with a maximum at y^* given by

$$w'(y^*) = Aa'(y^*) + (B - y^*)b'(y^*) = 0 \quad (11)$$

and generating enough surplus that even low types can be profitably employed, namely

$$w(y^*) + \theta_L B > \bar{U}. \quad (12)$$

⁸We make the usual assumption that when a worker is indifferent between an employer's offer and his reservation utility he chooses the former. We will also assume that \bar{U} is high enough that the constraint $z \geq 0$ is never binding, but not so large that hiring one of the worker types is unprofitable (see Appendix D for the exact conditions).

⁹For instance, we can think of firms' output as being sold on a perfectly competitive product market. It is, however, very easy to incorporate social spillovers into the analysis, as we explain below.

¹⁰Such is the case, in particular, with quadratic costs; see Appendix A.

In cases where (underprovision of) the “ethical” activity a also has spillovers on the rest of society –be they technological (pollution), pecuniary (imperfect competition in the product market) or fiscal (cost of government bailouts, taxes or subsidies)– total social welfare becomes $w(y) + e \cdot a(y)$, where e is the per-unit externality. Clearly, this will only strengthen our main results about the competitive overincentivization of the other activity, b .

3 Competing for talent

Throughout most of the paper (except for Section 4), v is known while $\theta = \{\theta_H, \theta_L\}$, with mean $\bar{\theta} \equiv q_L \theta_L + q_H \theta_H$. We consider in sequence monopsony, perfect and imperfect competition. Before proceeding, it is worth noting that if agents’ types $k = H, L$ were observable, the only impact of market structure would be on the fixed wages z_k , whereas incentives would always remain at the efficient level, $y_k = y^*$.

3.1 Monopsony employer

A monopsonist (or set of colluding firms) selects a menu of contracts (y_i, z_i) aimed at type $i \in \{L, H\}$. We assume that it wants to attract both types, which, as we will show, is equivalent to q_L exceeding some threshold.

The monopsonist maximizes expected profit

$$\max_{\{(y_i, z_i)\}_{i=H,L}} \left\{ \sum_{i=H,L} q_i [\pi(y_i) + (B - y_i)\theta_i - z_i] \right\}$$

subject to the incentive constraints

$$u(y_i) + \theta_i y_i + z_i \geq u(y_j) + \theta_j y_j + z_j \quad \text{for all } i, j \in \{K, L\} \quad (13)$$

and the low-productivity type’s participation constraint

$$u(y_L) + \theta_L y_L + z_L \geq \bar{U}. \quad (14)$$

This program is familiar from the contracting literature. First, incentive constraints, when added up, yield $(\theta_i - \theta_j)(y_i - y_j) \geq 0$, so the power of the incentive scheme must be non-decreasing in type –a more productive agent must receive a higher fraction of his measured output. Second, the low type’s incentive constraint is binding, and the high type’s rent above \bar{U} is given by the extra utility obtained by mimicking the low type: $(\Delta\theta)y_L$. Rewriting profits, the monopsonist solves:

$$\max_{\{(y_i, z_i)\}_{i=H,L}} \left\{ \sum_{i=H,L} q_i [w(y_i) + B\theta_i] - \bar{U} - q_H(\Delta\theta)y_L, \right\}$$

yielding $y_H^m = y^*$ (no distortion at the top) and¹¹

$$w'(y_L^m) = \frac{q_H}{q_L} \Delta\theta, \quad \text{implying } y_L < y^*. \quad (15)$$

The principal reduces the power of the low-type's incentive scheme, so as to limit the high-type's rent. It is optimal for the firm to hire both types if and only if

$$q_L [w(y_L^m) + B\theta_L - \bar{U}] \geq q_H y_L^m \Delta\theta, \quad (16)$$

meaning that the profits earned on low types exceed the rents abandoned to high types. By (15), the difference of the left- and right-hand sides is increasing in q_L , so the condition is equivalent to $q_L \geq \underline{q}_L$, where \underline{q}_L is defined by equality in (16).

Proposition 1 (monopsony) *Let (16) hold, so that the monopsonist wants to employ both types. Then $y_H^m = y^*$ and $y_L^m < y^*$ is given by*

$$w'(y_L^m) = \frac{q_H}{q_L} \Delta\theta,$$

with corresponding fixed payments $z_H^m = \bar{U} + y_L^m \Delta\theta - u(y^*) - \theta_H y^*$ and $z_L^m = \bar{U} - u(y_L^m) - \theta_L y_L^m$. The resulting welfare loss is equal to

$$L^m = q_L [w(y^*) - w(y_L^m)]. \quad (17)$$

It increases with $\Delta\theta$, but need not be monotonic in A or B .

Note that since total social welfare is $q_H [w(y_H) + B\theta_H] + q_L [w(y_L) + B\theta_L]$, a mean-preserving increase in the distribution of θ always reduces it, by worsening the informational asymmetry. In contrast, an increase in A (or a decrease in B) has two opposing effects on L^m : (i) it makes any given amount of underincentivization on the B task less costly, as the alternative task A is now more valuable; (ii) the efficient bonus rate y^* given to the high types declines, and to preserve incentive compatibility so must y_L^m , worsening low types' underincentivization. In the quadratic case, the two effects cancel out, shown in Appendix A.

3.2 Perfect competition in the labor market

A large number of firms now compete for workers, each one offering an incentive-compatible menu of contracts. We first look for a separating competitive allocation, defined as one in which: (i) each worker type chooses a different contract, respectively (y_L, z_L) and (y_H, z_H) for $i = H, L$, with resulting utilities U_L and U_H ; (ii) each of these two contracts makes zero profits, implying in

¹¹To exclude uninteresting corner solutions we shall assume that $w'(0) > q_H \Delta\theta / q_L$. Since later on we shall impose various other upper bounds on q_H , this poses no problem.

particular the absence of any cross-subsidy.¹² Then, in a second stage, we investigate the conditions under which this allocation is indeed an equilibrium, and even the unique one.

In a separating competitive equilibrium, any contract that operates must make zero profit:

$$\Pi(\theta_H, y_H, z_H) = 0 \iff \pi(y_H) + (B - y_H)\theta_H = z_H, \quad (18)$$

$$\Pi(\theta_L, y_L, z_L) = 0 \iff \pi(y_L) + (B - y_L)\theta_L = z_L, \quad (19)$$

which pins down z_H and z_L . Furthermore, a simple Bertrand-like argument implies that the low type must receive his symmetric-information efficient allocation,¹³

$$y_L^c = y^* \quad \text{and} \quad z_L^c = \pi(y^*) + (B - y^*)\theta_L.$$

He should then not benefit from mimicking the high type, nor vice-versa,

$$w(y^*) + B\theta_L \geq u(y_H) + \theta_L y_H + z_H = w(y_H) + B\theta_H - y_H\Delta\theta, \quad (20)$$

$$w(y_H) + B\theta_H \geq w(y^*) + B\theta_L + y^*\Delta\theta, \quad (21)$$

implying in particular that $y_H \geq y^*$. Among all such contracts, the most attractive to the high types is the one involving minimal distortion, namely such that (20) is an equality:

$$w(y_H^c) = w(y^*) - (B - y_H^c)\Delta\theta. \quad (22)$$

By strict concavity of w , this equation has a unique solution y_H^c to the right of y^* , satisfying $y^* < y_H^c < B$. The inequality in (20) is then strict, meaning that only the low type's incentive constraint is binding. Note that, as illustrated in Figure I, this is exactly the reverse of what occurred under monopsony.¹⁴

Existence and uniqueness. When is this least-cost separating (LCS) allocation indeed an equilibrium, or the unique equilibrium of the competitive-offer game? The answer, which is reminiscent of Rothschild and Stiglitz (1976), hinges on whether or not a firm could profitably deviate to a contract that achieves greater total surplus by using a cross-subsidy from high to low types to

¹²One can then can indifferently think of each firm offering a menu and employing both types of workers, or of different firms specializing in a single type by offering a unique contract.

¹³Absent cross-subsidies, the low type cannot receive more than the total surplus $w(y^*) + \theta_L B$ he generates under symmetric information, or else his employer would make a negative profit. Were he to receive less, conversely, another firm could attract him by offering $(y^*, z_L = z_L^c - \varepsilon)$ for ε small, leading to a profit ε on this type (and an even larger one on any high type who also chose this contract). Low types must thus be offered utility equal to $w(y^*) + \theta_L B$, which only their symmetric-information efficient allocation achieves.

¹⁴To build intuition for this reversal, one can look at how the symmetric-information outcome fails incentive compatibility. Under monopsony no employee obtains any rent, so type k 's symmetric information contract is: $(y_k = y^*, z_k = \bar{U} - va(y^*) + C(a(y^*), b(y^*)) - [\theta_k + b(y^*)]y^*)$. If it were still offered under asymmetric information, the high type could then obtain a positive rent $(\Delta\theta)y^*$ by choosing (y^*, z_L) ; we thus expect the downward incentive constraint to bind. Under perfect competition, the symmetric-information contract for type k is $(y_k = y^*, z_k = \pi(y^*) + (\theta_k - y^*)B)$. If it prevailed under asymmetric information, the low type could obtain extra utility $B\Delta\theta$ by choosing (y_H, z_H) . We thus now expect the upward incentive constraint to bind.

ensure incentive compatibility.

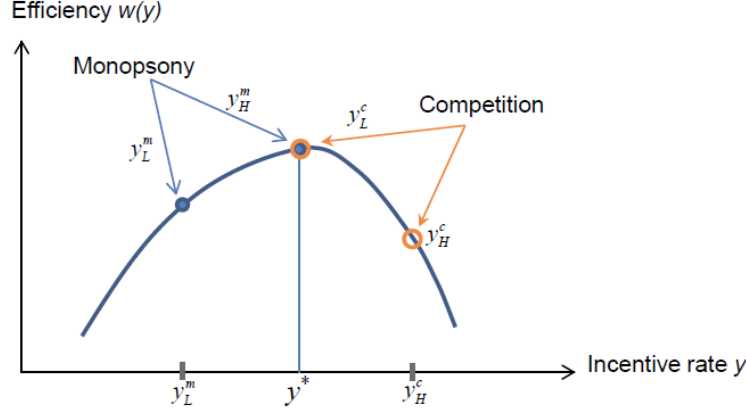


Figure I: Distortions under monopsony and perfect competition

Definition 1 An incentive-compatible allocation $\{(U_i^*, y_i^*)\}_{i=H,L}$ is interim efficient if there exists no other incentive-compatible $\{(U_i, y_i)\}_{i=H,L}$ that

- (i) Pareto dominates it: $U_H \geq U_H^*$, $U_L \geq U_L^*$, with at least one strict inequality.
- (ii) Makes the employer break even on average: $\sum_i q_i [w(y_i) + \theta_i B - U_i] \geq 0$.

For the LCS allocation to be an equilibrium, it must be interim efficient. Otherwise, there is another menu of contracts that Pareto dominates it, which one can always slightly modify (while preserving incentive-compatibility) so that both types of workers and the employer share in the overall gain; offering such a menu then yields strictly positive profits. The converse result is also true: under interim efficiency there can clearly be no positive-profit deviation that attracts both types of agents, and by a similar type of “small redistribution” argument one can also exclude those that attract a single type. These claims are formally proved in the appendix, where we also show that when the LCS allocation is interim efficient, it is in fact the unique equilibrium. Furthermore, we identify a simple condition for this to be case:

Lemma 1 The least-cost separating allocation is interim efficient if and only if

$$q_H w'(y_H^c) + q_L \Delta \theta \geq 0. \quad (23)$$

This condition holds whenever q_L exceeds some threshold $\tilde{q}_L < 1$.

The intuition is as follows. At the LCS allocation, we saw that the binding incentive constraint is the low type’s: $U_L^c = U_H^c - y_H^c \Delta \theta$. Consider now an employer who slightly reduces the power of the high type’s incentive scheme, $\delta y_H = -\varepsilon$, while using lump-sum transfers $\delta z_H = (b(y_H) + \theta_H) \varepsilon + \varepsilon^2$ to slightly more than compensate them for the reduction in incentive pay and $\delta z_L = \varepsilon \Delta \theta + \varepsilon^2$

to preserve incentive compatibility. Such a deviation attracts both types ($\delta U_H = \delta U_L = \varepsilon^2$, since $u'(y) = b(y)$) and its first-order impact on profits is

$$q_H [(\pi'(y_H) - \theta_H) \delta y_H - \delta z_H] - q_L \delta z_L = [q_H w'(y_H^c) + q_L \Delta\theta] (-\varepsilon)$$

Under (23) this net effect is strictly negative, hence the deviation unprofitable. When (23) fails, conversely, the increase in surplus generated by the more efficient effort allocation of the high types is sufficient to make the firm and all its employees strictly better off. A higher $q_L = 1 - q_H$ means fewer high types to generate such a surplus and more low types to whom rents (cross-subsidies) must be given to maintain incentive compatibility, thus making (23) more likely to hold.

We can now state this section's main results.

Proposition 2 *Let $q_L \geq \tilde{q}_L$. The unique competitive equilibrium involves two separating contracts, both resulting in zero profit:*

1. *Low-productivity workers get (y^*, z_L^c) where z_L^c is given by (19).*
2. *High-productivity ones get (y_H^c, z_H^c) , where z_H^c is given by (18) and $y_H^c > y^*$ by*

$$w(y^*) - w(y_H^c) = (B - y_H^c) \Delta\theta.$$

3. *The efficiency loss relative to the social optimum is*

$$L^c = q_H [w(y^*) - w(y_H^c)] = (B - y_H^c) q_H \Delta\theta. \tag{24}$$

It increases with $\Delta\theta$ and A , but need not be monotonic in B .

These results confirm and formalize the initial intuition that competition for talent will result in an overincentivization of higher-ability types. As shown on Figure I, this is the opposite distortion from that of the monopsony case, which featured underincentivization of low-ability types. When the degree of competition is allowed to vary continuously (Section 3.4), we therefore expect that there will be a critical point at which the nature of the distortion (reflecting which incentive constraint is binding) tips from one case to the other.

The model also shows how *skill-biased technical change* (a higher θ_H) exacerbates the competition for talented types, resulting in a higher bonus rate y_H^c that makes their performance-based pay rise *more than proportionately* with their productivity $B\theta_H$. This is in line with Lemieux et al.'s (2009) findings about the contribution of performance pay to rising earnings inequality. This equilibrium market response to technical or human-capital change is *inefficient*, however, as it worsens the underprovision of long-term investments and other voluntary efforts inside firms, thereby reducing the social value of the productivity increase. For a mean-preserving spread in the distribution of θ 's only the deadweight loss remains, so overall social welfare declines.

What happens when the LCS allocation is not interim efficient, that is, when $q_L < \tilde{q}_L$? We saw that it is then not an equilibrium, since there exist profitable deviations to incentive-compatible

contracts (involving cross-subsidies) that Pareto-dominate it. We also show in the appendix that no other pure-strategy allocation is immune to deviations, a situation that closely parallels the standard Rothschild and Stiglitz (1976) problem: the only equilibria are in mixed strategy.¹⁵ Since such an outcome is not really plausible as a stable labor-market outcome, we assume from here on

$$q_L \geq \max\{\tilde{q}_L, \underline{q}_L\} \equiv q_L^*. \quad (25)$$

3.3 Welfare: monopsony versus perfect competition

Single task ($A = 0$). As a benchmark, it is useful to recall that competition is always socially optimal with a single task. The competitive outcome is then the single contract $y^c = y^* = B$, $z^c = 0$; agents of either type are residual claimants for their production and therefore choose the efficient effort allocation.¹⁶ By contrast, leads to a downward distortion in the power of the incentive scheme. Hence competition is always strictly welfare superior.

Multitasking. Let us now compare the welfare losses under monopsony and competition. From (17) and (22), $L^m < L^c$ is larger if and only if:

$$q_L [w(y^*) - w(y_L^m)] < q_H [w(y^*) - w(y_H^c)]. \quad (26)$$

Intuitively, this occurs when competition pushes y_H^c above y^* by (sufficiently) more than monopsony depresses y_L^m below y^* . In the quadratic-cost case, we are able to obtain an explicit condition.

Proposition 3 *Let $C(a, b) = a^2/2 + b^2/2 + \gamma ab$. Social welfare is lower under competition than under monopsony if and only if $q_L \geq q_L^*$ and*

$$\frac{q_H}{2q_L} + \sqrt{\frac{q_H}{q_L}} < \left(\frac{\gamma}{1 - \gamma^2} \right) \left(\frac{A}{\Delta\theta} \right). \quad (27)$$

The underlying intuitions are quite general.¹⁷ First, competition entails a larger efficiency loss when the unrewarded task (long-run investments, cooperation, avoidance of excessive risks, etc.) is important enough and the two types of effort sufficiently substitutable; if they are complements ($\gamma < 0$), in contrast, competition is always efficiency-promoting. Second, the productivity differential $\Delta\theta$ scales the severity of the asymmetric-information problem that underlies both the monopsony and the competitive distortions. A monopsonistic firm optimally trades off total surplus

¹⁵The alternative approach is to assume that it is workers who make take-it-or-leave offers, instead of a competitive industry making offers to them. From Maskin and Tirole (1992) we know that for $q_L \geq \tilde{q}_L$, the unique equilibrium of the resulting “informed-principal game” is the LCS allocation, so the result is the same as here with competitive offers. By contrast, for $q_L < \tilde{q}_L$, the set of equilibrium interim utilities is the set of feasible utilities (incentive compatible and satisfying budget balance in expectation) that Pareto dominate (U_L^c, U_H^c) .

¹⁶Similar results holds if $A > 0$ but $v = 0$: since $a \equiv 0$ for all y , the socially optimal bonus rate is $y^* = B$, even though it results in an inefficient effort allocation. This is also the competitive outcome, since the B task is the sole source of revenue for the firms.

¹⁷Indeed, the model’s solution with quadratic costs (given in Appendix A) also correspond to Taylor approximations of the more general case when $\Delta\theta$ is small, provided $1/(1 - \gamma^2)$ is replaced everywhere by $-w''(y^*)$.

versus rent-extraction, so (by the envelope theorem) a small $\Delta\theta$ has only a second-order effect on overall efficiency. Under competition the effect is first-order, because a firm raising its y_H does not internalize the deterioration in the workforce quality it inflicts on its competitors –or, equivalently, the fact that in order to retain their “talent” they will also have to distort incentives and the allocation of effort. This general intuition explains why (27) is more likely to hold when $\Delta\theta$ decreases.¹⁸ Finally, as can also be seen from (15) and (22), the monopsony distortion $y^* - y_L^m$ is increasing with q_H/q_L (limiting the high types’s rents becomes more important), whereas the competitive one, $y_H^c - y^*$, is independent of it (being determined by an incentive constraint across types). The ratio L^m/L^c is therefore of order $q_L(q_H/q_L)^2/q_H = q_H/q_L$, so the greater the proportion of L types, the better monopsony does relative to competition.

3.4 Imperfect competition

Monopoly and perfect competition are useful benchmarks to illustrate key mechanisms, but of limited empirical relevance. To understand more generally how the intensity of labor market competition affects the equilibrium structure of wages, workers’ task allocation, firms’ profits and social welfare, we now develop a variant of the Hotelling model in which competitiveness can be appropriately parametrized.

As illustrated in Figure II, a unit continuum of workers are uniformly distributed along the unit interval, $x \in [0, 1]$. Two firms, $k = 0, 1$, are located respectively at the left and right extremities and recruit them to produce, with the same production function as before. When a worker located at x chooses to work for Firm 0 (resp., 1), he incurs a cost equal to the distance tx (resp., $t(1-x)$) that he must travel. We assume that θ and x are independent and that a worker’s position is not observable by employers, who therefore cannot condition contracts on this characteristic.

In the standard Hotelling model, agents also have an outside option (e.g., staying put) that yields a fixed level of utility, \bar{U} . This implies, however, that a change in t affects not only competitiveness *within* the market (firm 1 vs. 2) but also, mechanically, that of the *outside* option –formally, agents’ participation constraints. To isolate the pure competitiveness of the market from that of other (in)activities, we introduce an intuitive but novel modeling device.

Co-located outside option. Instead of receiving the outside option \bar{U} for free, agents must also “go and get it” at either the end of the unit interval, which involves paying the same cost tx or $t(1-x)$ as if they chose Firm 0 or Firm 1, respectively. One can think of two business districts, each containing both a multitask firm of the type studied here and a competitive fringe or informal sector in which all agents have productivity \bar{U} . Alternatively, each agent could produce \bar{U} “at home” but then have to travel (or adapt) to one or the other marketplace to sell his output.

¹⁸As we show in Appendix B, for small $\Delta\theta$ the lower bound \tilde{q}_L above which (23) holds (and the competitive equilibrium thus exists) is such that $1 - \tilde{q}_L$ is of order $\sqrt{\Delta\theta}$. Thus, to rigorously apply the above reasoning involving first- versus second-order losses for small $\Delta\theta$, one needs to also let q_H become small enough. This further reduces $y^* - y_L^m$ while leaving $y_H^c - y^*$ unchanged. This further increases L^c/L^m , making it of order $(\Delta\theta)^{-3/2}$ rather than $(\Delta\theta)^{-1}$.

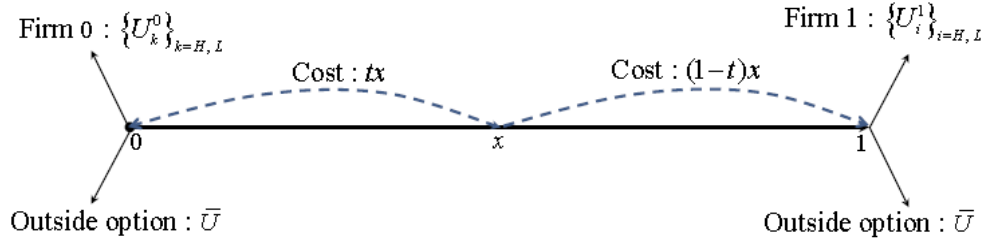


Figure II: Hotelling with co-located outside options

Without loss of generality, we can assume that each firm $k = 0, 1$ offers an incentive-compatible menu of compensation schemes $\{y_i^k, z_i^k\}_{i=H,L}$, in which workers who opt for this employer self-select. Let U_i^k denote the utility provided by firm k to type i :

$$U_i^k \equiv u(y_i^k) + \theta_i y_i^k + z_i^k. \quad (28)$$

A worker of type i , located at x , will choose Firm $k = 0$ (say) if and only if

$$U_i^k - tx \geq \max \left\{ \bar{U} - tx, \bar{U} - t(1-x), U_i^\ell - t(1-x), \right\} \quad (29)$$

The first inequality reduces to $U_i^k \geq \bar{U}$: a firm must at least match its local outside option. In any equilibrium in which both firms attract L -type workers, therefore, $U_i^\ell \geq \bar{U}$ and so the second inequality is redundant.

We shall initially assume that: (i) both firms indeed prefer to employ the two types of workers than only the more productive ones; (ii) no firm wants to “corner” the market on either type of worker, i.e. move the corresponding cutoff value of x all the way to 0 or 1. Intuition suggests that if (i) is true for a monopsonist, meaning that $q_L \geq \underline{q}_L$ defined by equality in (16), it is a fortiori true when there is some degree of competition. At the other end of the spectrum, intuition also suggests that if the perfectly competitive equilibrium involves separating contracts (LCS allocation), which occurs when $q_L \geq \tilde{q}_L$ defined in Lemma 1, then such will also be the case when agents are less mobile. Verifying these intuitions in Appendix D, we shall assume from here on that

$$q_L \geq \max \left\{ \underline{q}_L, \tilde{q}_L \right\} \equiv q_L^* \quad (30)$$

and focus on looking for a symmetric equilibrium in separating contracts, in which each firm employs half of each type of worker.

Given utilities (U_i^k, U_i^ℓ) resulting in interior cutoffs, firm k 's share of workers of type i is then

$$x_i^k(U_i^k, U_i^\ell) = \frac{U_i^k - U_i^\ell + t}{2t}. \quad (31)$$

and it chooses (U_L, U_H, y_L, y_H) to solve the program:

$$\max \left\{ q_H(U_H - U_H^\ell + t)[w(y_H) + \theta_H B - U_H] + q_L(U_L - U_L^\ell + t)[w(y_L) + \theta_L B - U_L] \right\} \quad (32)$$

subject to:

$$U_H \geq U_L + y_L \Delta \theta \quad (\mu_H) \quad (33)$$

$$U_L \geq U_H - y_H \Delta \theta \quad (\mu_L) \quad (34)$$

$$U_L \geq \bar{U} \quad (\nu) \quad (35)$$

To shorten the notation, let $m_i \equiv w(y_i) + \theta_i B - U_i$ denote the firm's margin on type $i = H, L$. The first-order conditions, together with the requirement that $U_i = U_i^l$ in a symmetric equilibrium, are:

$$q_H(m_H - t) + \mu_H - \mu_L = 0 \quad (36)$$

$$q_L(m_L - t) + \mu_L - \mu_H + \nu = 0 \quad (37)$$

$$tq_H w'(y_H) + \mu_L \Delta \theta = 0 \quad (38)$$

$$tq_L w'(y_L) - \mu_H \Delta \theta = 0. \quad (39)$$

It is easily seen that μ_H and μ_L cannot both be strictly positive: otherwise (33) and (34) would both bind, so $y_H = y_L$, but this renders (38)-(39) mutually incompatible.

Solving for the equilibrium: key intuitions. Because solving the above problem for all values of t is complicated, we provide here the underlying intuitions and relegate the proofs to the appendix.

(a) For large t , the equilibrium should resemble the monopsonistic one: the main concern is limiting high types' rent, so firms distort $y_L < y^* = y_H$ to make imitating low types unattractive. Conversely, for small t , the equilibrium should resemble perfect competition: the main concern is attracting the H types, leading employers to offer them high-powered incentives, $y_H > y^* = y_L$.

(b) As t declines over the whole real line, the high types' responsiveness to higher offered utility U_H rises, so firms are forced to leave them more rent. Since that rent is either $y_L \Delta \theta$ or $y_H \Delta \theta$ (depending on which of the above two concerns dominates, i.e. on which types' incentive constraint is binding), y_L and y_H should both be nonincreasing in t .

(c) Firms 0 and 1 are always actively competing for the high types. If t is low enough, they also compete for L types, offering them a surplus above their outside option: $U_L > \bar{U}$. At the threshold t_1 below which U_L starts exceeding \bar{U} , y_H has a convex kink: since the purpose of keeping y_H above y^* is to maintain a gap $U_H - U_L = y_H \Delta \theta$ just sufficient to dissuade low types from imitating high ones, as U_L begins to rise above \bar{U} , the rate of increase in y_H can be smaller.

These intuitions translate into a characterization of the equilibrium in terms of three regions, illustrated in Figure III and formally stated in Proposition 4 below.¹⁹

¹⁹With quadratic costs one can furthermore show (see Appendix A) that each of the curves is convex, as drawn on the figure.

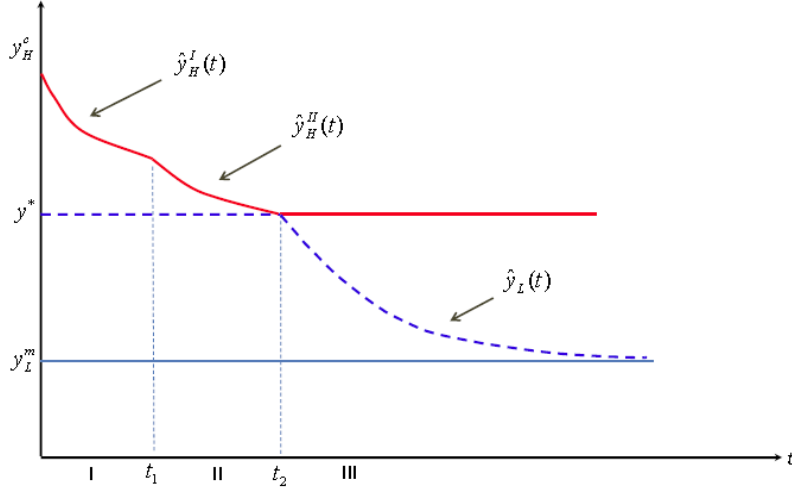


Figure III: equilibrium incentives under imperfect competition

Proposition 4 (Imperfect competition) *Let $q_L \geq q_L^*$. There exist unique thresholds $t_1 > 0$ and $t_2 > t_1$ such that, in the unique symmetric market equilibrium:*

1. *Region I (strong competition): for all $t < t_1$, bonuses are $y_L = y^* < \hat{y}_H^I(t)$, strictly decreasing in t , starting from $\hat{y}_H^I(0) = y_H^c$. The low type's participation constraint is not binding, $U_L > \bar{U}$, while the high type's incentive constraint constraint is: $U_H - U_L = \hat{y}_H^I(t)\Delta\theta$.*
2. *Region II (medium competition): for all $t \in [t_1, t_2)$, bonuses are $y_L = y^* < \hat{y}_H^{II}(t)$, with $\hat{y}_H^{II}(t) < \hat{y}_H^I(t)$ except at t_1 and strictly decreasing in t . The low type's participation constraint is binding, $U_L = \bar{U}$, and so is the high types' incentive constraint: $U_H - U_L = \hat{y}_H^{II}(t)\Delta\theta$.*
3. *Region III (weak competition): for all $t \geq t_2$, bonuses are $y_L = \hat{y}_L(t) < y^* = y_H$, with $\hat{y}_L(t)$ strictly decreasing in t and $\lim_{t \rightarrow +\infty} \hat{y}_L(t) = y_L^m$. The low type's participation and incentive constraints are both binding : $U_L = \bar{U}$, $U_H - U_L = \hat{y}_L(t)\Delta\theta$.*

Welfare. For each value of t , either y_L or y_H is equal to the (common) first-best value y^* , while the other bonus rate, which creates the distortion, is strictly increasing in t . Therefore, as illustrated on Figure IV:

Proposition 5 (optimal degree of competition) *Social welfare is hill-shaped as a function of the degree of competition in the labor market, reaching the first-best at $t_2 = w(y^*) + \theta_H(B - y^*) + \theta_L y$, where $y_L = y^* = y_H$.*

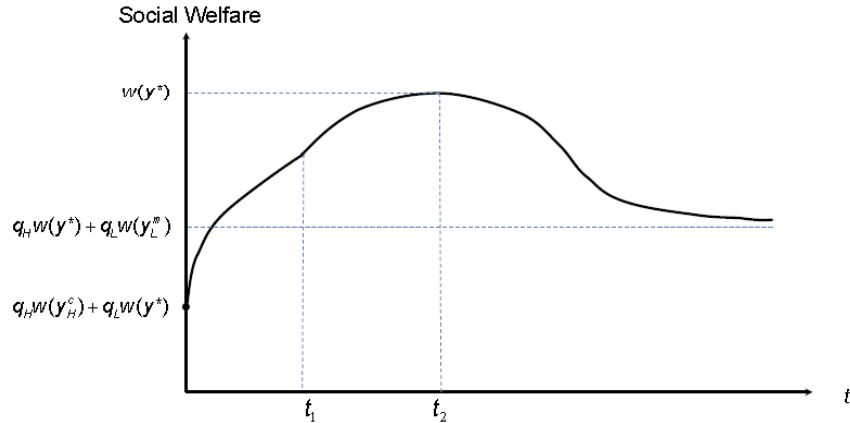


Figure IV: competition and social welfare

We can also say how these gains and losses in total welfare are distributed among the different actors in the market.

Proposition 6 (individual welfare and firm profits) *As the labor market becomes more competitive (t declines), both U_H and U_L increase (weakly for the latter), but inequality in workers' utilities, $U_H - U_L$ always strictly increases; firms' total profits strictly decline.*

In Regions III and II, $U_L = \bar{U}$. In Region I, U_L is decreasing in t , as we show in the appendix. Since $(U_H - U_L) / \Delta\theta$ is equal to $\hat{y}_H(t)$ over Regions I and II and to $\hat{y}_L(t)$ over Region III, it follows directly from Proposition 4 that $\partial U_H / \partial t \leq \partial (U_H - U_L) / \partial t < 0$. As to profits, they must clearly fall as t declines over Regions II and I, since overall surplus is shrinking but all workers are gaining. In Region III, as $\hat{y}_L(t)$ rises firms reap some of the efficiency gains from low-type agent's more efficient effort allocation, but the rents they must leave to high types increase even faster (as shown in the appendix), so total profits decline here as well.

Income inequality. Let us now examine the effects of a more competitive labor market on earnings, which is what is measured in practice. While we analyze these comparative statics over all $t \in \mathbb{R}_+$, the empirically relevant range for most sectors in a modern market economy is that of *medium to high mobility*, namely Regions I and II on Figure III. Indeed, this is where firms are more concerned with retaining and bidding away from each other the workers of high ability who can easily switch (x close to $1/2$) than with exploiting their more "captive" local markets (forcing down the rents of those with x close to 0 or 1). We compare how the two types of workers $i = H, L$ fare in terms of total earnings $Y_i \equiv [b(y_i) + \theta_i] y_i + z_i$, as well as the separate contributions of performance-based and fixed pay.

Proposition 7 (income inequality) *Let $q_L \geq q_L^*$. As the labor market becomes more competitive (t declines), both Y_H and Y_L increase (weakly for the latter). Furthermore,*

1. Over Regions I and II (medium and high competition), inequality in total pay $Y_H - Y_L$ rises, as does its performance-based component. Inequality in fixed wages declines, so changes in performance pay account for more than 100% of the rise in total inequality.
2. Over Region III (low competition), inequality in performance pay declines, while inequality in fixed wages rises. As a result, inequality in total pay need not be monotonic. With quadratic costs, a sufficient condition for it to rise as t declines is $B \leq \gamma A + (1 - \gamma^2)\Delta\theta$.

These results are broadly consistent with the findings of Lemieux et al. (2009) about the driving role of performance pay in rising earnings inequality, as well as their hypothesis that the increased recourse to performance pay also serves a screening purpose. They are also in line with Frydman’s (2007) evidence linking increased mobility (skills portability) of corporate executives to the rise in both the level and the variance in their compensation.²⁰

We demonstrate here the results for total earnings, leaving the others to the appendix. Since $z_i = U_i - u(y_i) - \theta_i y_i$, we can write $Y_i = U_i + b(y_i)y_i - u(y_i)$, for $i = H, L$. As t declines, U_i and y_i increase (at least weakly) and therefore so does Y_i , since $u'(y) = b$. Furthermore,

$$Y_H - Y_L = U_H - U_L + b(y_H)y_H - u(y_H) + u(y_L) - b(y_L)y_L.$$

Over Regions I and II this becomes $[\Delta\theta + b(y_H)]y_H - u(y_H)$ plus a constant term, with $y_H = \hat{y}_H(t)$; the result then follow from $u'(y) = b$. Over Region III, $Y_H - Y_L = [\Delta\theta - b(y_L)]y_L + u(y_L)$ plus a constant term, with $y_L = \hat{y}_L(t)$; therefore, $\partial(Y_H - Y_L)/\partial t < 0$ if and only if $b'(y_L)y_L < \Delta\theta$, which need not hold in general. With quadratic costs, $b'(y_L) = 1/(1 - \gamma^2)$ so it holds on $[t_2, +\infty)$ if and only if $y^* = B - \gamma A < (1 - \gamma^2)\Delta\theta$.

3.5 Regulating compensation (preliminary)

We focus the analysis on the case of perfect competition, which is the most relevant. Since the source of distortion is the “bonus culture”, a first question is whether the government or regulator is able to distinguish and treat differently the performance-related and fixed parts of compensation. If it can, then a first simple result is that the first-best can be restored by capping bonuses at y^* : the unique equilibrium under such a cap is a pooling one where all firms offer, and all workers take, the single contract $(y^*, \pi(y^*) + (B - y^*)\bar{\theta})$.

Of course, things in practice may not be so simple. First, firms may switch to alternative forms of rewards that (at the margin) still appeal differentially to different types but may be even less efficient screening devices than untaxed bonuses: latitude to serve on other companies’ boards, to engage in own practice (doctors) or consulting, lower lock-in to company (low clawbacks, easier terms for quitting), etc. Second, they may relabel fixed and variable compensation, or uses the above-mentioned alternative rewards to blur that line. While characterizing the optimal nonlinear tax in this setting is somewhat complicated and left for future work, we can show:

²⁰See also Frydman and Saks’ (2005) evidence about the rising share of performance-related pay in top CEO’s compensation.

Proposition 8 *A small tax τ on total compensation always improves welfare: $dW/d\tau|_{\tau=0} > 0$.*

The intuition for this result is as follows. Taxing total earnings has two effects. First, under symmetric information, it distorts (net) incentives downward relative to the private and social optimum, y^* . Second, it shrinks the compensation differential received by the two types under any given contract. This reduces low types' incentive to mimic high ones, thus dampening firms' need to screen through high-powered (net) incentives and thereby alleviating the misallocation of effort. For small τ the first effect is of second-order (a standard Harberger triangle), whereas the second one is of first order, due again to the externality between firms discussed earlier.

4 Competition for the motivated

We now look at the polar case, where all agents have the same productivity θ (normalized to 0 without loss of generality) in task B but differ in their “ethical” motivations v for task A : a fraction p_L has $v = v_L$ and the remaining p_H have $v = v_H$.

When facing compensation scheme (y, z) , an agent of type v_i has net utility $u_i(y) + z$, where

$$u_i(y) \equiv \max_{(a,b)} \{v_i a + yb - C(a, b)\} \quad (40)$$

Let $a_i(y)$ and $b_i(y)$ denote the corresponding efforts, namely the solutions to the system $\{C_a(a, b) = v_i$ and $C_b(a, b) = y\}$, and note that $u'_i(y) = b$. Concavity of the cost function implies here again that $a'_i(y) < 0 < b'_i(y)$, as well as $a_L(y) < a_H(y)$ and $b_H(y) < b_L(y)$ in response to any given incentive rate $y > 0$.

The employer of an agent with type v_i makes net profit $\pi_i(y) - z$, where

$$\pi_i(y) \equiv Aa_i(y) + (B - y)b_i(y). \quad (41)$$

Finally, we denote

$$w_i(y) \equiv u_i(y) + \pi_i(y), \quad (42)$$

$$y_i^* = \arg \max \{w_i(y)\}. \quad (43)$$

In contrast to the case of heterogeneity in talent θ , there are now generally different optimal incentive rates for each type of worker.²¹

Next, we note that when confronted with an incentive-compatible menu of options, the more pro-social type (v_H) chooses a less powerful incentive scheme:²²

$$y_H \leq y_L \quad (\text{and so } z_H \geq z_L).$$

²¹In the quadratic-cost benchmark, $y_L^* = y_H^* = B - \gamma A$. In general, the variation of y with v involves the third derivatives of C and is thus ambiguous.

²²Adding up the two incentive constraints, $U_H \geq U_L + u_H(y_L) - u_L(y_L)$ and $U_L \geq U_H - u_H(y_H) + u_L(y_H)$, yields $0 \geq \int_{y_H}^{y_L} [u'_H(y) - u'_L(y)] dy = \int_{y_H}^{y_L} [b_H(y) - b_L(y)] dy$. Since $b_H(y) < b_L(y)$ for all y , the result follows.

This, in turn, implies that if a_L and a_H denote the two types' equilibrium efforts on task A , then

$$a_L \leq a_H.$$

The more pro-socially inclined employee exerts more effort on task A both because she is more motivated for it and because she chooses a lower-powered incentive scheme.

Monopsony. The monopsonist offers an incentive-compatible menu (y_L, z_L) and (y_H, z_H) , or equivalently (y_L, U_L) and (y_H, U_H) so as to solve:

$$\begin{aligned} & \max_{\{(y_i, z_i)\}_{i=H,L}} \left\{ \sum_{i=H,L} p_i [w_i(y_i) - U_i] \right\} \\ & \text{s.t.} \\ & U_L \geq \bar{U} \\ & U_H \geq U_L + u_H(y_L) - u_L(y_L), \\ & U_L \geq U_H + u_L(y_H) - u_H(y_H), \end{aligned}$$

The first two constraints must clearly be binding, while the third imposes $y_H \leq y_L$, as seen above. Substituting in, the solution satisfies

$$y_H^m = y_H^* \tag{44}$$

$$w'_L(y_L) = \frac{p_H}{p_L} [u'_H(y_L) - u'_L(y_L)] = \frac{p_H}{p_L} [b_H(y) - b_L(y)] \tag{45}$$

when it is interior; more generally, the left-hand side of (45) must be no greater than the right-hand side. Since the latter is strictly negative, one must have in any case

$$y_L^m > y_L^*. \tag{46}$$

The monopsonist offers a higher-powered incentive scheme than under symmetric information so as to limit the rent of the more prosocial types, who clearly benefits less from an increase in y than the less prosocial one.

Perfect competition. Because employees' intrinsic-motivation benefits va are private, firms have no reason to compete to select more prosocial types. As a result, the kind of incentive distortion seen earlier does not arise, and the competitive equilibrium is the symmetric-information outcome. Employers offer the menu $\{(y_i^*, z_i^*)\}_{i=H,L}$, where for each type y_i^* is the efficient incentive rate defined by (43) and $z_i^* \equiv \pi_i(y_i^*)$, leaving the firm with zero profit. Type $i = H, L$ then chooses

$$\max_{j \in \{H,L\}} \{u_i(y_j^*) + \pi_i(y_j^*) = w_i(y_j^*)\}. \tag{47}$$

By definition, $j = i$ is the optimal choice, so the symmetric-information outcome is indeed incentive compatible.

Proposition 9 *When agents are similar in measurable talent θ but differ in their ethical values v , monopsony leads to an overincentivization of low-motivation types, $y_L^m > y_L^*$ (with $y_H^m = y_H^*$), whereas competition leads to the first-best outcome, $y_L^c = y_L^*$, $y_H^c = y_H^*$.*

Would conclusions differ under an alternative specification of the impact of prosocial heterogeneity? Suppose that instead of enjoying task A more, a more prosocial agent supplies more unmeasured positive externalities on the firm (or on her coworkers, so that their productivity is higher, or their wages can be reduced due to a better work environment). In other words, agents $i = H, L$ share the same preferences,

$$u_i(y) = \max_{(a,b)} \{\bar{v}a + yb - C(a, b)\} \equiv u(y)$$

but have different productivities in the A activity,

$$\pi_i(y) = A(a(y) + \nu_i) + (B - y)b(y). \quad (48)$$

Under this formulation there is no way to screen an agent's type, so the outcome under both monopsony and competition is full pooling at the efficient incentive power:

$$y_i^* = y^* \equiv \arg \max \{Aa(y) + Bb(y) - C(a(y), b(y))\}. \quad (49)$$

Sorting will occur, on the other hand, when agents' intrinsic motivation is not unconditional, as we have assumed, but reciprocal –that is, dependent on the presence in the same firm of other people who act cooperatively (e.g., Kosfeld and von Siemens (2011)), or on the firm fulfilling a socially valuable mission rather than merely maximizing profits (e.g., Besley and Ghatak 2005, 2006, Brekke and Nyborg 2006). The fact that the benefits of “competing for the motivated” are somewhat attenuated in our model with respect to those only reinforces the contrast with the potentially very distortionary effects of competition for “talent”, thus further strengthening our main message.

5 Conclusion

To be completed.

Appendix A: Quadratic-Cost Case

Let the cost function be

$$C(a, b) = a^2/2 + b^2/2 + \gamma ab. \quad (\text{A.1})$$

with $\gamma^2 < 1$, ensuring convexity. The main case of interest is $\gamma > 0$ (efforts are substitutes), but all derivations and formulas hold with $\gamma < 0$ (complements) as well. The first-order conditions for (2) yield $v = a + \gamma b$, and $y = b + \gamma a$, hence

$$a(y) = \frac{v - \gamma y}{1 - \gamma^2}, \quad b(y) = \frac{y - \gamma v}{1 - \gamma^2}, \quad \text{and} \quad a'(y) = \frac{-\gamma}{1 - \gamma^2}, \quad b'(y) = \frac{1}{1 - \gamma^2}. \quad (\text{A.2})$$

Equations (10)-(11) then lead to

$$y^* = B - \gamma A, \quad (\text{A.3})$$

$$w(y^*) - w(y) = - \int_{y^*}^y w'(z) dz = - \int_{y^*}^y \left(\frac{y^* - z}{1 - \gamma^2} \right) dz = \frac{(y - y^*)^2}{2(1 - \gamma^2)}. \quad (\text{A.4})$$

1. *Monopsony.* Substituting the last two expressions into Proposition 1 yields

$$y_L^m = y^* - (1 - \gamma^2) \frac{q_H}{q_L} \Delta\theta, \quad (\text{A.5})$$

$$L^m = \frac{1}{2} \frac{q_H^2}{q_L} (1 - \gamma^2) (\Delta\theta)^2. \quad (\text{A.6})$$

2. *Perfect competition.* From (A.4) and (22), we get:

$$\frac{1}{2(1 - \gamma^2)} (y_H^c - y^*)^2 = (B - y_H^c) \Delta\theta. \quad (\text{A.7})$$

Let $\nu \equiv y_H^c - y^* = y_H^c - B + \gamma A > 0$ and $\omega \equiv (1 - \gamma^2) \Delta\theta$. Then $\nu^2 + 2\omega(\nu - \gamma A) = 0$ and solving this polynomial yields $\nu = -\omega + \sqrt{\omega^2 + 2\omega\gamma A} > 0$, or

$$y_H^c = B - \gamma A - \omega + \sqrt{\omega^2 + 2\omega\gamma A}. \quad (\text{A.8})$$

Note that $y_H^c < B$, since $\omega + \gamma A > \sqrt{\omega^2 + 2\omega\gamma A}$. The resulting efficiency loss relative to the social optimum is

$$L^c = q_H [w(y^*) - w(y_H^c)] = (B - y_H^c) q_H \Delta\theta = \left(\gamma A + \omega - \sqrt{\omega^2 + 2\omega\gamma A} \right) q_H \Delta\theta. \quad (\text{A.9})$$

Finally, the least-cost separating allocation is interim efficient if (23) holds, which here becomes

$$\frac{1}{1 - q_L} \geq \sqrt{1 + \frac{2\gamma A}{\omega}}. \quad (\text{A.10})$$

It will be useful to rewrite the condition as

$$\frac{\gamma A}{\omega} < \frac{1}{2} \left(\frac{q_L}{q_H} \right)^2 + \frac{q_L}{q_H}. \quad (\text{A.11})$$

3. *Welfare under monopsony versus competition.* Using (A.6) and (A.9), condition (26) becomes:

$$\left(\frac{q_H^2}{2q_L} \right) (1 - \gamma^2) (\Delta\theta)^2 < (B - y_H^c) q_H \Delta\theta \iff \frac{q_H}{2q_L} \omega < \gamma A - \nu \iff \nu < \gamma A - \frac{q_H}{2q_L} \omega.$$

Substituting into the polynomial whose positive root is ν , this is equivalent to:

$$\left(\gamma A - \frac{q_H}{2q_L} \omega \right)^2 > 2\omega \left(\frac{q_H}{2q_L} \omega \right) = \frac{q_H}{q_L} \omega^2,$$

which yields (27). This inequality and the interim efficiency condition (A.11) are simultaneously satisfied if and only if

$$\underline{M}(q_L) \equiv \frac{1 - q_L}{2q_L} + \sqrt{\frac{1 - q_L}{q_L}} < \frac{\gamma}{1 - \gamma^2} \frac{A}{\Delta\theta} \leq \frac{1}{2} \left(\frac{q_L}{1 - q_L} \right)^2 + \frac{q_L}{1 - q_L} \equiv \bar{M}(q_L). \quad (\text{A.12})$$

Note that:

(i) $\underline{M}(q_L) < \bar{M}(q_L)$ if and only if $q_H/q_L < 1$, so for any $q_L > 1/2$, (A.12) defines a nonempty range for $(A/\Delta\theta) [\gamma/(1 - \gamma^2)]$.

(ii) As $q_L \rightarrow 1$, $\underline{M}(q_L) \rightarrow 0$ and $\bar{M}(q_L) \rightarrow +\infty$, so arbitrary values of $(A/\Delta\theta) [\gamma/(1 - \gamma^2)]$ become feasible, including arbitrarily large values of A or arbitrarily low values of $\Delta\theta$. In particular, imposing $\gamma A < B(1 - \gamma^2) - q_H \Delta\theta/q_L$ to ensure $0 < y^* < y_L^m$ is never a problem for q_L large enough.

Imperfect competition. In Region I, $\hat{y}_H(t)$ is defined as the solution to (C.22) in Appendix C, which here becomes:

$$\begin{aligned} \frac{(y_H - y^*)^2}{2(1 - \gamma^2)} - (B - y_H) \Delta\theta + \frac{t}{q_L \Delta\theta} \frac{(y_H - y^*)}{(1 - \gamma^2)} &= 0 \iff \\ \nu^2 + 2(t/q_L \Delta\theta) \nu - 2\omega(\gamma A - \nu) &= \nu^2 + 2(\omega + t/q_L \Delta\theta) \nu - 2\omega \gamma A = 0, \end{aligned}$$

with the above definitions of ω and $\nu = y_H - y^*$. Solving, we have:

$$\hat{y}_H(t) = B - \gamma A - t/q_L \Delta\theta - \omega + \sqrt{(\omega + t/q_L \Delta\theta)^2 + 2\omega \gamma A}. \quad (\text{A.13})$$

It is easily verified that $\hat{y}_H(0) = y_H^c$ and

$$q_L \Delta\theta \cdot \hat{y}'_H(t) = -1 + \frac{\omega + t/q_L \Delta\theta}{\sqrt{(\omega + t/q_L \Delta\theta)^2 + 2\omega \gamma A}} < 0.$$

Moreover, this expression is increasing in t/q_L , so $\hat{y}_H(t)$ is decreasing and convex over Region I.

In Region II, $\hat{y}_H(t)$ is defined as the solution to (C.25) in Appendix 5, which here becomes:

$$\begin{aligned}
w(y^*) + \theta_L B - \frac{(y_H - y^*)^2}{2(1 - \gamma^2)} + (B - y_H)\Delta\theta - \frac{t}{\Delta\theta} \frac{(y_H - y^*)}{(1 - \gamma^2)} - \bar{U} - t &= 0 \iff \\
\frac{\nu^2}{2(1 - \gamma^2)} + (\nu - \gamma A)\Delta\theta + \frac{t}{\Delta\theta} \frac{\nu}{(1 - \gamma^2)} + \theta_L B + \bar{U} + t - w(y^*) &= 0 \iff \\
\nu^2 + 2(\nu - \gamma A)\omega + \frac{2t\nu}{\Delta\theta} + 2(1 - \gamma^2) (\bar{U} + t + \theta_L B - w(y^*)) &= 0 \iff \\
\nu^2 + 2\nu(\omega + t/\Delta\theta) - 2\gamma A\omega + 2(1 - \gamma^2) (\bar{U} + t + \theta_L B - w(y^*)) &= 0.
\end{aligned}$$

Solving, we have:

$$\hat{y}_H(t) = B - \gamma A - t/\Delta\theta - \omega + \sqrt{(\omega + t/\Delta\theta)^2 + 2[\omega\gamma A + (1 - \gamma^2) (\bar{U} + t + \theta_L B - w(y^*))]}. \quad (\text{A.14})$$

Moreover,

$$\hat{y}'_H(t) = -1 + \frac{\omega + t/\Delta\theta}{\sqrt{(\omega + t/\Delta\theta)^2 + 2[\omega\gamma A + (1 - \gamma^2) (\bar{U} + t + \theta_L B - w(y^*))]}} < 0$$

and it is increasing in $t/\Delta\theta$, so $\hat{y}_H(t)$ is decreasing and convex over Region II.

In Region III, $\hat{y}_L(t)$ is defined as the solution to (C.29). Denoting $\nu = y_L - y^*$, this now becomes:

$$\begin{aligned}
w(y^*) + \theta_L B - \bar{U} - t + (\gamma A - \nu)\Delta\theta - \frac{tq_L}{q_H\Delta\theta} \frac{\nu}{(1 - \gamma^2)} &= 0 \iff \\
(1 - \gamma^2) [w(y^*) + \theta_L B - \bar{U} - t + \gamma A\Delta\theta] &= \nu \left(\omega + \frac{tq_L}{q_H\Delta\theta} \right).
\end{aligned}$$

Solving, we have:

$$\hat{y}_L(t) = B - \gamma A - \frac{(1 - \gamma^2) [\bar{U} + t - w(y^*) - \theta_L B] - \omega\gamma A}{\omega + tq_L/q_H\Delta\theta}. \quad (\text{A.15})$$

It is easily verified that $\hat{y}_L(+\infty) = B - \gamma A - (1 - \gamma^2)q_H\Delta\theta/q_L = y_L^m$. Moreover, recalling (12),

$$\hat{y}'_H(t) = \frac{(q_L/q_H\Delta\theta) [\bar{U} - w(y^*) - \theta_L B - \omega\gamma A\Delta\theta] - \omega}{(\omega + tq_L/q_H\Delta\theta)^2/(1 - \gamma^2)} < 0$$

and this function is increasing in $t/\Delta\theta$, implying that $\hat{y}_L(t)$ is convex. ■

Appendix B: Nonlinear Contracts

We allow here for general reward functions $Y(r)$, where $r \equiv b + \theta$ and Br is the employer's revenue on the verifiable task. As it will be more convenient to work directly with effort b rather than the marginal incentive $y(b) \equiv Y'(b)$, let us define the pseudo-cost function

$$\hat{C}(b) \equiv \min_a \{C(a, b) - va\}, \quad (\text{B.1})$$

and denote $\hat{a}(b)$ the minimizing choice. It is easily verified that $\hat{a}'(b) < 0$ when $C_{ab} < 0$ and that $\hat{C}(b)$ is strictly convex. To preclude unbounded solutions, we shall assume that $\lim_{b \rightarrow +\infty} \hat{C}'(b) \geq B$. The allocative component of total surplus is

$$\hat{w}(b) \equiv A\hat{a}(b) + Bb - \hat{C}(b), \quad (\text{B.2})$$

which shall take to be strictly quasiconvex and maximized at $b^* > 0$. The effort levels b_i chosen by each type $i = H, L$ are given by $\hat{C}'(b_i) = Y'(b_i + \theta_i)$, with resulting utilities are $U_i = Y(b_i + \theta_i) - \hat{C}(b_i)$, so the relevant participation and incentive constraints are now:

$$U_L = Y(b_L + \theta_L) - \hat{C}(b_L) \geq \bar{U}, \quad (\text{B.3})$$

$$U_H \geq U_L + \hat{C}(b_L) - \hat{C}(b_L - \Delta\theta), \quad (\text{B.4})$$

$$U_L \geq U_H - \hat{C}(b_H + \Delta\theta) + \hat{C}(b_H). \quad (\text{B.5})$$

Summing up the last two yields $\hat{C}(b_L) - \hat{C}(b_L - \Delta\theta) \leq \hat{C}(b_H + \Delta\theta) - \hat{C}(b_H)$, which by strict convexity of \hat{C} requires that $r_L \equiv b_L + \theta_L < b_H + \theta_H \equiv r_H$.

1. *Monopsony.* As before, the low types' participation and high type's incentive constraints must be binding in an optimum, so the firm solves

$$\max_{\{b_H, b_L\}} \left\{ \sum_{i=H,L} q_i [\hat{w}(b_i) + B\theta_i - \bar{U}] - q_H [\hat{C}(b_L) - \hat{C}(b_L - \Delta\theta)] \right\},$$

leading to $b_H^m = b^*$ and

$$\hat{w}'(b_L^m) = \frac{q_H}{q_L} \left[\hat{C}'(b_L^m) - \hat{C}'(b_L^m - \Delta\theta) \right] \quad (\text{B.6})$$

when the solution is interior and to $b_L^m = \Delta\theta$ when it is not; in either case, $b_H^m < b^*$.

As before, the firm is willing to employ the low types provided the profits they generate exceed the rents abandoned to the high types:

$$q_L [\hat{w}(b_L^m) + B\theta_L - \bar{U}] \geq q_H [\hat{C}(b_L) - \hat{C}(b_L - \Delta\theta)], \quad (\text{B.7})$$

which defines a lower bound $\underline{q}_L < 1$ for q_L . Compared with the case of linear incentives, the greater flexibility afforded by nonlinear schemes allows the monopsonist to reduce the high type's rents,

$$U_H = \bar{U} + \hat{C}(b_L^m) - \hat{C}(b_L^m - \Delta\theta) < \bar{U} + \hat{C}'(b_L^m)\Delta\theta \equiv \bar{U} + y_L^m \Delta\theta \quad (\text{B.8})$$

while keeping the distortion –underincentivization of the low types– unchanged, thereby increasing his profits.

2. *Perfect competition.* In a separating competitive equilibrium (with no cross-subsidies), $U_i = w(b_i) + \theta_i B$ for $i = L, H$, and by the same argument as in Section 3.2 the low type must get his efficient symmetric-information allocation, so $b_L^c = b^*$. Furthermore, among all such allocations that are incentive-compatible, the most attractive to high types is the LCS one, defined by:

$$b_H \equiv \arg \max_b \left\{ \hat{w}(b) \mid \hat{w}(b^*) - \hat{w}(b) \geq B\Delta\theta - \hat{C}(b + \Delta\theta) + \hat{C}(b) \right\}. \quad (\text{B.9})$$

Denoting the Lagrange multiplier as $\lambda \geq 0$, the first-order condition is $-\hat{w}'(b)(1 - \lambda) = \lambda[\hat{C}'(b + \Delta\theta) - \hat{C}'(b)]$, requiring that $b \geq b^*$ and leading to two cases:

- (i) If $\hat{C}(b^* + \Delta\theta) - \hat{C}(b^*) \geq B\Delta\theta$, then $b = b^*$.
- (ii) If $\hat{C}(b^* + \Delta\theta) - \hat{C}(b^*) < B\Delta\theta$, then b_H is uniquely given by

$$\hat{w}(b^*) - \hat{w}(b_H^c) = B\Delta\theta - \hat{C}(b_H^c + \Delta\theta) + \hat{C}(b_H^c), \quad (\text{B.10})$$

since $\Omega(b) \equiv \hat{w}(b^*) - \hat{w}(b) + \hat{C}(b + \Delta\theta) - \hat{C}(b)$ is strictly increasing on $(b^*, +\infty)$, with $\Omega(b^*) < B\Delta\theta$ and $\Omega(b) \geq \hat{C}'(b) > B\Delta\theta$ for all b large enough. Competition now leads to overincentivization only if $\hat{C}(b^* + \Delta\theta) - \hat{C}(b^*)$ is not too large, meaning that b^* itself is not too large, which in turn occurs when A is large enough.²³ Whereas the greater flexibility afforded by nonlinear contracts allows a monopsonist to reduce the high types' rents and increase his profits, in a competitive industry those benefits are appropriated by the high types, in the form of more efficient contracts:

$$\hat{w}(b^*) - \hat{w}(b_H^c) \leq B\Delta\theta - \hat{C}(b_H^c + \Delta\theta) + \hat{C}(b_H^c) < B\Delta\theta - \hat{C}'(b_H^c)\Delta\theta \equiv (B - b_H^c)\Delta\theta. \quad (\text{B.11})$$

As before, the above separating allocation is the equilibrium if and only if it is interim efficient, meaning that there is no profitable deviation consisting of lowering b_H by a small amount $\delta b_H = -\varepsilon$ to increase total surplus while offering compensating transfers $\delta Y_L = \hat{C}'(b_H + \Delta\theta) - \hat{C}'(b_H)$ to low types so as to maintain incentive compatibility (note that $\delta U_H = [Y'(b_H + \theta_H) - \hat{C}(b_H)]\varepsilon = 0$ to the first order). In other words,

$$q_H \hat{w}'_H(b_H) + q_L \left[\hat{C}'(b_H + \Delta\theta) - \hat{C}'(b_H) \right] \geq 0, \quad (\text{B.12})$$

which defines a lower bound $\tilde{q}_L < 1$ for q_L . We denote again $q_L^* \equiv \max\{ \underline{q}_L, \tilde{q}_L \}$.

The following proposition summarizes the above results.

Proposition 10 *Let $q_L > q_L^*$. With unrestricted nonlinear contracts, it remains the case that:*

1. *A monopsonist distorts downward the measurable effort of low types: $b_L^m < b^* = b_H^m$, with b_L^m given by (B.6).*

²³Note that $\hat{C}(b^* + \Delta\theta) - \hat{C}(b^*) < \hat{C}'(b^* + \Delta\theta)\Delta\theta$, so $\hat{C}(b^* + \Delta\theta) - \hat{C}(b^*) < B\Delta\theta$ if $b^* < \hat{C}'^{-1}(B) - \Delta\theta$. This, in turn, occurs when $A\hat{a}'[\hat{C}'^{-1}(B) - \Delta\theta] + B < \hat{C}'[\hat{C}'^{-1}(B) - \Delta\theta]$, for which it suffices that a be large enough.

2. Under perfect competition, if $\hat{C}(b^* + \Delta\theta) - \hat{C}(b^*) < B\Delta\theta$, firms distort upwards the measurable effort of high types: $b_L^c = b^* < b_H^c$, with b_H^c given by (B.10). If $\hat{C}(b^* + \Delta\theta) - \hat{C}(b^*) \geq B\Delta\theta$, both effort levels are efficient: $b_L^c = b^* = b_H^c$.

For small $\Delta\theta$, Taylor expansions show that: (i) $b^* - b_L^m$ is, as before, of order $\Delta\theta$, with coefficient q_H/q_L ; (ii) under competition, since $\hat{C}'(b^*) < B\Delta\theta$, $b_H^c - b^*$ is positive and again of order $\sqrt{\Delta\theta}$, with coefficient independent of q_H and q_L ; (iii) $1 - \tilde{q}_L$ is also of order $\Delta\theta$. Therefore, the ratio L^c/L^m is at least of the same order as $q_H\Delta\theta/q_L [(q_H/q_L)\Delta\theta]^2 = (q_L/q_H)\Delta\theta$, or equivalently $(\Delta\theta)^{-3/2}$, and hence arbitrarily large.

3. *Quadratic-cost case.* Here again, the cost function (A.1) leads to explicit and transparent solutions, including for the comparison of efficiency losses under monopsony and competition. Minimizing $C(a, b) - va$ over a leads to $\hat{a}(b) = v - \gamma b$ and

$$\hat{C}(b) = \frac{1}{2}[b^2 - (v - \gamma b)^2] = \frac{1}{2}[(1 - \gamma^2)b^2 + 2v\gamma b - v^2]. \quad (\text{B.13})$$

Therefore $\hat{C}'(b) = (1 - \gamma^2)b + v\gamma$ and $\hat{w}'(b) = B - \gamma(A + v) - (1 - \gamma^2)b$, leading to

$$(1 - \gamma^2)b^* = B - \gamma(A + v), \quad (\text{B.14})$$

$$\begin{aligned} \hat{w}(b^*) - \hat{w}(b) &= -w''(b^*) \frac{(b - b^*)^2}{2} = \frac{1 - \gamma^2}{2}(b - b^*)^2, \\ \hat{C}(b) - \hat{C}(b^*) &= \hat{w}(b^*) - \hat{w}(b) + (B - \gamma A)(b - b^*) \\ &= \frac{1 - \gamma^2}{2}(b - b^*)^2 + (B - \gamma A)(b - b^*). \end{aligned} \quad (\text{B.15})$$

Under *monopsony*, b_L^m is defined (when interior) by (B.6), which now becomes:

$$\begin{aligned} B - \gamma(A + v) - (1 - \gamma^2)b_L^m &= \frac{q_H}{q_L}(1 - \gamma^2)\Delta \iff \\ b_L^m &= b^* - \frac{q_H}{q_L}\Delta. \end{aligned} \quad (\text{B.16})$$

This is the same outcome as with linear contracts, and the marginal incentives $y_i \equiv Y'(b_i + \theta_i) = C'(b_i)$ are also unchanged. The rent left to high types is now lower, however, implied by (B.8).

Under *competition*, b_H^c is given by (B.9). Now, for all b , $\hat{w}(b^*) - \hat{w}(b) \geq B\Delta\theta - \hat{C}(b + \Delta\theta) + \hat{C}(b_H)$ if and only if

$$\begin{aligned} B\Delta\theta - [\hat{w}(b^*) - \hat{w}(b)] &\leq [\hat{C}(b_H + \Delta\theta) - C(b^*)] - [\hat{C}(b) - C(b^*)] \iff \\ B\Delta\theta - \frac{1 - \gamma^2}{2}(b - b^*)^2 &= \frac{1 - \gamma^2}{2}(b + \Delta\theta - b^*)^2 - \frac{1 - \gamma^2}{2}(b - b^*)^2 + (B - \gamma A)\Delta\theta \iff \\ \frac{2\gamma A\Delta\theta}{1 - \gamma^2} &\leq \Delta\theta [2(b - b^*) + \Delta\theta] + (b - b^*)^2 = (b - b^* + \Delta\theta)^2 \end{aligned}$$

Therefore (B.9) becomes

$$b_H^c = b^* + \max \left\{ 0, -\Delta\theta + \sqrt{\frac{2\gamma A \Delta\theta}{1-\gamma^2}} \right\} \quad (\text{B.17})$$

or, in terms of marginal incentives, $y_H^c = y^* + \max \{0, -\omega + \sqrt{2\omega\gamma A}\}$. Comparing these expressions to (A.8), it is clear that competition leads to less distortion when nonlinear contracts are allowed (possibly none, for low A), whereas the monopsony distortion remains the same. Nonetheless, when A is high enough, or $\Delta\theta$ small enough, competition can still be efficiency-reducing:

Proposition 11 *Let $C(a, b) = a^2/2 + b^2/2 + \gamma ab$. When employers can offer arbitrary nonlinear contracts, social welfare is lower under competition than under monopsony if and only if q_L is high enough and*

$$\frac{1 + q_H}{2q_L} < \left(\frac{\gamma}{1-\gamma^2} \right) \left(\frac{A}{\Delta\theta} \right). \quad (\text{B.18})$$

Proof. We have

$$\begin{aligned} L^m &= \frac{q_L (b^* - b_L^m)^2}{2(1-\gamma^2)} < \frac{q_H (b_H^c - b^*)^2}{2(1-\gamma^2)} = L^c \iff \\ \sqrt{\frac{q_L}{q_H} \frac{q_H}{q_L} \Delta\theta} &< \sqrt{\frac{2\gamma A \Delta\theta}{1-\gamma^2}} - \Delta\theta \iff \frac{1}{2} \left(1 + \sqrt{\frac{q_H}{q_L}} \right)^2 < \frac{\gamma}{1-\gamma^2} \frac{A}{\Delta\theta} \iff \\ \hat{M}(q_L) &\equiv \frac{1 + q_H}{2q_L} < \frac{\gamma}{1-\gamma^2} \frac{A}{\Delta\theta} \end{aligned}$$

Thus, together with the interim-efficiency condition, which remains unchanged, we require:

$$\hat{M}(q_L) \equiv \frac{2 - q_L}{2q_L} < \frac{\gamma}{1-\gamma^2} \frac{A}{\Delta\theta} \leq \frac{1}{2} \left(\frac{q_L}{1 - q_L} \right)^2 + \frac{q_L}{1 - q_L} \equiv \bar{M}(q_L).$$

While the upper bound $\bar{M}(q_L)$ is unchanged from (A.12), the lower bound is higher than in the linear-contracts case, $\hat{M}(q_L) > \underline{M}(q_L)$, since

$$\frac{1 - q_L}{2q_L} + \sqrt{\frac{1 - q_L}{q_L}} < \frac{1 - q_L}{2q_L} + \frac{1}{2q_L}.$$

Nonetheless, it remains the case that:

(i) $\underline{M}(q_L) < \bar{M}(q_L)$ if and only if $q_H/q_L < 1$, so for any $q_L > 1/2$, (A.12) defines a nonempty range for $(A/\Delta\theta) [\gamma/(1-\gamma^2)]$.

(ii) As $q_L \rightarrow 1$, and $\bar{M}(q_L) \rightarrow +\infty$ so arbitrary large values of $(A/\Delta\theta) [\gamma/(1-\gamma^2)]$ become feasible, including arbitrarily large values of A or arbitrarily low values of $\Delta\theta$. In particular, imposing $\gamma A < B(1-\gamma^2) - q_H \Delta\theta/q_L$ to ensure $0 < y^* < y_L^m$ is never a problem for q_L large enough. ■

Appendix C: Main Proofs

Proof of Proposition 1. Only the comparative-statics results remains to prove. First, differentiating (10) and (17) with respect to $\Delta\theta$ yields

$$\frac{\partial y_L^m}{\partial \Delta\theta} = \frac{q_H}{q_L} \frac{1}{w_{yy}(y_L^m)} < 0 < q_H \frac{w_y(y_L^m)}{-w_{yy}(y_L^m)} = \frac{\partial L^m}{\partial \Delta\theta}. \quad (\text{C.1})$$

Turning next to A , we have

$$\frac{\partial y_L^m}{\partial A} = \frac{w_{yA}(y_L^m)}{-w_{yy}(y_L^m)} = \frac{a'(y_L^m)}{-w_{yy}(y_L^m)} < 0 \quad (\text{C.2})$$

$$\begin{aligned} \frac{1}{q_L} \frac{\partial L^m}{\partial A} &= w_A(y^*; A, B) - w_A(y_L^m; A, B) + w_y(y^*; A, B) \frac{\partial y^*}{\partial A} - w_y(y_L^m; A, B) \frac{\partial y_L^m}{\partial A} \\ &= a(y^*) - a(y_L^m) - \frac{q_H}{q_L} \Delta\theta \frac{\partial y_L^m}{\partial A}, \end{aligned} \quad (\text{C.3})$$

showing clearly the two opposing effects discussed in the text, and which exactly cancel out in the quadratic-cost case (see (A.5)). Similarly, for B :

$$\frac{\partial y_L^m}{\partial B} = \frac{w_{yB}(y_L^m)}{-w_{yy}(y_L^m)} = \frac{b'(y_L^m)}{-w_{yy}(y_L^m)} > 0 \quad (\text{C.4})$$

$$\begin{aligned} \frac{1}{q_L} \frac{\partial L^m}{\partial B} &= w_B(y^*; A, B) - w_B(y_L^m; A, B) + w_y(y^*; A, B) \frac{\partial y^*}{\partial B} - w_y(y_L^m; A, B) \frac{\partial y_L^m}{\partial B} \\ &= b(y^*) - b(y_L^m) - \frac{q_H}{q_L} \Delta\theta \frac{\partial y_L^m}{\partial B}, \end{aligned} \quad (\text{C.5})$$

showing again two offsetting effects on L^m . ■

Proof of Lemma 1. Denote U_L^c and U_H^c the two types' utilities in the LCS allocation, and recall that the former takes the same value as under symmetric information.

Claim 1 *The LCS allocation is interim efficient if and only if it solves the following program*

$$\begin{aligned} (\mathcal{P}) : \quad & \max_{(U_L, U_H, y_H, y_L)} \{U_H\}, \quad \text{subject to:} \\ & U_L \geq U_L^c = w(y^*) + \theta_L B \end{aligned} \quad (\text{C.6})$$

$$U_L \geq U_H - y_H \Delta\theta \quad (\text{C.7})$$

$$U_H \geq U_L + y_L \Delta\theta \quad (\text{C.8})$$

$$0 \leq \sum_{i=H,L} q_i [w(y_i) + B\theta_i - U_i], \quad (\text{C.9})$$

Proof. Conditions (C.7) and (C.8) are the incentive constraints for types L and H respectively, and condition (C.9) the employers' interim break-even constraint. Now, note that:

- (i) If the LCS allocation does not achieve the optimum, interim efficiency clearly fails.

(ii) If the LCS allocation solves (\mathcal{P}) , there can clearly be no incentive-compatible Pareto improvement in which $U_H > U_H^c$, but neither can there be one in which $U_H = U_H^c$ and $U_L > U_L^c$. Indeed, if there was such an allocation, note first that one could without loss of generality take it to satisfy $y_H \geq y^*$; otherwise, replacing y_H by y^* while keeping U_H and U_L unchanged strictly increases profits, so the LCS allocation remains (even more) dominated. Starting from such an allocation with $y_H \geq y^*$, let us now reduce U_L by some small $\eta > 0$ and increase U_H by some small ε , while also increasing y_H by $(\varepsilon + \eta) \Delta\theta$ to leave (C.6) unchanged (while (C.8) is only strengthened). This results in extra profits of $q_H [w'(y_H) (\varepsilon + \eta) \Delta\theta - \varepsilon] + q_L \eta$, which is positive as long as

$$\eta > \frac{q_H [1 - w'(y_H) \Delta\theta] \varepsilon}{q_L + q_H w'(y_H) \Delta\theta}.$$

Both types and the firm are now strictly better off than in the LCS allocation, contradicting the fact that it is a solution to (\mathcal{P}) . ■

To study interim efficiency and prove Lemma 1, let us therefore analyze the solution(s) to (\mathcal{P}) . First, condition (C.9) must be binding, otherwise U_L and U_H could be increased by the same small amount without violating the other constraints. Second, (C.7) must also be binding. Indeed, solving (\mathcal{P}) without that constraint leads to $U_L = U_L^c$ and $y_H = y^*$; maximizing U_H subject to (C.9) then leads to $y_L = y^*$ and $U_L = w(y^*) + B\theta_H$, so (C.6) holds with equality, a contradiction. Third, (C.8) now reduces to $y_H \geq y_L$. Two cases can then arise:

(i) If (C.6) is binding, the triple (U_L, U_H, y_H) is uniquely given by the same three equality constraints as the LCS allocation, and thus coincides with it.

(ii) If (C.6) is not binding, the solution to (\mathcal{P}) is the same as when that constraint is dropped. Substituting (C.7) into (C.9), and both being equalities, we have $U_H = \Sigma q_i [w(y_i) + B\theta_i] + q_L y_H \Delta\theta$, so (\mathcal{P}) reduces to

$$\max_{y_H, y_L} \{q_H w(y_H) + q_L [w(y_L) + y_H \Delta\theta] \mid y_H \geq y_L\}. \quad (\text{C.10})$$

For all $x \geq 0$, define the function $\tilde{y}(x) \equiv \arg \max_y \{w(y) + xy\}$ and let $\bar{x} \equiv -w'(B)$. On the interval $[0, \bar{x}]$ the function \tilde{y} is given by $w'(\tilde{y}(x)) = -x$, so it is strictly increasing up to $\tilde{y}(\bar{x}) = B$, while for $x > \bar{x}$, $\tilde{y}(x) \geq B$. Furthermore, it is clear that $\tilde{y}(x) \geq y^*$ with equality only at $x = 0$, so the pair $(y_H = \tilde{y}(q_L \Delta\theta / q_H), y_L = y^*)$ is the solution to (C.10). It is then indeed the case that (C.6) is non-binding, $U_L > U_L^c = U_H - y_H \Delta\theta$, if and only if

$$w(y^*) + \theta_L y^* < q_H [w(y_H) + B\theta_H] + q_L [w(y^*) + B\theta_L] - q_H y_H \Delta\theta$$

or equivalently $H(q_L \Delta\theta / q_H) \geq 0$, where

$$H(x) \equiv w(\tilde{y}(x)) - w(y^*) + [B - \tilde{y}(x)] \Delta\theta \geq 0. \quad (\text{C.11})$$

Note that $H(0) > 0$ and $H(x) < 0$ for if $x \geq \bar{x}$, while on $[0, \bar{x}]$ we have $H'(x) = [w'(\tilde{y}(x)) - \Delta\theta] \tilde{y}'(x) = (q_L / q_H - 1) \Delta\theta \tilde{y}'(x) = -(\Delta\theta / q_H) \tilde{y}'(x)$. Therefore, there exists a unique $\tilde{x} \in (0, \bar{x})$ such that (C.6) is non-binding –and the solution to (\mathcal{P}) thus differs from the LCS allocation– if and only

if $\Delta\theta q_L/q_H < \tilde{x}$. Equivalently, the LCS allocation is the unique solution to (\mathcal{P}) , and therefore interim efficient, if and only if $q_L/(1 - q_L) \geq \tilde{x}/\Delta\theta \equiv \tilde{q}_L/(1 - \tilde{q}_L)$, hence the result. For small $\Delta\theta$ it is easily verified from $H(\tilde{x}) \equiv 0$ and $w'(\tilde{y}(x)) = -x$ (implying $\tilde{y}'(x) = 1/w''(x)$) that $-\tilde{x}^2/w''(y^*) \approx 2(B - y^*) \Delta\theta$, so that $\tilde{q}_L \approx 1 - \chi\sqrt{\Delta\theta}$, where $\chi \equiv \sqrt{-2w''(y^*)(y^* - B)} > 0$. ■

Proof of Proposition 2. To complete the proof Results (1) and (2), it just remains to show that: (a) If the LCS allocation is interim efficient, it is a competitive equilibrium; (b) It is then the unique one. We shall also prove here that: (c) If the LCS allocation is not interim efficient, there exists no competitive equilibrium in pure strategies.

Claim 2 *In any competitive equilibrium, the utilities (U_L, U_H) must satisfy*

$$U_L \geq U_L^c = U_L^{SI} \equiv w(y^*) + \theta_L B, \quad (\text{C.12})$$

$$U_H \geq U_H^c \equiv w(y_H^c) + \theta_H B, \quad (\text{C.13})$$

Proof. If $U_L < U_L^c$, a firm could offer the single contract $(y = y^*, z = z_L^c - \varepsilon)$ for ε small, attracting and making a profit ε on type θ_L (and perhaps also attracting the more profitable type θ_H). Similarly, if $U_H < U_H^c$, a firm could offer the incentive-compatible menu $\{(y^*, z_L^c - \varepsilon), (y_H^c, z_H^c - \varepsilon)\}$ thereby attracting and making a profit ε on type θ_H (and perhaps also attracting and making zero profit on type θ_L). ■

Claim 3 *If an allocation Pareto dominates (in the interim-efficiency sense) the least-cost separating one, it must involve a cross-subsidy from high to low types, meaning that*

$$w(y_H) + B\theta_H - U_H > 0 > w(y_L) + B\theta_L - U_L, \quad (\text{C.14})$$

Proof. If $U_L \leq w(y_L) + \theta_L B$, then $U_L \geq U_L^c$ requires that $y = y^*$ and $U_L = w(y_L) + \theta_L B$; the break-even condition then requires that $w(y_H) + \theta_H B \geq U_H$, and Pareto-dominance that $U_H > U_H^c$. Therefore $w(y_H) + \theta_H B > U_H^c = U_L + y_H \Delta\theta = w(y^*) + B\theta_L + y_H \Delta\theta$, or finally $w(y_H) < w(y^*) - (B - y_H) \Delta\theta$. By (22) this means $y_H > y_H^c$, but then $w(y_H) + \theta_H B < w(y_H^c) + \theta_H B = U_H^c$, a contradiction. Thus, $w(y_L) + B\theta_L - U_L < 0$, meaning that low types get more than the total surplus they generate. For the employer to break even, it must be that high types get strictly less, $w(y_H) + B\theta_H - U_H > 0$. ■

We are now ready to establish the properties (a)-(c) listed above, and thereby complete the proof of Results (1) and (2) in Proposition 2.

(a) Suppose that the LCS allocation, defined by (18)-(22), is offered by all firms. Could another one come in and offer a different set of contracts, leading to new utilities (U_L, U_H) and a strictly positive profit? First, note that we can without loss of generality assume that $U_L \geq U_L^c$: if $U_L < U_L^c$ and U_H is indeed selected (with positive probability) by type H , then (U_L^c, U_H) is incentive-compatible. By offering U_L^c to L types (via their symmetric-information allocation), the deviating firm does not alter its profitability. Second, if $U_H < U_H^c$, the deviating employer does

not attract type H ; since it cannot make money on type L while providing $U_L \geq U_L^c$, the deviation is not profitable. Finally, suppose that $U_L \geq U_L^c$ and $U_H \geq U_H^c$. If at least one inequality is strict, then interim efficiency of the LCS allocation implies that the deviating firm loses money. If both are equalities, let us specify (for instance) that both types workers, being indifferent, do not select the deviating firm.

(b) By (C.12)-(C.13), in any equilibrium both types must be no worse off than in the LCS allocation, and similarly for the firm, which must make non-negative profits. If any of these inequalities is strict there is Pareto dominance, so when the LCS allocation is interim efficient, they must all be equalities, giving the LCS allocation as the unique solution.

(c) Suppose now that LCS allocation is not interim efficient. The contract that solves (\mathcal{P}) is then such that $y_H = \tilde{y}(x)$, $y_L = y^*$, $U_L > U_L^c$ (equation (C.6) is not binding) and $U_H > U_H^c$ (since the LCS does not solve (\mathcal{P})). A firm can then offer a contract with the same y_H and y_L but reducing both U_H and U_L by the same small amount, resulting in positive profits; the LCS allocation is thus not an equilibrium. Suppose now that some other allocation, with utilities U_L and U_H , is an equilibrium. As seen in (b), it would have to Pareto-dominate the LCS allocation, which by Claim 3 implies:

$$\begin{aligned} U_L &\geq U_H - y_H \Delta\theta, \\ w(y_H) + B\theta_H - U_H &> 0. \end{aligned}$$

Consider now a deviating employer offering a single contract, aimed at the high type: $y'_H = y_H + \varepsilon$ and $U'_H = U_H + (\varepsilon\Delta\theta)/2 < w(y'_H) + B\theta_H$. The low type does not take it up, as it would yield $U_L = U_L^c - (\varepsilon\Delta\theta)/2$. The high type clearly does, leading to a positive profit for the deviating employer. \parallel

The only part of Proposition 2 remaining to prove are the comparative statics. Differentiating (22) and (24) with respect to $\Delta\theta$ yields

$$\frac{\partial y_H^c}{\partial \Delta\theta} = \frac{B - y_H^c}{\Delta\theta - w_y(y_H^c; A, B)} > 0, \quad \frac{\partial L^c}{\partial \Delta\theta} = -q_H w_y(y_H^c; A, B) \frac{\partial y_H^c}{\partial \Delta\theta} > 0. \quad (\text{C.15})$$

Turning next to A ,

$$\begin{aligned} -\Delta\theta \frac{\partial y_H^c}{\partial A} &= w_A(y^*; A, B) - w_A(y_H^c; A, B) + w_y(y^*; A, B) \frac{\partial y^*}{\partial A} - w_y(y_H^c; A, B) \frac{\partial y_H^c}{\partial A} \\ &= a(y^*) - a(y_H^c) - w_y(y_H^c; A, B) \frac{\partial y_H^c}{\partial A} \Rightarrow \\ \frac{\partial y_H^c}{\partial A} &= \frac{a(y_H^c) - a(y^*)}{\Delta\theta - w_y(y_H^c; A, B)} < 0 < -q_H \Delta \frac{\partial y_H^c}{\partial A} = \frac{\partial L^c}{\partial A}. \end{aligned} \quad (\text{C.16})$$

Again there is a direct and an indirect effect of A on L^c , but now the direct one always dominates. For B , in contrast, the ambiguity remains

$$\begin{aligned}\Delta\theta - \Delta\theta \frac{\partial y_H^c}{\partial B} &= w_B(y^*; A, B) - w_B(y_H^c; A, B) + w_y(y^*; A, B) \frac{\partial y^*}{\partial B} - w_y(y_H^c; A, B) \frac{\partial y_H^c}{\partial B} \\ &= b(y^*) - b(y_H^c) - w_y(y_H^c; A, B) \frac{\partial y_H^c}{\partial B} \Rightarrow \\ \frac{\partial y_H^c}{\partial B} &= \frac{\Delta\theta + b(y_H^c) - b(y^*)}{\Delta\theta - w_y(y_H^c; A, B)} > 0\end{aligned}\tag{C.17}$$

$$\frac{1}{q_H \Delta\theta} \frac{\partial L^c}{\partial B} = 1 - \frac{\partial y_H^c}{\partial B} = \frac{-w_y(y_H^c; A, B) - b(y_H^c) + b(y^*)}{\Delta\theta - w_y(y_H^c; A, B)}.\tag{C.18}$$

In the quadratic case, L^c is independent of B and the last term thus equal to zero; see (A.9). ■

Proof of Proposition 4. We solve for the symmetric equilibrium under the assumption that market shares are always interior, and thus given by (31). In Appendix D [[to be completed]] we verify that individual deviations to corner solutions (one firm grabbing the whole market for some worker type, or on the contrary dropping them altogether) can indeed be excluded.

To characterize the symmetric solution to (32)-(35), we distinguish three regions.

Region I. Suppose first that the low type's individual rationality constraint is not binding, $U_L > \bar{U}$, so that $\nu = 0$.

Lemma 2 *If $\nu = 0$, then $\mu_H = 0 \leq \mu_L$ and $y_L = y^* \leq y_H$.*

Proof. (i) If $\mu_H = \mu_L = 0$, then $y_H = y_L = y^*$ by (38)-(39), so (33)-(34) imply that $U_H - U_L = y^* \Delta\theta$. Next, from (36)-(37) we have $m_H - m_L = t - m_L = 0$, whereas $m_H - m_L \equiv B \Delta\theta - (U_H - U_L) = (B - y^*) \Delta\theta > 0$, a contradiction.

(ii) If $\mu_H > 0 = \mu_L$ condition (39) implies $w'(y_L) > 0$, hence $y_L < y^*$, and condition (38) $y_H = y^*$. Moreover, (36)-(37) and $\mu_H > \mu_L$ require that $m_H < t < m_L$. However,

$$m_H - m_L = w(y^*) - w(y_L) + (B - y_L) \Delta\theta > 0,$$

a contradiction. We are thus left with $\mu_H = 0 < \mu_L$, which implies $y_L = y^* < y_H$ by (39) and (38) respectively. ■

Let us now derive and characterize y_H as a function of t . We can rewrite (38) as

$$t q_H w'(y_H) = -\mu_L \Delta\theta = -q_L (m_L - t) \Delta\theta.\tag{C.19}$$

Combining (36)-(37) and recalling that $m_i \equiv w(y_i) + \theta_i B - U_i$ yields

$$\begin{aligned}U_L + t &= q_H [w(y_H) + \theta_H B - (U_H - U_L)] + q_L [w(y^*) + \theta_L B] \\ &= q_H [w(y_H) + \theta_H B - y_H \Delta\theta] + q_L [w(y^*) + \theta_L B],\end{aligned}\tag{C.20}$$

where the second equality reflects the fact that (33) is an equality, since $\mu_L > 0$. Therefore:

$$\begin{aligned}
m_L - t &= w(y^*) + \theta_L B - U_L - t \\
&= w(y^*) + \theta_L B - q_L [w(y^*) + \theta_L B] - q_H [w(y_H) + \theta_H B - y_H \Delta\theta] \\
&= q_H [w(y^*) - w(y_H) - (B - y_H) \Delta\theta]
\end{aligned} \tag{C.21}$$

Substituting into (C.19) yields

$$\Phi(y_H; t) \equiv w(y_H) - w(y^*) + (B - y_H) \Delta\theta + \frac{tw'(y_H)}{q_L \Delta\theta} = 0. \tag{C.22}$$

The following lemma characterizes the equilibrium value of y_H over Region I, denoted $\hat{y}_H^I(t)$.

Lemma 3 *For any $t \geq 0$ there exists a unique $\hat{y}_H^I(t) \in (y^*, B)$ to (C.22). It is strictly decreasing in t , starting from the perfectly competitive value $\hat{y}_H^I(0) = y_H^c$.*

Proof. The function $\Phi(y; t)$ is strictly decreasing in y on $[y^*, B)$, with $\Phi(y^*) > 0 > \Phi(B)$, hence existence and uniqueness. Strict monotonicity then follows from the fact that Φ is strictly decreasing in t , while setting $t = 0$ in (C.22) shows that $\hat{y}_H^I(0)$ must equal y_H^c , defined in (22) as the unique solution to

$$w(y^*) - w(y_H^c) = (B - y_H^c) \Delta\theta.$$

It only remains to verify that the solution $\hat{y}_H^I(t)$ is consistent with the initial assumption that $\nu = 0$, or equivalently $U_L > \bar{U}$. By (C.20), we have for all y_H

$$\begin{aligned}
U_L + t &= q_H [w(y_H) + \theta_H B - y_H \Delta\theta] + q_L [w(y^*) + \theta_L B] \\
&= w(y^*) + \theta_L B + q_H [(B - y_H) \Delta\theta + w(y_H) - w(y^*)].
\end{aligned}$$

■

For $y_H = \hat{y}_H^I(t)$, the corresponding value of U_L is strictly above \bar{U} if and only if $\psi(t) > \bar{U} + t$, where we define for all t :

$$\psi(t) \equiv w(y^*) + \theta_L B + q_H [(B - \hat{y}_H^I(t)) \Delta\theta - w(y^*) + w(\hat{y}_H^I(t))]. \tag{C.23}$$

Lemma 4 *There exists a unique $t_1 > 0$ such that $\psi(t) \geq \bar{U} + t$ if and only if $t \leq t_1$. On $[0, t_1]$, the low type's utility U_L is strictly decreasing in t , reaching \bar{U} at t_1 .*

Proof. At $t = 0$ the bracketed term is zero by definition of $\hat{y}_H^I(0) = y_H^c$, so $\psi(0) = w(y^*) + \theta_L B > \bar{U}$ by (16), which stated that a monopsonist hires both types, and $\lim_{t \rightarrow +\infty} [\psi(t) - \bar{U} - t] = -\infty$, there exists at least one solution to $\psi(t) = \bar{U} + t$. To show that it is unique and the monotonicity

of U_L , we establish that, $\psi'(t) < 1$ for all $t > 0$. From (C.22) and (C.23), this means that

$$\begin{aligned}
q_H [\Delta\theta - w'(\hat{y}_H)] \left(\frac{-w'(\hat{y}_H)/q_L\Delta\theta}{\Delta\theta - w'(\hat{y}_H) - tw''(\hat{y}_H)/q_L\Delta\theta} \right) &< 1 \iff \\
q_H [\Delta\theta - w'(\hat{y}_H)] (-w'(\hat{y}_H)/q_L\Delta\theta) &< \Delta\theta - w'(\hat{y}_H) - tw''(\hat{y}_H)/q_L\Delta\theta \iff \\
q_H [\Delta\theta - w'(\hat{y}_H)] (-w'(\hat{y}_H)) &< q_L\Delta\theta [\Delta\theta - w'(\hat{y}_H)] - tw''(\hat{y}_H) \iff \\
tw''(\hat{y}_H) &< [\Delta\theta - w'(\hat{y}_H)] [q_H w'(\hat{y}_H) + q_L\Delta\theta]
\end{aligned}$$

where we abbreviated $\hat{y}_H^I(t)$ as \hat{y}_H . In the last expression, the left-hand side is always non-negative, whereas on the right hand side $y^* < \hat{y}_H < \hat{y}_H^c$ implies that $w'(\hat{y}_H) < 0$ and $q_H w'(\hat{y}_H) + q_L\Delta\theta > q_H w'(\hat{y}_H^c) + q_L\Delta\theta > 0$, by (23). ■

In summary, Region I consists of the interval $[0, t_1]$, where t_1 is uniquely defined by $\psi(t_1) = t_1$. Over that interval, $y_L = y^*$ while $y_H = \hat{y}_H^I(t)$ is strictly decreasing in t , and therefore so is the high type's relative rent, $U_H - U_L = \hat{y}_H^I(t)\Delta\theta$. The low type's utility level U_L need not be declining, but it starts at a positive value and reaches \bar{U} exactly at t_1 .

For $t \geq t_1$, the constraint $U_L \geq \bar{U}$ is binding. Recalling that $\mu_H\mu_L$ must always equal zero, we distinguish two subregions, depending on whether $\mu_H = 0$ (Region II) or $\mu_L = 0$ (Region III), and show that these are two intervals, respectively $[t_1, t_2]$ and $[t_2, +\infty)$, with $t_1 < t_2$. Thus, inside Region II the low type's incentive constraint is binding but not the high type's ($\mu_L > 0 = \mu_H$ for $t \in (t_1, t_2)$), whereas inside Region 2 it is the reverse ($\mu_H > 0 = \mu_L$ for $t > t_2$).

Region II. Consider first the values of t where $\mu_H = 0 < \mu_L$. As before, this implies that $y_L = y^* < y_H$ and $U_H - U_L = y_H\Delta\theta$, or $U_H = \bar{U} + y_H\Delta\theta$ since $U_L = \bar{U}$. Therefore:

$$\mu_L = q_H(m_H - t) = q_H [w(y_H) + \theta_H B - U_H - t] = q_H [w(y_H) + \theta_H B - y_H\Delta\theta - \bar{U} - t]. \quad (\text{C.24})$$

Substituting into condition (38), the latter becomes

$$\Gamma(y_H; t) \equiv w(y_H) + \theta_H B - y_H\Delta\theta - \bar{U} - t + \frac{tw'(y_H)}{\Delta\theta} = 0. \quad (\text{C.25})$$

On the interval $[y^*, B)$, the function $\Gamma(y; t)$ is strictly decreasing in y_H and t , with

$$\begin{aligned}
\Gamma(\hat{y}_H^I(t); t) &\equiv w(\hat{y}_H(t)) + \theta_H B - \hat{y}_H^I(t)\Delta\theta + \frac{tw'(\hat{y}_H^I(t))}{\Delta\theta} - \bar{U} - t \\
&= w(y^*) + \theta_L B + \left(1 - \frac{1}{q_L}\right) \frac{tw'(y_H^I(t_1))}{\Delta\theta} - \bar{U} - t \\
&= w(y^*) + B\theta_L - t \left(1 + \frac{q_H w'(y_H)}{q_L \Delta\theta}\right) - \bar{U}.
\end{aligned} \quad (\text{C.26})$$

At $t = t_1$, substituting (C.22) into (C.23) yields $\Gamma(\hat{y}_H^I(t_1); t_1) = 0$. Furthermore, as t rises above t_1 , $\hat{y}_H^I(t)$ decreases, so $w'(\hat{y}_H^I(t))$ increases. Since

$$q_L \Delta\theta + q_H w'(\hat{y}_H(t)) > q_L \Delta\theta + q_H w'(\hat{y}_H(0)) = q_L \Delta\theta + q_H w'(y_H^c) > 0$$

by (23), $t [q_L \Delta\theta + q_H w'(\hat{y}_H^I(t))]$ is also increasing in t , implying that $\Gamma(\hat{y}_H^I(t); t)$ is decreasing in t and therefore negative over $(t_1 + \infty)$. Next, observe that

$$\Gamma(y^*; t) = w(y^*) + \theta_H(B - y^*) + \theta_L y^* - \bar{U} - t.$$

Define therefore

$$t_2 \equiv w(y^*) + \theta_H(B - y^*) + \theta_L y^* - \bar{U}, \quad (\text{C.27})$$

and note that

$$\begin{aligned} t_1 &= w(y^*) + \theta_L B + q_H [(B - \hat{y}_H^I(t_1)) \Delta\theta - w(y^*) + w(\hat{y}_H^I(t_1))] - \bar{U} \\ &< w(y^*) + \theta_L B + q_H (B - \hat{y}_H^I(t_1)) \Delta\theta \bar{U} \\ &< w(y^*) + \theta_L B + (B - y^*) \Delta\theta \bar{U} = t_2. \end{aligned}$$

Lemma 5 *For all $t \in [t_1, t_2]$, there exists a unique $\hat{y}_H^{II}(t) \in [y^*, \hat{y}_H^I(t_1)]$ such that $\Gamma(\hat{y}_H^{II}(t); t) = 0$. Furthermore, $\hat{y}_H^{II}(t)$ is strictly decreasing in t , starting at $\hat{y}_H^{II}(t_1) = \hat{y}_H^I(t_1)$ and reaching y^* at $t = t_2$. For all $t > t_2$, $\Gamma(y_H; t) < 0$ over all $y_H \geq y^*$.*

Proof. For $t \in [t_1, t_2]$ we have shown that $\Gamma(\hat{y}_H(t_1); t) \leq 0 \leq \Gamma(y^*; t)$, with the first equality strict except at t_1 and the second one strict except at t_2 . Since $\Gamma(y; t)$ is strictly decreasing in y_H and t , the claimed results follow. The fact that on $(t_1, t_2]$ the graph of $\hat{y}_H^{II}(t)$ lies strictly below that of $\hat{y}_H^I(t)$ also means that if there is a kink between the two curves at t_1 it is a convex one, as illustrated on Figure III. And indeed, differentiating (C.22) and (C.25), we have $-(\hat{y}_H^I)'(t_1) < -(\hat{y}_H^{II})'(t_1)$ if and only if

$$\begin{aligned} -\frac{-w'}{q_L \Delta\theta(\Delta\theta - w') - t w''} &< \frac{\Delta\theta - w'}{\Delta\theta(\Delta\theta - w') - t w''} \iff \\ q_L \Delta\theta \left(\frac{\Delta\theta}{-w'} + 1 \right) + t \left(\frac{-w''}{-w'} \right) &> \Delta\theta + t \left(\frac{-w''}{\Delta\theta - w'} \right) \iff \\ t(-w'') \left[\frac{1}{-w'} - \frac{1}{\Delta\theta - w'} \right] &> [q_H(-w') - q_L \Delta\theta] \Delta\theta = \frac{-(q_H w' + q_L \Delta\theta)}{-w'} \\ t \left(\frac{-w''}{-w'} \right) \left[\frac{1}{\Delta\theta - w'} \right] &> -(q_H w' + q_L \Delta\theta). \end{aligned}$$

with all derivatives evaluated at $\hat{y}_H^I(t_1) = \hat{y}_H^{II}(t_1)$. Since $y^* < \hat{y}_H^I(t_1) < y_H^c$ the term on the left is positive and that on the right negative. ■

As to t_2 , note that it is the only point where $\mu_H = 0 = \mu_L$ (the only intersection of Regions II and III). Indeed, this requires $y_H = y^* = y_L$ by (38)-(39) and condition (39) together with $U_L = \bar{U}$

then implies that $t = m_L = w(y^*) + \theta_L B - y^* \Delta \theta_L - \bar{U} = t_2$.

Region II thus consists of the interval $[t_1, t_2]$. Over that interval, $y_L = y^*$ while $y_H = \hat{y}_H^I(t)$ is strictly decreasing in t , and therefore so is the high type's utility, $U_H = \bar{U} + y_H^*(t) \Delta \theta$, while the low type's utility remains fixed at $U_L = \bar{U}$. Furthermore, we can show.

Putting together Regions I and II, we shall define:

$$\hat{y}_H(t) = \begin{cases} \hat{y}_H^I(t) & \text{for } t \in [0, t_1] \\ \hat{y}_H^II(t) & \text{for } t \in [t_1, t_2] \end{cases}. \quad (\text{C.28})$$

Region III. Inside this region, namely for $t > t_2$, we have $U_L = \bar{U}$ but now $\mu_H > \mu_L = 0$. This implies that $y_H = y^* > y_L$ by (38)-(39) and $U_H = \bar{U} + y_L \Delta \theta$ by (33). Furthermore,

$$\mu_H = q_H (t - m_H) = q_H [t + \bar{U} + y_L \Delta \theta - w(y^*) - \theta_H B]$$

Substituting into condition (39), the latter becomes

$$\Lambda(y_L; t) \equiv q_H [w(y^*) + \theta_H B - y_L \Delta \theta - \bar{U} - t] + \frac{t q_L w'(y_L)}{\Delta \theta} = 0. \quad (\text{C.29})$$

On the interval $[0, y^*]$, the function $\Lambda(y; t)$ is strictly decreasing in y_L , with

$$\Lambda(y^*; t) = q_H [w(y^*) + \theta_H B - y^* \Delta \theta - \bar{U} - t] = q_H (t_2 - t) < 0.$$

Recall now that the monopsony price y_L^m is uniquely defined by $w'(y_L^m) = (q_H/q_L) \Delta \theta$. Therefore:

$$\Lambda(y_L^m; t) = q_H [w(y^*) + \theta_H B - \bar{U} + (B - y_L^m \Delta) \theta] > 0.$$

Lemma 6 For all $t \geq t_2$ there exists a unique $\hat{y}_L(t)$ such that $\Lambda(\hat{y}_L(t); t) = 0$, and $y_L^m < \hat{y}_L(t) \leq y^*$, with equality at $t = t_2$. Furthermore, $\hat{y}_L(t)$ is strictly decreasing in t and $\lim_{t \rightarrow +\infty} \hat{y}_L(t) = y_L^m$.

Proof. Existence and uniqueness have been established. Next, $\partial \Lambda(y; t) / \partial t = q_L w'(y) / \Delta \theta - 1$. At $y = \hat{y}_L(t)$, this equals $1/t$ times

$$-q_H [w(y^*) + \theta_H B - y_L \Delta \theta - \bar{U} - t] - t = -q_L t - q_H [w(y^*) + \theta_H B - \bar{U} - y_L \Delta \theta] < 0,$$

so the function $\hat{y}_L(t)$ is strictly decreasing in t . Taking limits in (C.29) as $t \rightarrow +\infty$, finally, yields as the unique solution $\lim_{t \rightarrow +\infty} \hat{y}_L(t) = y_L^m$. ■

Proof of Proposition 6 The fact that $\partial U_L / \partial t < 0$ over Region I was shown in Lemma 4. To show the last result, note that over Region III, we have

$$\begin{aligned} 2\Pi &= q_H [w(y^*) + \theta_H B - \hat{y}_L \Delta \theta] + q_L [w(\hat{y}_L) + \theta_L B] - \bar{U} \Rightarrow \\ \frac{2}{q_L} \frac{\partial \Pi}{\partial \hat{y}_L} &= w'(\hat{y}_L) - \frac{q_H}{q_L} \Delta \theta > w'(y_L^m) - \frac{q_H}{q_L} \Delta \theta = 0, \end{aligned}$$

so profits fall as t declines, as was shown to be the case over Regions I and II. ■

Proof of Proposition 7

The result for $Y_H - Y_L$ was shown in the text. For performance-based pay, we have

$$\frac{\partial([b(y_H) + \theta_H]y_H - [b(y_L) + \theta_L]y_L)}{\partial t} = [b(y_H) + \theta_H + y_H b'(y_H)] \frac{\partial y_H}{\partial t} - [b(y_L) + \theta_L + y_L b'(y_L)] \frac{\partial y_L}{\partial t}.$$

In Regions I and II the first term is negative and the second zero; in Region III it is the reverse. Turning finally to fixed wages,

$$z_H - z_L = U_H - U_L - \theta_H y_H + \theta_L y_L - u(y_H) + u(y_L)$$

In Regions I and II, $z_H - z_L = -\theta_L(y_H - y^*) - u(y_H) + u(y^*)$ is decreasing in y_H , hence increasing in t . In Region III, $z_H - z_L = (y_L - y^*)\theta_H - u(y^*) + u(y_L)$, so the opposite holds. ■

Appendix D: Additional Proofs

Bounds on \bar{U} ensuring non-negative wages. We make explicit here the joint restrictions placed on \bar{U} by the requirements that z_H be non negative²⁴ and that firms want to keep low types one board, then provide sufficient conditions for all of them to hold jointly. In Region III, $z_H(t) = \bar{U} + \hat{y}_L(t)\Delta\theta - u(y^*) - \theta_H y^*$ is decreasing, so it suffices that $\lim_{t \rightarrow +\infty} z_H(t) = z_H^m \geq 0$. In Regions I and II, $z_H(t) = U_L(t) - \theta_L \hat{y}_H(t) - u(\hat{y}_H(t))$; its variations with t are ambiguous, but since $U(t)$ is (weakly) declining toward \bar{U} and $\hat{y}_H(t)$ strictly decreasing from $\hat{y}_H^c(0) = y_H^c$, it is bounded below by $\bar{U} - \theta_L y_H^c - u(y_H^c)$. Combining this with condition (16), we therefore require:

$$\bar{U}_{\min} \equiv \max \{u(y^*) + \theta_H y^* - y_L^m \Delta\theta, u(y_H^c) + \theta_L y_H^c\} \leq \bar{U} \leq w(y_L^m) + \theta_L B - (q_H/q_L)y_L^m \Delta\theta \equiv \bar{U}_{\max}. \quad (\text{C.1})$$

This defines a nonempty interval for \bar{U} as long as $\bar{U}_{\min} < \bar{U}_{\max}$, which can be insured in at least two ways. First, for q_L close enough to 1, y_L^m is close to y^* , so $\bar{U}_{\min} \approx u(y_H^c) + \theta_L y_H^c < w(y_H^c) + \theta_L y_H^c < w(y^*) + \theta_L B \approx \bar{U}_{\max}$, ensuring the result.

Alternatively, one can slightly modify firms' technology so that the revenue generated by each worker of type θ becomes instead $Aa + B(\theta + b) + \bar{d}$, where \bar{d} is a constant reflecting some other "basic" activity, performed at a fixed (e.g., perfectly monitored) level by all employees, and for which their compensation is therefore part of the fixed wage z . This augments total surplus $w(y)$ by the same amount \bar{d} , which can be made large enough to ensure that $\bar{U}_{\min} < \bar{U}_{\max}$. ■

[[Appendix D to be completed]].

²⁴Recall that $z_L \geq z_H$ everywhere by incentive compatibility. As to the bonus rates y_i , they all are bounded below by y_L^m , which is positive since we have assumed that $w'(0) > (q_H/q_L)\Delta\theta$.

REFERENCES

- Acharya, V., Gabarro, M., and P. Volpin (2012) “Competition for Managerial Talent, Corporate Governance and Incentive Compensation,” CEPR W.P. 8936.
- Acharya, V., Pagano, P. and P. Volpin (2012) “Seeking Alpha: Excess Risk Taking and Competition for Managerial Talent,” NUY mimeo, May.
- Akerlof, G. and and Kranton, R. (2005) “Identity and the Economics of Organizations,” *Journal of Economic Perspectives*, 19(1), 9-32.
- Armstrong, M. and J. Vickers (2001) “Competitive Price Discrimination”, *Rand Journal of Economics*, 32(4), 579-605.
- Baker, G., R. Gibbons and K.J. Murphy (1994) “Subjective Performance Measures in Optimal Incentive Contracts”, *Quarterly Journal of Economics*, 109, 1125-56.
- Bandiera, O., Barankay, I. and I. Rasul (2007) “Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment”, *The Quarterly Journal of Economics*, 122, 729-773.
- Baker, G. (1992) “Incentive Contracts and Performance Measurement,” *Journal of Political Economy*, 100(3), 98-614.
- Bartling, B., Fehr, E. and K. Schmidt (2012) “Screening, Competition and Job Design”, *American Economic Review*, forthcoming.
- Baumol, W. J. (1990): “Entrepreneurship: Productive, Unproductive, and Destructive,” *Journal of Political Economy*, 98, 893-921.
- Bebchuk, L. and J. Fried (2004) *Pay without Performance: The Unfulfilled Promise of Executive Compensation*. Cambridge: Harvard University Press, 2004.
- Bénabou, R., and J. Tirole (2003), “Intrinsic and Extrinsic Motivation,” *Review of Economic Studies*, 70, 489-520.
- Bénabou, R., and J. Tirole (2006), “Incentives and Prosocial Behavior,” *American Economic Review*, 96(5), 1652-1678.
- Bénabou, R., and J. Tirole (2011), “Laws and Norms,” NBER Working Paper 17579, November.
- Bénabou and Tirole (2012) “Competitive Pay, Work Ethics, and Corporate Image,” mimeo.
- Bertrand, M. and S. Mullainathan (2001) “Are CEOs Rewarded for Luck? The Ones without Principals Are”, *Quarterly Journal of Economics*, 116, 901-932.
- Besley, T., and M. Ghatak (2005) “Competition and Incentives with Motivated Agents,” *American Economic Review*, 95(3), 616-636.
- Besley, T., and M. Ghatak (2006) “Sorting with Motivated Agents: Implications for School Competition and Teacher Incentives,” *Journal of the European Economics Association*, 4, 404-414.
- Besley, T. and M. Ghatak (2011) “Taxation and Regulation of Bonus Pay”, mimeo, LSE.
- Bolton, P., Scheinkman, J. and W. Xiong (2006) “Executive Compensation and Short-Termist Behaviour in Speculative Markets”, *Review of Economic Studies*, 73, 577-610.

- Brekke, K. A., and K. Nyborg (2008) "Attracting Responsible Employees: Green Production as Labor Market Screening," *Resource and Energy Economics*, 39, 509-526.
- Crocker, K. and A. Snow (1985) "The Efficiency of Competitive Equilibria in Insurance Markets with Asymmetric Information", *Journal of Public Economics*, 26, 207-219.
- Desai, M. (2012) "The Incentive Bubble," *Harvard Business Review*, 90(3), 124-132,
- Dewatripont, M., I. Jewitt and J. Tirole (1999a) "The Economics of Career Concerns, Part 1: Comparing Information Structures", *Review of Economic Studies*, 66, 183-198.
- Fabbri, F. and D. Marin (2012) "What explains the Rise in CEO Pay in Germany? A Panel Data Analysis for 1977-2009", CESifo WW.P. 3757.
- Fehr, E. and K. Schmidt (2004) "Fairness and Incentives in a Multi-Task Principal-Agent Model," *Scandinavian Journal of Economics*, 106, 453-474.
- Fehr, E. and List, J. (2004) "The Hidden Costs and Returns of Incentives –Trust and Trustworthiness Among CEO's," *Journal of the European Economic Association*, 2(5), 743-771.
- Frydman, C. (2007) "Rising Through the Ranks: The Evolution of the Market for Corporate Executives, 1936-2003," MIT mimeo, September.
- Frydman, C. and D. Jenter (2010) "CEO Compensation," *Annual Review of Financial Economics*, 2(1), 75-102.
- Frydman, C. and R. E. Saks (2010) "Executive Compensation: A New View from a Long-Term Perspective, 1936-2005," *Review of Financial Studies*, 23, 2099-2138.
- Gabaix, X. and A. Landier (2008) "Why has CEO Pay Increased So Much?" *Quarterly Journal of Economics*, 123: 49-100.
- Garicano, L. and E. Rossi-Hansberg (2006) "Organization and Inequality in a Knowledge Economy," *Quarterly Journal of Economics* 121(4), 1383-1435.
- Gawande, A. (2009) "The Cost Conundrum: What a Texas Town Can Teach Us About Health Care," *The New Yorker*, June 1.
- Holmström B. (1982) "Moral Hazard in Teams," *The Bell Journal of Economics*, 13(2), 324-340.
- Holmström B., and P. Milgrom (1991) "Multi-Task Principal-Agent Analyzes: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics and Organization*, 7 (special issue): 24-52.
- Itoh, H. (1991) "Incentives to Help in Multi-Agent Situations", *Econometrica*, 59, 611-636.
- Kosfeld, M. and F. von Siemens (2011) "Competition, Cooperation and Corporate Culture", *Rand Journal of Economics*, 42, 23-43.
- Laffont, J.J., and J. Tirole (1986), "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 94, 614-641.
- Lazear, E. (2000) "Performance Pay and Productivity," *American Economic Review*, 90(5), 1346-1361.
- Lemieux, T., MacLeod, B. and D. Parent (2009) "Performance Pay and Wage Inequality," *Quarterly Journal of Economics*, 104 (1), 1-49.
- Marin, D. and T. Verdier (2009) "Globalization and the Empowerment of Talent", *Journal of International Economics*, 86, 209-223.

- Maskin, E., and J. Tirole (1992), “The Principal-Agent Relationship with an Informed Principal, II: Common Values,” *Econometrica*, 60(1), 1-42.
- Nalebuff, B. and J. Stiglitz (1983) “Prices and Incentives: Towards a General Theory of Compensation and Competition,” *Bell Journal of Economics*, 14(1), 21-43.
- Philippon, T. and A. Reshef (2012) “Wages and Human Capital in the U.S. Financial Industry: 1909-2006”, *Quarterly Journal of Economics*, forthcoming.
- Prendergast, C. (1992) “Career Development and Specific Human Capital Collection,” *Journal of the Japanese and International Economies*, 6, 207-227.
- Prendergast, C. (2007) “The Motivation and Bias of Bureaucrats,” *American Economic Review*, 97(1), 180–196.
- Raith, M. (2003) “Competition, Risk, and Managerial Incentives,” *American Economic Review*, 93(4), 1425–1436.
- Ramalingam, A. and Rauh, M. (2010) “The Firm as a Socialization Device,” *Management Science*, 56(12), 2191-2206.
- Rochet, J.C. and L. Stole (2002) “Nonlinear Pricing with Random Participation,” *Review of Economic Studies*, 69, 1, 277-311.
- Rosen, S. (1981) “The Economics of Superstars”, *American Economic Review*, 71: 845-858.
- Rothschild, M., and J. Stiglitz (1976) “Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information,” *The Quarterly Journal of Economics*, 90(4), 629–649.
- Schmidt, K. (1997) “Managerial Incentives and Product Market Competition,” *Review of Economic Studies*, 64(2), 191-213.
- Shearer, B. (2004) “Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment,” *Review of Economic Studies*, 71, 513-534.