RELATIVE AMBITION AND THE ROLE OF WAGE SECRECY IN LABOR CONTRACTS

Tomer Blumkin and David Lagziel

Discussion Paper No. 19-02

September 2019

Monaster Center for
Economic Research
Ben-Gurion University of the Negev
P.O. Box 653
Beer Sheva, Israel

Fax: 972-8-6472941
Tel: 972-8-6472286

Relative Ambition and the Role of Wage Secrecy in Labor Contracts*

Tomer Blumkin† and David Lagziel‡

September 25, 2019

Abstract:

In this paper we study the properties of optimal labor contracts in an efficiency-wage setting with homogeneous workers whose utilities depend both on their absolute and relative wages, compared to their co-workers. Assuming that relative wages carry a two-sided effect over workers' incentives, we characterize necessary and sufficient conditions for wage dispersion and wage secrecy to be part of the optimal labor contract. We show the important role played by the extent of complementarity exhibited by the production function, and further demonstrate the robustness of our results to the incorporation of general equilibrium stability considerations.

Journal of Economic Literature classification numbers: E24, D82, J30, J31, J71.

Keywords: secrecy; wages; relative wage; labor contracts; wage compression; wage dispersion.

*The authors wish to thank Leif Danziger, Ada Gonzalez-Torres, Ori Heffetz, Nadav Levy, Omer Moav, Tal Regev, Moses Shayo, Ran Spiegler, Yoram Weiss, David Wettstein, Ro'i Zultan, as well as the participants of the IDC Economics seminar, the Brown-Bag Economics seminar at BGU, the Behavioral and Experimental Economics seminar at the Hebrew University, the Economic Theory seminar at Tel-Aviv University, the Economic Theory seminar at Bar-Ilan University, and the Strategy seminar at Coller Business School at Tel-Aviv University for their valuable comments and suggestions. Blumkin and Lagziel acknowledge the support of the Israel Science Foundation, Grant #513/19.

†Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel. e-mail: Tomerblu@bgu.ac.il; CESifo, Poschingerstrasse 5, 81679 Munich, Germany; IZA, Schaumburg-Lippe Strasse 7/9, 53113 Bonn, Germany.

‡Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel. e-mail: Davidlag@bgu.ac.il.

"...rank among our equals is, perhaps, the strongest of all our desires, and our anxiety to obtain the advantages of fortune is accordingly much more excited and irritated by this desire..."

—Adam Smith (1759), *The Theory of Moral Sentiments*, VI.I.4.

1 Introduction

Envy and competitiveness are core emotions at the human need to compare oneself to others. While competitiveness directs an individual to gain an advantage over others, envy dictates a basic desire to level a relative shortage. Essentially, the two emotions produce complementary effects - either downwards or upwards. In this paper we offer some insights into the repercussions of these elements (henceforth, referred to as *relative ambition*) through a natural economic setting.

Our problem begins with a simple principal-agent interaction in which a firm designs incomplete contracts so that employees receive fixed wages. We augment a standard gift-exchange framework by assuming that the ex-ante homogenous workers exhibit some form of relative ambition: their utility functions depend, in addition to individual effort levels and personal wages, on their counterparts' wages. In particular, we assume that an increase in others' wages and enhanced relative ambition adversely affects employees' incentives. Given a general CES production function, we address the problem of optimal labor contracts, focusing on the role played by two key parameters: (i) the extent of relative ambition exhibited by the workers; and, (ii) the degree of complementarity exhibited by the production function. In each aspect we offer a specific insight, explained as follows.

The first part concerns the variation in optimal incentives as employees' relative ambition intensifies. As long as ambition levels are low, employees pay little attention to others' wages, and the firm exercises a simple uniform-pay policy. This is anticipated in light of the presumed symmetry in the production function and workers' homogeneity. However, once relative ambition considerations become more manifest (i.e., employees are more concerned with their relative lot vis-a-vis their co-workers), a uniform-pay policy becomes increasingly costly. There is simply no way to efficiently incentivize an entire group of people, whenever their main concern is their ordinal ranking. A superior remuneration strategy involves wage dispersion which discriminates among ex-ante homogeneous employees. In essence, we prove that wage gaps are a natural reaction to the human need of competition.

The second part focuses on different levels of substitution in production. Under high level of complementarity, employees cooperation weighs heavily on production, and the firm must resort to a uniform compensation policy even in the backdrop of high-level relative ambition. Since the common solution of wage compression becomes increasingly costly, we propose a novel approach - the use of *secrecy*. We show that a firm can, and sometimes must, employ an undisclosed-wage policy, such that only the average wage of workers is common knowledge. In other words, the firm uses secrecy to reduce the wage benchmark that employees use to assess their relative reward. To differ, under high level of substitution in production, little cooperation is needed and the firm only gains from fully-disclosed wage-gaps to address the relative ambition concerns.

The combined presentation of the two stated results is somewhat misleading since the two are, in fact, independent. First, the use of a confidential-pay policy is relevant as long as some form of relative ambition exists. Whether employees are homogeneous or not, the firm can use confidentiality to reduce the employees wage benchmark and optimize incentives. The same result would hold even if wage dispersion follows from employees heterogeneity. Second, wage dispersion is a natural response to the employees' concerns about their relative pay. It goes to the core problem of optimal incentives for agents whose primary concern is their ordinal ranking, and it is irrelevant to any disclosure concerns.

In the last part of the paper, we extend the basic set-up to a general-equilibrium setting, with ex-ante and ex-post stability considerations and free entry. Our results indicate that secrecy is potentially beneficial to individuals' expected welfare, thus bearing non-standard policy implications defying the conventional wisdom, which calls for enhanced wage transparency. In particular, we prove that fully disclosed policies are not (necessarily) supported in equilibrium, since firms' have profitable deviations towards confidential wage policies, that are supported in equilibrium. Thus, a regulatory restriction to eliminate secrecy could give rise to suboptimal equilibria from the employees' perspective.

1.1 Related literature

Our paper relates to a broad literature in the field of labor economics that explores the impact of relative remuneration concerns, addressing the issue both from a theoretical perspective, focusing on the optimal design of labor contracts, and from an empirical perspective, measuring the effect of pay inequality on labor market outcomes. In particular, our paper relates to the (mostly empirically geared) strand in this literature that focuses on the impact of pay transparency.

The extent to which workers' perception of fairness and, consequently, their motivation to exert effort are affected by relative remuneration consideration bears important implications for optimal labor contracts and may provide a positive rationale for compensation practices such as wage-secrecy and wage-compression. A large strand in the labor economics literature examines the optimal design of incentive contracts with status-concerned agents [see, e.g., Frank (1984a,b), and Fershtman et al. (2003, 2006)]. Akerlof and Yellen (1990) were the first to suggest the potential role of wage compression as a means to alleviate co-workers' equity concerns in an efficiency-wage setting. More recently, Charness and Kuhn (2007) employed a reduced form gift-exchange framework, with heterogeneous workers (differing in productivity) whose effort choices are affected by their co-workers' wages. They explicitly demonstrate how profit-maximizing firms would respond to an increase in the responsiveness of workers to their co-workers' compensation by compressing wages.

Somewhat surprisingly there is a paucity in theoretical studies examining wage-secrecy arrangements in optimal labor contracts, or exploring the implications of such arrangements on labor market outcomes. A notable exception is the early study by Danziger and Katz (1997) that demonstrates how a wage-secrecy convention facilitates risk shifting between firms and workers in response to productivity shocks. More recently, Cullen and Pakzad-Hurson (2018) employed a dynamic wage-negotiation model to explore the impacts of enhanced pay-transparency along both the demand (wage-setting and hiring policies) and supply (workers' bargaining strategies) channels. In their set-up, workers stochastically learn the wages of their peers and can voluntarily choose to re-negotiate their contracts. Cullen and Pakzad-Hurson (2018) show that an increase in pay transparency induces an information externality that shifts, de-facto, the bargaining power from the workers to the firm, resulting in lower wage rates and higher hiring rates. Blumkin and Lagziel (2018) study the strategic role of wage secrecy arrangements in a labor market with matching frictions. They demonstrate how firms employ pay-secrecy policies to control the dissemination of wage-related information to job applicants, via professional social networks. Consequently, wage dispersion, both within and across firms, arises in equilibrium, ensuring that firms derive positive rents.

On the empirical side, relative pay reciprocity has been the subject of many experimental studies (for a recent survey see Charness and Kuhn (2011)). The studies that are particularly relevant for our analysis focus on 'horizontal comparability', namely the effect of peers' compensation. Their conclusions are rather mixed. Charness and Kuhn (2007), Fischer and Steiger (2009) and Hennig-Schmidt et al. (2010) find no effect of others remuneration on the level of effort; whereas, Gächter and Thöni (2010), Cohn et al. (2011), Greiner et al. (2011), Ku and Salmon (2012), and Bracha et al. (2015) do find an effect.

The incentivizing role of relative pay has been examined outside the lab in several recent studies. In a study of employees at the University of California, Card et al. (2012) find that giving workers access to a database listing the salary of their peers, results in a decrease in job satisfaction and a rise in self-reported job search among those with relatively low wages. Perez-Truglia (2015) use the 2001 policy change in the on-line availability of Norwegian tax records to find that income transparency increases the satisfaction gap between low- and high-earning individuals by more than 20%. Rege and Solli (2015) use the same Norwegian tax records policy change to study how relative compensation affects workers' incentives. They find that the information shock increases job separation for low-earning workers relative to high-earning ones. Cullen and Perez-Truglia (2018) use a field experiment to control the exposure of employees to wage related information about their peers and show that higher perceived peer salary decreases effort and output, as well as retention. Breza et al. (2018) use a field experiment with heterogeneous workers to show that when co-workers' productivity is difficult to observe, pay inequality reduces output by 0.45 standard deviations and attendance by 18 percentage points. Notably, the studies by Card et al. (2012), Perez-Truglia (2015), Rege and Solli (2015) and Cullen and Perez-Truglia (2018) all allude to the significant role played by pay transparency in determining labor market outcomes.

1.2 Structure of the paper

The paper is organized as follows. In Section 2 we present the basic set-up and discuss the key assumptions. Section 3 provides the main results for the single-firm problem. Section 4 extends the analysis to a general-equilibrium framework with free entry, incorporating both ex-ante and ex-post stability considerations. Section 5 provides concluding remarks.

2 The model

Consider a group of employees working for a single representative firm. The goal of the firm is to minimize costs, subject to a fixed level of production. The firm rewards employees through fixed individual payments. In turn, the (ex-ante) homogenous employees choose their efforts levels, taking into account their own pay structure as well as that of their counterparts.

Acknowledging that employees are concerned with their relative compensation, firms may desire to employ some form of a non-disclosure policy regarding wages, that prevents employees from sharing wage related information with their peers, so as to mitigate the associated dis-incentivizing effects. The firm can exercise a *confidential-pay policy* as an additional tool, besides the ability to fix wages, in the quest for an optimal outcome. We capture this secrecy aspect by assuming that employees are matched into couples such that individual partners are used as *benchmarks* where, under a confidential pay policy, a worker is only informed about the average level of remuneration of his/her teammate.

More formally, the process begins when all employees are matched into couples. For simplicity, we assume that there is a continuum of employees with a mass of 2, such that the total mass of couples is normalized to unity.[1] Next, the firm commits to a *feasible policy* $F \in \Delta\mathbb{R}^2_+$ which dictates the distribution of wages across all matched couples. A feasible policy must sustain two conditions. First, the marginals must be identical to reflect the true distribution of wages among employees. Second, F is either a product distribution (i.e., featuring independent marginals) or satisfies a property that both coordinates are deterministically dependent. A product distribution is considered a *confidential-pay policy* (CP policy) since partners' wages are independent, whereas the case of deterministically dependent wages reflects an *observable-pay policy* (OP policy) among matched employees.[2] Denote the set of feasible policies by $\mathcal{F}$.

Once wages are distributed according to F, the employees are privately informed of their realized pay.

[1] A continuum of workers provides a technical simplification as it allows the firm to employ any wage distribution, including a non-atomic one.

[2] An alternative way to consider the firm's strategy is through a two-stage decision process. First, the firm chooses a distribution over $\mathbb{R}_+$, which dictates the wages of all employees before they are matched. Next, employees are either randomly matched to generate independent wages within couples (a CP policy), or employees are deterministically matched to produce an OP policy.

All employees possess the same utility function

$$U(e, w; w_{\text{ref}}) \quad = \quad w - \frac{e^2}{2} + e(w - w_{\text{ref}}),$$

where $e \geqslant 0$ is the employee's effort level, $w \geqslant 0$ is the employee's fixed wage, and $w_{\text{ref}} \geqslant 0$ is the employee's reference wage, namely, the perceived fair-wage which also depends on the firm's secrecy arrangements. Following the formulation of Akerlof and Yellen (1990), we assume that the perceived fair-wage is a weighted average $w_{\text{ref}} = \gamma w_p + (1-\gamma)\underline{w}$, where w_p is the wage of the employee's partner, $\underline{w}$ is the employee's reservation wage (common to all workers, by the homogeneity assumption), and $\gamma \in [0,1]$ is a measure of the employees' sensitivity towards their partners' income. To simplify the exposition and analysis, we normalize the reservation wage to zero, $\underline{w} = 0$. Thus, with a slight abuse of notation, the employee's utility function is

$$U(e, w; w_p) = w - \frac{e^2}{2} + e(w - \gamma w_p).$$

Wages should be interpreted as real rates, an issue we revisit later when discussing the market equilibrium.[3] For concreteness, we assume the existence of a single consumption good produced by the representative firm.

The utility function has two components. The first is standard and measures the worker's net payoff, $w - \frac{e^2}{2}$; namely, the compensation minus the disutility from the effort exerted. The second component of the utility function, $e(w - \gamma w_p)$, captures our notion of relative ambition under incomplete contracts, through the concept of gift exchange.[4] Namely, for a fixed level of relative ambition, denoted by γ, the employee weighs-in his relative pay compared to the partner's pay, and privately decides on the level of effort to exert. The range of γ suggests that the employees' preferences vary from no relative ambition ($\gamma = 0$) to full comparability ($\gamma = 1$).

The output per matched couple, exerting efforts e and e_p, is determined by a CES production function $Q(e, e_p) = \left(\frac{1}{2}e^\rho + \frac{1}{2}e_p^\rho\right)^{1/\rho}$, where $\rho \leqslant 1$ (the notation $\rho = 0$ refers to the limit value, a Cobb-Douglas production function). The firm's total production is given by a strictly positive, increasing, and weakly

[3] With a slight re-arrangement of $w - \gamma w_p$, we obtain $(1-\gamma)w + \gamma(w - w_p)$, in line with Charness and Kuhn (2007) who studied a set-up where effort is a separable linear function of w and $w - w_p$. The normalized weights, $1-\gamma$ and γ, are a matter of technical simplification, rather than an essential one.

[4] The gift exchange literature, that stems from the seminal contribution of Akerlof (1982), views employer-worker relations, in the context of an incomplete labor contract, as a gift-exchange in which the employer's gift in the form of a fair wage (exceeding the competitive rate) is being reciprocated by the worker's gift in the form of efforts exceeding the minimum contractible levels. There is wide support for the predicted reciprocal labor relations in the lab, following the pioneering work of Fehr et al. (1993). The field evidence for the existence of gift exchange in the workplace is somewhat mixed (see Esteves-Sorenson (2018)). For a similar utility specification of the relative-ambition component see Breza et al. (2018), which augments DellaVigna et al. (2016) to capture the effect of peer wages on workers' morale (the psychological utility associated with exerting effort).

concave function H over the aggregate production of all teams.[5] Thus, assuming that the required production is exogenously fixed to X, the firm is confronted with the following cost-minimization problem,

$$\begin{aligned} \min_F C &= \min_{F \in \mathcal{F}} \mathbf{E}[w + w_p], \\ \text{s.t.} \quad H(\mathbf{E}[Q(e, e_p)]) &\geqslant X > 0, \\ e &= \operatorname{argmax}_{\tilde{e}} \mathbf{E}[U(\tilde{e}, w; w_p)], \\ e_p &= \operatorname{argmax}_{\tilde{e}} \mathbf{E}[U(\tilde{e}, w_p; w)], \end{aligned}$$

where the expectation operator $\mathbf{E}[\cdot]$ represents the aggregation across all matched couples at the firm's level, and taken w.r.t. F. A direct optimization shows that $e = \max\{w - \gamma \mathbf{E}[w_p], 0\}$ and a similar equality holds for e_p. Notice the crucial difference between OP and CP policies in the eyes of the employees. Under an OP policy, every worker knows his partner's realized wage w_p, whereas a CP policy suggests that employees only know their partners' expected wage based on F.

Reservation wage and minimal effort are fixed at zero in our model. This may seem prima facie unrealistic since zero-wage employees are, de-facto, not employed by the firm, and such wages are not implementable due to minimum-wage legislation. Thus, we emphasize that the chosen base levels are a matter of a simplifying normalization. One can consider an alternative scenario where wage and effort are strictly positive, supported on some reservation levels, reflecting industry norms and/or pertinent legislation. More importantly, our set-up could be interpreted as *a bonus-plan model* eliciting extra efforts, rather than an entire compensation scheme. With no loss in the generality of our qualitative insights, we abstract away from these elements, as the forces that apply under the following analysis would still hold given the updated setting.

3 Micro foundations

The first part of our analysis deals with the two extreme cases of the CES production function: perfect complements vs. perfect substitutes. Under the latter set-up, we prove that an observable-pay policy dominates any confidential one in terms of lower expected costs per unit of production. However, if γ is sufficiently high, then perfect complementarity yields the opposite outcome. Specifically, in case the production depends on the minimal amount of effort among the matched employees (a Leontief production function), the firm can use a confidential policy to reduce expected costs.

The driving force behind this result is the combination of perfect complementarity with the disutility from partners' wages. On the one hand, perfect complementarity requires high effort levels on the side of both employees simultaneously. On the other hand, the disutility from others' wages makes it extremely

[5] The function H is essential to the partial-equilibrium analysis as it eliminates trivial solutions (i.e., no production or an unbounded level of production). Meanwhile, one can assume H is the identity function.

costly (prohibitively, in the limiting case where γ goes to unity) to incentivize all employees at the same time using a fully disclosed compensation policy. Secrecy balances these two forces to produce an adequate number of high-wage matched employees, whose disutility from the expected average pay (and not their actual counterparts) is sufficiently low.

Theorem 1. *A confidential-pay policy is suboptimal in case of perfect substitutes. However, if $\gamma > \frac{1}{2}$, then an observable-pay policy is suboptimal in case of perfect complements.*

All proofs are deferred to the Appendix.

A key observation in the construction of the proof for the case of perfect complementarity is the role of *benchmarkers*. The latter refers to workers whose productive effort is sacrificed by the firm (which offers them a relatively low remuneration), so as to reduce the benchmark of their productive counterparts, who are offered a relatively high level of compensation. By employing a confidential wage policy, the firm can reduce the measure of benchmarkers and, at the same time, entail a significant reduction in the benchmark wage rate.

A notable implication of the proof of Theorem 1 is that, under perfect substitutes, one can restrict attention to CP policies supported on only two wage levels. This observation relates to a broader result, given in Lemma 1, stating that the entire analysis could be restricted to at most two wage levels. The result is based on the concavity of the production function which allows us to contract wage levels and increase expected productivity. Note that the same contraction would be amplified by any convex cost function (of the firm), as it will reduce expected costs.[6]

Lemma 1. *For every finite ρ, any optimal policy, either confidential or observable, could be induced by at most two wage levels.*

Intuitively, one needs only two wage levels: a high level to elicit productive efforts and a low level to reduce the benchmark. Any wage dispersion, either within the group of workers that exert a positive effort level or amongst the pool of benchmarkers, can be reduced and save costs hinging on the concavity of the production function that dictates a symmetric compensation structure. The only deviation to a non-symmetric structure is due to the role of benchmark reduction, hence the separation between the two wage levels. In line with our interpretation of the optimal remuneration policy as a differential bonus plan, the firm saves on its production costs by offering only a fraction of the workers a bonus, while paying the remainder of the workers a base level of compensation. Notably the latter property holds across the board and is independent of the degree of complementarity/substitutability between the effort levels.

Using Lemma 1, we can go beyond the two extreme cases of perfect complements and perfect substitutes, and extend Theorem 1 to any CES production function. Theorems 2 and 3 show that an ex-post

[6]Under the current formulation, the cost is linear in output. Notice that with a piece-rate remuneration scheme (see our discussion in Section 3.1), assuming a fixed wage per unit of effort and maintaining our assumptions w.r.t. the production function, the resulting cost function is quadratic.

discriminating and confidential policy cannot be optimal whenever employees are not zealous to compare wages. However, once relative ambition rises, the firm must employ some form of discrimination to reduce costs while motivating a sufficient amount of workers to produce X. The point at which the firm deviates from symmetric wages to non-symmetric ones depends on the level of complementarity among workers' production efforts, thus we split the results into two parts: $\rho \leqslant 0$ and $\rho \in (0, 1]$, considered in Theorems 2 and 3 respectively.

Theorem 2 relates to the case where levels of complementarity are relatively high, namely $\rho \leqslant 0$ (see Figure 1). It states that for low level of ambition (i.e., $\gamma \leqslant \frac{1}{2}$), wage dispersion is suboptimal and the firm should use a symmetric OP policy. However, if relative ambition rises (i.e.,$\gamma > \frac{1}{2}$), then the firm should use a discriminating CP policy, where a proportion of employees receive a high wage, while others receive nothing.

Theorem 2. *Fix $\rho \leqslant 0$. The symmetric observable-pay policy is optimal if and only if $\gamma \leqslant \frac{1}{2}$. If $\gamma > \frac{1}{2}$, then an ex-post asymmetric* CP *policy is optimal.*

The case characterized in Theorem 2 focuses on production functions exhibiting a relatively high degree of complementarity. In these cases the only feasible wage policy that induces a benchmark reduction is the confidential pay regime. Any discriminatory OP policy would imply that total output produced by the pair of workers would be zero (this reflects an extreme manifestation of the complementarity property - see our earlier discussion of the limiting Leontief case). With a confidential pay policy, the benchmark can be reduced probabilistically and not for each pair across the board. In the backdrop of a sufficiently high degree of relative ambition this turns out to be the cost-minimizing strategy.

Theorem 3 focuses on the second part of $\rho \in (0, 1]$, showing that there are three possible solutions: a symmetric OP policy; an asymmetric OP policy; and an asymmetric CP policy. The variation between these three solutions is best exemplified by Figure 1. There exists a continuously-decreasing function, $\gamma_1(\rho)$, such that wage compression is optimal at (γ, ρ) if and only if $\gamma \leqslant \gamma_1(\rho)$ (this is consistent with the result of Theorem 2 in the sense that $\gamma_1(0) = \frac{1}{2}$). Alternatively, for $\gamma > \gamma_1(\rho)$, wage gaps become imminent due to the high-level of competitiveness. The firm implements wage dispersion either with secrecy, a CP policy, or without it, an OP policy. The distinction between the two is a function $\gamma_2(\rho)$ of the complementarity level, where relatively high complementarity (i.e., a low ρ) requires secrecy.

Theorem 3. *Fix $\rho \in (0, 1]$ and $\gamma \in [0, 1]$. There exist two continuously-decreasing functions γ_1 and γ_2 such that the confidential-pay policy is optimal if and only if $\gamma_1(\rho) < \gamma < \gamma_2(\rho)$. Moreover, the symmetric observable-pay policy is optimal if and only if $\gamma \leqslant \min\{\gamma_1(\rho), \gamma_2(\rho)\}$; and the asymmetric observable-pay policy is optimal if and only if $\gamma \geqslant \max\{\gamma_1(\rho), \gamma_2(\rho)\}$.*

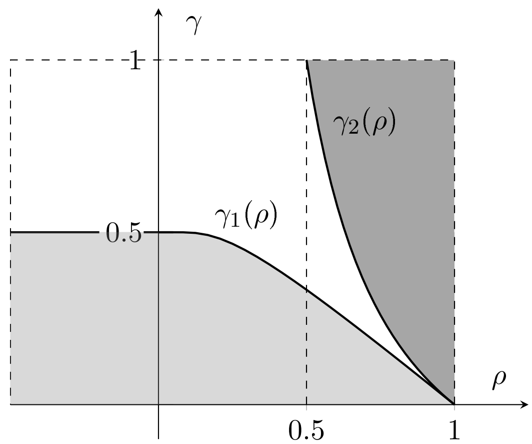

Figure 1: Optimal policies as a function of γ and ρ: light gray area = symmetric OP policy; white area = asymmetric CP policy; dark gray area = asymmetric OP policy. The proof of Theorem 3 also shows that $\gamma_1(0) = 0.5$, $\gamma_2(0.5) = 1$, and $\gamma_1(1) = \gamma_2(1) = 0$.

3.1 Robustness and (possible) extensions

Before we extend our model to partial- and general-equilibrium settings, we briefly discuss the robustness of our previous results. Four remarks are in order.

First, our derivations and conclusions are invariant to an alternative formulation that abstracts from specifying a utility function and considers, instead, a generic effort-response function (see Claim 1 in Appendix E). The utility function will be needed, however, for the general-equilibrium analysis and for deriving normative implications under the partial-equilibrium set-up, as will be shown below.

Second, for tractability reasons, our formulation assumes that relative ambition has a *symmetric* two-sided effect over incentives, upwards and downwards. Yet, asymmetries between upwards and downwards effects could arise, e.g., in case employees' possess a high level of envy with a low level of competitiveness (see the related discussion in Charness and Kuhn (2007) and Bracha et al. (2015)). Such asymmetries do not change the qualitative nature of our results. In addition, recent field studies provide some justification to our assumption. Cullen and Perez-Truglia (2018) provide evidence that the negative externality associated with peers' level of remuneration along the intensive margin (effort and performance) holds across the board, and affects both underpaid and overpaid workers. In their own words: "...the evidence does confirm that the symmetric approach from the baseline model is a reasonable approximation." In contrast, Breza et al. (2018) show that underpaid workers reduce their output, but find little evidence that performance improves when workers are paid more than their peers. Further note that previous experimental studies that tested the incentivizing effect of relative pay considered set-ups with heterogeneous workers. Such heterogeneity in productivity may justify differential remunerations and mitigate the incentivizing role of relative pay.

Third, though we believe that the gift-exchange formulation is a natural framework to explore the role of wage secrecy policies, we emphasize that our analysis can be readily extended to a complete-contract set-up. For instance, consider an alternative set-up with piece-rate remuneration and the following utility function $U(e, w; w_p) = ew - C(e, w_p)$, where the cost function C is assumed to satisfy the following properties $C_e > 0,\ C_{ee} > 0,\ C_{w_p} > 0$, and $C_{ew_p} > 0$. The first two properties are standard and imply a positive and increasing marginal cost of effort. The last two properties capture a novel behavioural component, such that the cost of effort depends on the partner's level of remuneration. With a few simple adjustments, all our qualitative results hold under the alternative specification.

Lastly, the use of couples for benchmark purposes could be easily extended to teams of more than two employees and different sizes. Such modelling choices would carry limited influence over our analysis and conclusions. The forces that lead the interaction and, specifically, the externalities of wages would still apply in a general-team setting.

4 Equilibrium analysis - stability and implementation

4.1 Partial equilibrium

The partial equilibrium analysis is straightforward and based on a fixed number of firms, not greater than half the number of employees (recalling that a single firm employs a mass of 2 employees). Under such conditions, all firms strive to maximize profit. By presumption that prices are normalized to unity and that wage rates are denoted in real terms, the profit is given by $X - C_{\gamma,\rho}(X)$ where $C_{\gamma,\rho}$ is the minimal cost given (γ, ρ). Previous results indicate that $C_{\gamma,\rho} = \alpha_{\gamma,\rho} H^{-1}(X)$ is a linear function of $H^{-1}(X)$. To eliminate trivial solutions (i.e., no production or an unbounded level of production), one can assume that H is strictly-concave such that the maximum-profit problem, $\max\{X - \alpha_{\gamma,\rho} H^{-1}(X)\}$, has a finite positive solution. Therefore, all firms would follow the previous analysis to minimize costs, while the production level is determined according to a maximum-profit condition. In particular, the characterization of the optimal pay policy and the necessary and sufficient conditions for the optimality of a CP policy, stated in Theorems 2 and 3, continue to hold under the partial equilibrium regime.

4.1.1 Policy implications

Our preceding analysis (Theorem 2) shows that, for $\gamma > 0.5$ and under perfect complementarity, firms would resort to a CP policy, in equilibrium. An interesting question is whether regulating pay transparency, which takes the form of restricting firms to OP policies, would be beneficial for workers. A straightforward computation reveals that such regulation would actually be harmful for all sides.

Observation 1. *Consider the case of perfect complementarity and, for tractability, fix $H(k) = k^{1/2}$. For every $\gamma > 0.5$, workers' ex-ante utility is higher under the optimal* CP *policy than under the optimal* (*restricted*) OP *policy.*

The fact that CP policies increase workers' ex-ante utility in a partial-equilibrium setting is intriguing. To differ, in a general-equilibrium setting firms must deal with free entry of other firms, so they must maintain a high level of satisfaction among employees. In a partial-equilibrium setting this concern is irrelevant. Yet, CP policies are ex-ante beneficial for workers since secrecy mitigates the dissatisfying effect of relative ambition. A win-win situation.

4.2 General equilibrium

Thus far we have confined our analysis to a simple partial equilibrium framework, in which the number of firms was fixed. Namely, incumbent firms were not threatened by potential entrants that might dissipate their rents. In addition, a key feature of our analysis was the desirability of confidential pay structures, that resulted in ex-post payoff differences amongst ex-ante identical workers. However, in the presence of free entry, workers can renegotiate their contracts with their current employer as long as firms derive positive rents. These rents ensure that employees have a credible threat to switch to an alternative firm that offers a more generous remuneration. Moreover, even in the absence of rents (i.e., whenever firms earn zero profits), workers who end up with a lower remuneration level would try to renegotiate their wages, since the superior possibility of switching remains feasible and credible. In this section, we incorporate these stability considerations into account. In particular, we show that our key insights carry over to the extended general equilibrium setting.

To guarantee these stability notions, the wage policy must maximize employees' expected utility, subject to two conditions: (i) the firm's profit is non-negative; and (ii) every realized wage weakly exceeds the expected utility minus the transaction costs associated with switching a firm.[7] Thus, given that all firms guarantee an expected utility of $\tilde{U}$, the single-firm maximization problem is reformulated as follows,

$$\begin{aligned} \max_{F} \ \mathbf{E}[U(e,w;w_p)] \quad &= \quad \max_{F\in\mathcal{F}} \ \mathbf{E}\left[w - \frac{e^2}{2} + e[w-\gamma w_p]\right], \\ \text{s.t.} \qquad H(\mathbf{E}[Q(e,e_p)]) \quad &\geqslant \quad \mathbf{E}[w+w_p], \\ U(e,w_0;\mathbf{E}[w_p]) \quad &\geqslant \quad \tilde{U}-T, \quad \forall w_0 \in \mathrm{Supp}(w), \end{aligned}$$

where all expected values are taken w.r.t. the pay policy F, the value $T>0$ denotes the transaction cost associated with switching a firm, and (e,e_p) are determined as before. The new optimization problem ensures that no firm would be able to attract employees by offering them a higher expected utility. Such a deviation would either be non-profitable (violating the first constraint) or turn out to be ex-post unstable as it would induce a positive measure of employees to renegotiate their terms (violating the second

[7]These costs may capture search frictions, costs of negotiating the new wage contract and the uncertainty due to the ex-ante stochastic compensation policy associated with switching to an alternative employer.

constraint). The objective function essentially imposes an ex-ante stability condition which suggests that all firms, in equilibrium, produce the same expected utility for workers. Thus, an equilibrium (à la Nash) induces a fixed point with respect to the objective function, and the solution for the above optimization must be $\tilde{U}$.

To simplify our exposition, we will focus on the case of perfect complementarity in production and, for tractability, fix $H(k) = k^{1/2}$. Clearly, any strictly-concave function H which ensures a positive finite level of production would apply as well. Another technical issue revolves around the existence of an equilibrium. For OP policies, the problem is straightforward since there is one possible profile to consider. Under symmetry in production within and across teams with free entry and workers' homogeneity, the optimal OP policy will specify a uniform wage rate. Therefore, one can derive the maximal wage from the first constraint (as the second becomes redundant) and solve the optimization problem. On the other hand, CP polices are more challenging. Note that the objective function is convex while the first constraint is concave, so the previously-used compression argument fails. Thus, we must work with general wage distributions, and it is unclear whether a solution to the problem exists.[8] To avoid technical issues of existence (which are beyond the scope of this paper) we assume that CP policies are supported on a finite number of bounded wage levels, such that the above problem could be embedded in a compact set (finite dimension, closed, and bounded).

Recall that in the partial equilibrium framework, we have shown that a CP policy dominates an OP regime in case γ is sufficiently high (i.e., $\gamma > 1/2$ in Theorem 1). Similarly, in the following theorem we provide a sufficient condition for the non-existence of an OP policy equilibrium due to firms' profitable deviation towards a CP policy, whereas a CP-policy equilibrium exists.

Theorem 4. *In the case of perfect complements and given a positive transaction cost, there exists a weakly-decreasing function, $\frac{1}{2} \leqslant \gamma(T) < 1$, such that for every $\gamma > \gamma(T)$ there exists no observable-pay policy equilibrium. Moreover, there exists $T^* > 0$ such that for every $T \geqslant T^*$ and every $\gamma > \gamma(T^*)$, a confidential-pay policy equilibrium exists.*

The rationale underlying the sufficient condition stated in Theorem 4 relies on two key insights. First, as in the partial-equilibrium setting, the optimality depends again on the degree of relative ambition. With a higher degree of relative ambition, a firm offering an OP policy (with uniform wage rates) would elicit lower levels of effort and consequently produce lower levels of output. Constrained by the zero profit condition, the latter implies that the maximized utility, under an OP policy, is decreasing in the degree of relative ambition. A switch to an alternative CP policy would serve to mitigate the relative ambition

[8] The set of probability distributions over $[0, \infty)$ is not necessarily compact, and the expectation functional need not be continuous. There is the possibility of reverting to a weak* topology, namely the vague topology, along with radon measures that guarantee tightness (i.e., vanishes at infinity; see Theorems 5.19-5.22 in Kallenberg (2002)). However, since this is not a mathematical paper dealing with modern methods in functional analysis, we revert to a simpler framework, without loss of *economic* generality.

dis-incentivizing effect, and therefore raise the workers' expected utility. Thus, when the degree of relative ambition is sufficiently high, a CP policy prevails over an OP policy. For example, in the limiting case where the parameter γ converges to unity, an OP policy elicits no effort and therefore no positive utility, hence it is dominated by a CP policy for any positive level of transaction costs.

The second insight relates to a main feature of the CP policy - wage dispersion. Workers with realized low wages reduce the 'benchmark' faced by their highly remunerated peers, thereby contributing to efforts and output. The presence of transaction costs limits the scope of wage dispersion, thus limiting the potential gains from a switch to a CP policy. In the absence of transaction costs, for instance, wage dispersion cannot hold in equilibrium. That is, in a frictions-less environment with zero transaction costs, the only pay policy that would be ex-post stable is an OP policy. However, if transaction costs are bounded away from zero, a CP policy becomes ex-post stable and can potentially sustain in equilibrium. The higher the transaction costs are, the larger the gains from switching to a CP policy. Thus, for a given degree of relative ambition and if transaction costs are sufficiently high, a CP policy prevails over an OP one.

4.2.1 Policy implications

As done in the context of partial equilibrium, we revisit the welfare implications of pay transparency regulations under the general equilibrium set-up. A straightforward observation that follows from the combination of the two parts of Theorem 4 is that for sufficiently high degree of relative ambition and large enough transaction cost, there exist only CP policy equilibria.

Observation 2. *For every $T \geqslant T^*$ and every $\gamma > \gamma(T^*)$, the only existing equilibria are confidential-pay policy ones.*

In the proof of Theorem 4 we have shown that an OP policy equilibrium does not exist due to a profitable deviation towards a CP policy. It therefore follows that a regulatory restriction which prohibits the use of pay secrecy clauses in wage contracts has a potential detrimental effect on workers' welfare. To see this consider the setting described in Observation 2, with a regulatory restriction of wage secrecy, such that firms are obligated to publish wages. Once CP polices become infeasible, the only supported equilibrium is the previously-unattainable suboptimal OP policy profile, which imposes an expected welfare loss to workers.

Observation 3. *For every sufficiently high transaction cost and relative ambition, the exclusion of* CP *policies lowers workers' expected utility.*

Observation 3 states that wage secrecy may be desirable from the workers' perspective. However, one should notice that wage secrecy entails ex-post wage differences amongst ex-ante homogeneous workers. Thus, in order to determine the social optimal, one should account for the trade-off between ex-ante efficiency considerations and ex-post equity concerns.

5 Concluding remarks

Our paper provides a positive explanation for a pay secrecy convention to arise in equilibrium. The key insight of our analysis hinges on the combination of the relative ambition exhibited by workers and the complementarity of the production function. A worker cares about the level of remuneration of a subset of her peers, which serves as her reference group and are complementary to her in the production process. The firm is employing a confidential pay policy to strike a balance between the desire to pay those in the reference group a high wage in order to elicit high productive efforts (by virtue of the complementarity) and the need to pay those workers a low wage in order to mitigate the dis-incentivizing effect of relative ambition. The idea underlying the wage secrecy policy is to use the workers outside the reference group as benchmarkers, serving to reduce the expected wage of the workers in the reference group without actually doing so. In our set-up, the worker's reference group was comprised of the members of her production team.

Clearly, a worker's reference group may be defined more broadly. Considering the employees' social network, one can assume that a worker's reference group is comprised of a weighted average of adjacent peers, where closer peers are weighted more heavily than others. Our set-up is, in essence, a specific version of this concept since only team members are positively weighted. Though the general set-up is left for future research, the key feature that would maintain our qualitative insights is that, as a whole, the reference group would exhibit sufficient complementarity with respect to the worker's productive effort.

Somewhat surprisingly, pay secrecy, which is often described as a strategic tool used by employers to improve their bargaining position in wage negotiations and as a means to mitigate the potentially demoralizing effect of pay gaps on employees, may actually improve employees' welfare. In particular, in case relative ambition considerations are sufficiently manifest, we demonstrate that in a general equilibrium setting with free entry (which implies full dissipation of firms' rents), a confidential pay policy would maximize the ex-ante utility of workers, as it serves to mitigate workers' relative ambition concerns.

Our focus in the paper was on the efficiency enhancing features of pay secrecy. Most of the popular debate on the desirability of pay transparency, however, revolves around equity aspects. A notable example is the ongoing public discourse on executive excessive pay (that hogged the limelight in the early 90's), which was the trigger for legislation mandating the disclosure of this information in financial statement of publicly traded firms and setting salary caps on executive levels of remuneration. More recently, the issue of pay transparency has resurfaced, in the context of gender pay gaps, where transparency has been suggested as a means to address persistent gender inequities in the labor market.[9]

To the extent that executives' high compensation schemes reflect economic rents (potentially driven by poor corporate governance) and gender gaps in the film industry are a byproduct of gender-based

[9] A recent exposure of e-mails by executives in Sony Pictures revealed substantial gender wage variation in wage contracts signed with top stars in the Film industry in Hollywood.

discrimination, pay transparency should be promoted as a means to reduce inequities without entailing efficiency costs (or better, mitigating those). Our analysis demonstrates, however, that pay secrecy may be desirable on efficiency grounds. Thus, determining the optimal extent of pay transparency involves resolving an equity-efficiency trade-off, which in our case is captured by the choice between an OP and a CP policy.

References

Akerlof, G. A. (1982): "Labor Contracts as Partial Gift Exchange," *The Quarterly Journal of Economics*, 97, 543.

Akerlof, G. A. and J. L. Yellen (1990): "The Fair Wage-Effort Hypothesis and Unemployment," *The Quarterly Journal of Economics*, 105, 255.

Blumkin, T. and D. Lagziel (2018): "Pay Secrecy in Labor Markets with Matching Frictions," .

Bracha, A., U. Gneezy, and G. Loewenstein (2015): "Relative Pay and Labor Supply," *Journal of Labor Economics*, 33, 297–315.

Breza, E., S. Kaur, and Y. Shamdasani (2018): "The Morale Effects of Pay Inequality," *The Quarterly Journal of Economics*, 133, 611–663.

Card, D., A. Mas, E. Moretti, and E. Saez (2012): "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction," *American Economic Review*, 102, 2981–3003.

Charness, G. and P. Kuhn (2007): "Does Pay Inequality Affect Worker Effort? Experimental Evidence," *Journal of Labor Economics*, 25, 693–723.

——— (2011): "Lab Labor: What Can Labor Economists Learn from the Lab?" *Handbook of Labor Economics*, 4, 229–330.

Cohn, A., E. Fehr, B. Herrmann, and F. Schneider (2011): "Social Comparison in the Workplace: Evidence from a Field Experiment," *IZA Discussion Paper No. 5550.*

Cullen, Z. B. and B. Pakzad-Hurson (2018): "Equilibrium Effects of Pay Transparency in a Simple Labor Market," .

Cullen, Z. B. and R. Perez-Truglia (2018): "How Much Does Your Boss Make? The Effects of Salary Comparisons," .

Danziger, L. and E. Katz (1997): "Wage Secrecy as a Social Convention," *Economic Inquiry*, 35, 59–69.

DellaVigna, S., J. List, U. Malmendier, and G. Rao (2016): "Estimating Social Preferences and Gift Exchange at Work," .

Esteves-Sorenson, C. (2018): "Gift Exchange in the Workplace: Addressing the Conflicting Evidence with a Careful Test," *Management Science*, 64, 4365–4388.

Fehr, E., G. Kirchsteiger, and A. Riedl (1993): "Does Fairness Prevent Market Clearing? An Experimental Investigation," *The Quarterly Journal of Economics*, 108, 437–459.

Fershtman, C., H. K. Hvide, and Y. Weiss (2003): "A Behavioral Explanation of the Relative Performance Evaluation Puzzle," *Annales d'Économie et de Statistique*, 349.

——— (2006): "Cultural Diversity, Status Concerns and the Organization of Work," in *The Economics of Immigration and Social Diversity (Research in Labor Economics)*, Emerald Group Publishing Limited, 361–396.

Fischer, S. and E.-M. Steiger (2009): "Exploring the effects of unequal and secretive pay," *Jena economic research papers*.

Frank, R. H. (1984a): "Are workers paid their marginal products?" *American Economic Review*, 74, 549–571.

——— (1984b): "Interdependent Preferences and the Competitive Wage Structure," *The RAND Journal of Economics*, 15, 510.

Gächter, S. and C. Thöni (2010): "Social comparison and performance: Experimental evidence on the fair wageeffort hypothesis," *Journal of Economic Behavior & Organization*, 76, 531–543.

Greiner, B., A. Ockenfels, and P. Werner (2011): "Wage transparency and performance: A real-effort experiment," *Economics Letters*, 111, 236–238.

Hennig-Schmidt, H., A. Sadrieh, and B. Rockenbach (2010): "In Search of Workers' Real Effort Reciprocitya Field and a Laboratory Experiment," .

Kallenberg, O. (2002): *Foundations of modern probability*, Springer.

Ku, H. and T. C. Salmon (2012): "The Incentive Effects of Inequality: An Experimental Investigation," *Southern Economic Journal*, 79, 46–70.

Perez-Truglia, R. (2015): "The Effects of Income Transparency on Well-Being: Evidence from a Natural Experiment," .

REGE, M. AND I. F. SOLLI (2015): "Lagging Behind the Joneses: The Impact of Relative Earnings on Job Separation," .

Appendices

A Proof of Theorem 1

Proof. We start with the case of perfect complements such that $Q(e, e_p) = \min\{e, e_p\}$. Take an OP policy where every matched couple is paid w and w_p w.p. 1. Due to the complementarity, it would be suboptimal to pay different wages, thus $w = w_p$, and $e = e_p = w(1-\gamma)$. Since $\gamma = 1$ is a trivial case, assume $\gamma \in \left(\frac{1}{2}, 1\right)$. This alternation is without loss of generality since the aggregate level of production is fixed. Thus, $\mathbf{E}[Q(e, e_p)] = w(1-\gamma) = \tilde{X}$ and $C = 2w = \frac{2\tilde{X}}{1-\gamma}$, where $\tilde{X} = H^{-1}(X)$.

Now, fix a CP policy where every employee receives either a wage of w w.p. $p \in (0, 1)$, or nothing, otherwise.[10] There is a probability of p^2 for a couple to be productive, hence $\mathbf{E}[Q(e, e_p)] = p^2(w - \gamma wp) = \tilde{X}$ and $C = 2pw = \frac{2\tilde{X}}{p(1-\gamma p)}$. Comparing the expected costs leads to the following sufficient condition $p(1-\gamma p) > 1-\gamma$. If p is sufficiently close to 1 (specifically, in case $p > \frac{1}{2\gamma}$), then the LHS of the inequality is decreasing, while $p = 1$ yields an equality. Thus, for $p > \frac{1}{2\gamma}$ and $\gamma > \frac{1}{2}$, the result follows.

We move on to the case of perfect substitutes such that $Q(e, e_p) = \frac{1}{2}e + \frac{1}{2}e_p$. Take an OP policy where every matched couple is paid (w, w_p) w.p. 1. A straightforward constrained optimization shows that asymmetric pay of either $w = 0$ or $w_p = 0$ is the optimal OP policy for every γ. This outcome is intuitive since costs and production are linear and asymmetric wages bypass the problem of comparison among employees. Therefore, we need to show that any CP policy produces an expected cost of at least $2\tilde{X}$, given production level $\tilde{X}$, since an OP $(w, 0)$-policy has an expected total cost of $2\tilde{X}$.

Take a CP policy F. Recall it is ex-ante symmetric w.r.t. all employees, thus they maintain the same benchmark of $\mathbf{E}[w] = \mathbf{E}[w_p]$. Denote $a = \gamma\mathbf{E}[w]$. The expected production and cost are given by $\tilde{X} = \mathbf{E}\left[(w-a)\mathbb{1}_{\{w \geqslant a\}}\right]$ and $C = 2\mathbf{E}[w]$. Define $Y = w - a$ and $q = \Pr(Y \geqslant 0)$. Let $Y_+ = \max\{Y, 0\}$ and $Y_- = \min\{Y, 0\} = Y - Y_+$ be the non-negative and non-positive parts of Y. Therefore, $\tilde{X} = \mathbf{E}[Y_+]$. In simple terms, the expected production equals the expectation of the marginal distribution above a.

Define an atomic distribution $\tilde{F} \in \Delta\mathbb{R}_+$ over the two values $\frac{\mathbf{E}[Y_+]}{q} + a$ and $\frac{\mathbf{E}[Y_-]}{1-q} + a$ (if $q = 1$, consider only the first term) with probabilities q and $1-q$, respectively. The distribution $\tilde{F}$ is, essentially, a contraction of F's marginals on the two sides of a, preserving the expected value $\mathbf{E}[w]$ and the probabilities of being above or below the benchmark a. Specifically, the expectation w.r.t. $\tilde{F}$ equals $\mathbf{E}[Y] + a = \mathbf{E}[w]$. Thus, an implementation of $\tilde{F}$, as a wage distribution with a random matching of employees, yields the same expected production of $\tilde{X} = \mathbf{E}[Y_+]$, and the same expected cost of $C = 2\mathbf{E}[w]$. So, the random

[10] Throughout the proof, we assume that fixed wages and probabilities meet the production requirement.

implementation of $\tilde{F}$ is equivalent to F, and (without loss of generality) we can restrict any CP policy to atomic marginal distributions supported on two values, above and below the benchmark level.

Consider an atomic distribution F' such that $w_i \geqslant 0$ is reached w.p. p_i, where $i = 1, 2$ and $p_1 + p_2 = 1$. Following the previous statement, let $\mu = w_1 p_1 + w_2 p_2$ be the expected value according to F', and assume that $w_1 < a = \gamma\mu \leqslant \mu \leqslant w_2$. Note that w_1-wage employees exert no effort, thus it is suboptimal to offer them a positive wage, and we can assume that $w_1 = 0$. An implementation of F' as a CP policy yields an expected production of $X = (w_2 - a)p_2 = w_2(1 - \gamma p_2)p_2$, and an expected cost of $C = 2w_2 p_2 = \frac{2\tilde{X}}{1-\gamma p_2}$. Hence, we need to verify that $\frac{2\tilde{X}}{1-\gamma p_2} \geqslant 2\tilde{X}$ or, equivalently, $\gamma p_2 \geqslant 0$. The last inequality is strict by the use of a CP policy (i.e., $p_2 > 0$) whenever $\gamma > 0$, and otherwise we get an equality, thus concluding the proof. ∎

B Proof of Lemma 1

Proof. Fix ρ and consider a CP policy F which marginals are supported on more than two values. Since it is ex-ante symmetric, all employees maintain the same benchmark of $\mathbf{E}[w] = \mathbf{E}[w_p]$. Denote $a = \gamma\mathbf{E}[w]$ and define the two random variables $Y = w - a$ and $Y^p = w_p - a$. We follow the standard $\pm$ notation for $Y_+ = \max\{Y, 0\}$ and for $Y_- = \min\{Y, 0\}$ such that $Y = Y_- + Y_+$. The expected production and cost are given by $\tilde{X} = \mathbf{E}[Q(Y_+, Y_+^p)]$ and $C = 2\mathbf{E}[w]$, respectively.

Denote $q = \Pr(Y > 0)$. Define an atomic distribution $\tilde{F}$ such that

$$W = \begin{cases} \frac{\mathbf{E}[Y_+]}{q} + a, & \text{w.p. } q, \\ \frac{\mathbf{E}[Y_-]}{1-q} + a, & \text{w.p. } 1-q, \end{cases}$$

for $W \sim \tilde{F}$ (in case $q = 1$, consider only the first term). The distribution $\tilde{F}$ is, essentially, a contraction of F's marginals on the two sides of a, which preserves several key values. Namely, for $Y^w = W - a$, it follows that

$$\begin{aligned}
\Pr\left(Y_+^w > 0\right) &= \Pr(W > a) = q = \Pr(w > a) = \Pr(Y_+ > 0); \\
\mathbf{E}[W] &= q\left[\tfrac{E[Y_+]}{q} + a\right] + (1-q)\left[\tfrac{E[Y_-]}{1-q} + a\right] = \mathbf{E}[w]; \\
\mathbf{E}\left[Y_+^w\right] &= \mathbf{E}\left[(W-a)_+\right] = q\left[\tfrac{E[Y_+]}{q} + a - a\right] = \mathbf{E}\left[Y_+\right]; \\
\mathbf{E}\left[Y_+^w | Y_+^w > 0\right] &= \frac{\mathbf{E}\left[Y_+^w\right]}{\Pr(Y_+^w > 0)} = \frac{\mathbf{E}\left[Y_+\right]}{\Pr(Y_+ > 0)} = \mathbf{E}\left[Y_+ | Y_+ > 0\right].
\end{aligned}$$

Thus, an implementation of $\tilde{F}$, as a wage distribution with a random matching of employees, preserves the expected cost along with $\mathbf{E}[Y_+]$ and a.

Now, fix a realization of Y^p and observe the expected production $\mathbf{E}[Q(Y_+, Y_+^p)]$.

$$\begin{aligned}
\mathbf{E}\left[Q\left(Y_+, Y_+^p\right)\right] &= (1-q)\mathbf{E}[Q(Y_+, Y_+^p)|Y_+ = 0] + q\mathbf{E}[Q(Y_+, Y_+^p)|Y_+ > 0] && (1)\\
&\leqslant (1-q)\mathbf{E}[Q(Y_+, Y_+^p)|Y_+ = 0] + qQ\left(\mathbf{E}[Y_+|Y_+ > 0], Y_+^p\right) && (2)\\
&= (1-q)\mathbf{E}[Q(Y_+^w, Y_+^p)|Y_+^w = 0] + qQ\left(\mathbf{E}[Y_+^w|Y_+^w > 0], Y_+^p\right) && (3)\\
&= (1-q)\mathbf{E}[Q(Y_+^w, Y_+^p)|Y_+^w = 0] + q\mathbf{E}\left[Q(Y_+^w, Y_+^p)|Y_+^w > 0\right] && (4)\\
&= \mathbf{E}\left[Q\left(Y_+^w, Y_+^p\right)\right], && (5)
\end{aligned}$$

where Eqs. (1) and (5) follow from the law of total expectation; Ineq. (2) follows from the concavity of the CES function and Jensen inequality; Eq. (3) follows from the preserving qualities of W and Y^w; and Eq. (4) follows from the fact that $Y_+^w|Y_+^w > 0$ is constant. A similar computation holds for Y^p. Therefore, the random implementation of $\tilde{F}$ dominates F (potentially weakly) by preserving the expected cost and increasing production. We conclude our analysis of CP policies. Since any OP policy is based on a deterministic matching of employees, it is optimal to uniquely implement the most efficient couple's wage-combination, thus maintaining at most two wage levels and the result follows. ■

C Proof of Theorem 2

Proof. The computation is trivial for $\gamma \in \{0, 1\}$, thus we consider $\gamma \in (0, 1)$, and denote $\tilde{X} = H^{-1}(X)$. We begin by explicitly writing the expected cost, in terms of production, for every policy. By Lemma 1 we can restrict ourselves to, at most, two pay levels. Note that the observable policy must be symmetric, since the overall production is set to zero once at least one agent is effortless. The expected cost and production from a symmetric OP policy and a CP policy (where a positive wage w is paid w.p. p) are:

$$\begin{aligned}
Q_{\text{OP}} &= \left[\tfrac{1}{2}(w-\gamma w)^\rho + \tfrac{1}{2}(w-\gamma w)^\rho\right]^{1/\rho} = w(1-\gamma) = \tilde{X},\\
C_{\text{OP}} &= 2w = \frac{2\tilde{X}}{1-\gamma},\\
Q_{\text{CP}} &= p^2\left[\tfrac{1}{2}(w-\gamma wp)^\rho + \tfrac{1}{2}(w-\gamma wp)^\rho\right]^{1/\rho} = p^2 w(1-\gamma p) = \tilde{X},\\
C_{\text{CP}} &= 2pw = \frac{2\tilde{X}}{p(1-\gamma p)},
\end{aligned}$$

similarly to the proof of Theorem 1. Note that the symmetric OP policy is nested in the CP policy given $p = 1$. Hence, the optimal policy is reached by maximizing the denominator of C_{CP}, and the CP policy is optimal whenever $p < 1$. Optimizing $p(1-\gamma p)$ w.r.t. $p \in (0, 1]$, we get that the optimal probability is $p = \min\{\frac{1}{2\gamma}, 1\}$. Thus, the CP policy is optimal if and only if $\gamma > 1/2$, as needed. ■

D Proof of Theorem 3

Proof. The computation for $\gamma = 0$ is trivial, showing that a symmetric OP policy is optimal. Thus, we consider $\gamma \in (0,1]$. Differing from Theorem 2, we also need to consider a non-symmetric OP policy where one agent receives a positive wage of w, while the other receives nothing. The expected cost and production from such a policy are $Q_{\text{OP}-\text{Asym}} = \left[\frac{1}{2}(w-\gamma\cdot 0)^{\rho} + \frac{1}{2}\cdot 0\right]^{1/\rho} = \frac{w}{2^{1/\rho}} = \tilde{X}$ and $C_{\text{OP}-\text{Asym}} = w = 2^{1/\rho}\tilde{X}$, where $\tilde{X} = H^{-1}(X)$. Recall, from Theorem 2, that the symmetric OP policy yields an expected cost of $C_{\text{OP}-\text{Sym}} = \frac{2\tilde{X}}{1-\gamma}$. On the other hand, the expected production and cost of a CP policy, where a wage of w is paid w.p. $p > 0$, are given by

$$\begin{aligned}
Q_{\text{CP}} &= p^2 Q(w-\gamma wp, w-\gamma wp) + 2p(1-p)Q(w-\gamma wp, 0) \\
&= p^2 w(1-\gamma p) + 2p(1-p)\frac{w-\gamma wp}{2^{1/\rho}} \\
&= wp(1-\gamma p)\left[p + 2^{1-1/\rho}(1-p)\right] = \tilde{X},
\end{aligned}$$

and $C_{\text{CP}} = 2pw = \frac{2\tilde{X}}{(1-\gamma p)\left[p+2^{1-1/\rho}(1-p)\right]}$. In case $\rho = 1$, we get $C_{\text{CP}} = \frac{2\tilde{X}}{1-\gamma p}$, and the asymmetric OP policy is superior for every $\gamma > 0$. Thus, define $\gamma_1(1) = \gamma_2(1) = 0$, and henceforth assume that $\rho \in (0,1)$.

First, we study the function $h_{\gamma,\rho}(p) = (1-\gamma p)\left[p + 2^{1-1/\rho}(1-p)\right]$. Denote the probability that maximizes $h_{\gamma,\rho}$ (and minimizes C_{CP}) by $\text{P}_{\gamma,\rho}$. Since $h_{\gamma,\rho}(p)$ is a parabolic function, it follows that $\text{P}_{\gamma,\rho} = 1$ if and only if $h'_{\gamma,\rho}(1) \geqslant 0$. The sign of $h'_{\gamma,\rho}(1) = 1 - 2^{1-1/\rho} + \gamma\left(2^{1-1/\rho} - 2\right)$ is crucial for our analysis since the symmetric OP policy is embedded in the CP policy, and the policies coincide in case $\text{P}_{\gamma,\rho} = 1$.

Define $\gamma_1(\rho)$ based on the indifference curve $h'_{\gamma,\rho}(1) = 0$. The LHS of $h'_{\gamma,\rho}(1) = 0$ is continuous and decreasing in γ and ρ, so every feasible (γ,ρ) above the curve maintains $h'_{\gamma,\rho}(1) < 0$ such that the CP policy is superior to the symmetric OP policy, as needed. The properties of $h'_{\gamma,\rho}(1)$ along with the values $\gamma_1(0) = 1/2$ and $\gamma_1(1) = 0$ assure that the function γ_1 is well-defined and continuously-decreasing.

We move on to γ_2. The asymmetric OP policy is (weakly) superior to the CP policy if and only if $2^{1/\rho-1}h_{\gamma,\rho}\left(\text{P}_{\gamma,\rho}\right) \leqslant 1$. A computation of $\text{P}_{\gamma,\rho}$ yields $\text{P}_{\gamma,\rho} = \frac{2^{1/\rho-1}-1-\gamma}{2\gamma\left(2^{1/\rho-1}-1\right)}$. In case, $\gamma \geqslant 2^{1/\rho-1}-1$, it follows that $\text{P}_{\gamma,\rho} \leqslant 0$, which means that positive wages are not distributed and production is impossible. Thus, any (γ,ρ) above the curve $\gamma = 2^{1/\rho-1}-1$ is irrelevant for a CP policy, and we can henceforth restrict ourselves to the area below the curve.

Denote $t = 2^{1-1/\rho}$, and plug $\text{P}_{\gamma,\rho} = \frac{1-t-t\gamma}{2\gamma(1-t)}$ into the function $2^{1/\rho-1}h_{\gamma,\rho}(p)$:

$$\begin{aligned}
2^{1/\rho-1}h_{\gamma,\rho}\left(\text{P}_{\gamma,\rho}\right) &= 2^{1/\rho-1}\left[1-\gamma\text{P}_{\gamma,\rho}\right]\left[2^{1-1/\rho} + \text{P}_{\gamma,\rho}(1-2^{1-1/\rho})\right] \\
&= t^{-1}\left[1-\gamma\tfrac{1-t-t\gamma}{2\gamma(1-t)}\right]\left[t + \tfrac{1-t-t\gamma}{2\gamma(1-t)}(1-t)\right] \\
&= t^{-1}\left[1-\tfrac{1-t-t\gamma}{2(1-t)}\right]\left[t+\tfrac{1-t-t\gamma}{2\gamma}\right] \\
&= t^{-1}\left[\tfrac{1}{2}-\tfrac{-t\gamma}{2(1-t)}\right]\left[\tfrac{t}{2}+\tfrac{1-t}{2\gamma}\right] \\
&= \tfrac{1}{4}\left[1+\tfrac{t\gamma}{1-t}\right]\left[1+\tfrac{1-t}{t\gamma}\right].
\end{aligned}$$

Thus, $2^{1/\rho-1}h_{\gamma,\rho}\left(\mathrm{P}_{\gamma,\rho}\right) = 1$ iff $\frac{t\gamma}{1-t} = 1$. The last equality translates to $\gamma = 2^{1/\rho-1} - 1$, which is exactly the curve that defines the non-positive probability-restriction $\mathrm{P}_{\gamma,\rho} = 0$.

Define $\gamma_2(\rho) = 2^{1/\rho-1} - 1$ and note that $\gamma_2(0.5) = 1$, $\gamma_2(1) = 0$. We now need to show that below the curve, the CP policy is superior to the asymmetric OP policy. Given $p > 0$, the function $2^{1/\rho-1}h_{\gamma,\rho}(p)$ is point-wise decreasing in ρ and γ. Thus, the function $2^{1/\rho-1}h_{\gamma,\rho}\left(\mathrm{P}_{\gamma,\rho}\right) = \max_{p\in(0,1]}\left[2^{1/\rho-1}h_{\gamma,\rho}\left(p\right)\right]$ is decreasing in both arguments, by virtue of an envelope argument. Let $S = \{(\gamma,\rho) \in [0,1]^2 : 2^{1/\rho-1}h_{\gamma,\rho}\left(\mathrm{P}_{\gamma,\rho}\right) \leqslant 1\}$ be the set of points (γ,ρ) such that asymmetric OP policy is superior to the CP policy. Hence, for every $(\gamma,\rho) \in S$, it follows that $(\gamma',\rho') \in S$ where $\gamma' \geqslant \gamma$ and $\rho' \geqslant \rho$. By the continuity and the monotonicity of $2^{1/\rho-1}h_{\gamma,\rho}\left(\mathrm{P}_{\gamma,\rho}\right)$, we establish that γ_2 is continuously-decreasing and, given ρ, the CP policy is optimal if and only if $\gamma_1(\rho) < \gamma < \gamma_2(\rho)$, as required. Note that the second part of the theorem follows directly from the definition of γ_1 and γ_2. ∎

E A generic effort-response function

To facilitate the interpretation of Theorem 1, one may abstract from our specific utility function and modelling choices, and consider a generic effort-response function $e(w, w_p)$ which dictates the optimal response of an employee. Starting with a standard set-up with no reference level ($\gamma = 0$, in our formulation), one expects an increasing and weakly concave effort level with respect to the individual wage. Taking into account the concave and symmetric production function, the optimal OP policy calls for wage compression. These assumptions and result arise naturally in the classic set-up, and completely change once relative ambition is embedded into the framework.

The existence of a reference point introduces a fixed cost to the firm's optimization problem. Namely, a worker exerts additional effort, above the base level, only if he is sufficiently compensated relative to the reference point (in our case, above γw_p). If the fixed cost is large enough, there are increasing returns to scale and the firm can do better by assigning the production to a subgroup of workers, thereby saving some of the fixed costs. The mathematical manifestation of a substantial fixed cost is reflected through an average-convexity property of the effort-response function. These sufficient conditions are summarized in Claim 1.

Claim 1. *Consider the case of perfect complements, and fix the employees' non-negative effort response function $e = e(w, w_p)$, where $e_w > 0 > e_{w_p}$. Given equal wages, assume that the effort function is increasing, and assume it is sufficiently convex (on average) in its first argument such that $e_w(w,w) > 2\frac{e(w,w)-e(0,w)}{w}$. Then, the* OP *policy is suboptimal.*

Note that all monotonicity assumptions are quite orthodox in either framework. The response function increases in individual wage w, decreases in the partner's wage w_p, and increases in wages subject to equal pay. The key ingredient for our result is the convexity assumption which follows from the previously mentioned fixed cost, due to relative ambition and the use of partners as reference points.

Proof. Under the Leontief production function $Q(e, e_p) = \min\{e, e_p\}$ and given the monotonicity properties of e, the optimal OP policy dictates equal wages such that $e(w, w) = X$ and $C = 2w$. Consider a CP policy where every employee receives either a wage of w w.p. $p \in (0, 1)$, or nothing, otherwise. The expected production is $p^2 e(w, wp) + (1 - p^2)e(0, wp) = X$ and the expected cost is $C(w, p) = 2wp$. Note that the optimal OP solution is embedded in the CP analysis for $p = 1$, so it is sufficient to show that the expected cost is increasing at $p = 1$, which implies that a strict CP policy where $p < 1$ is superior.

The equation $p^2 e(w, wp) + (1 - p^2)e(0, wp) - X = 0$ defines w as an implicit function of p. Hence,

$$w'(p) = -\frac{2pe(w, wp) + wp^2 e_{w_p}(w, wp) - 2pe(0, wp) + w(1 - p^2)e_{w_p}(0, wp)}{p^2 e_w(w, wp) + p^3 e_{w_p}(w, wp) + p(1 - p^2)e_{w_p}(0, wp)}.$$

Differentiating the expected cost, plugging in $w'(p)$, and taking $p = 1$, we get

$$\begin{aligned}
[2w'(p)p + 2w]_{p=1} &= \frac{-4e(w, w) - 2we_{w_p}(w, w) + 4e(0, w)}{e_w(w, w) + e_{w_p}(w, w)} + 2w \\
&= \frac{-4e(w, w) + 4e(0, w) + 2we_w(w, w)}{e_w(w, w) + e_{w_p}(w, w)} \\
&> \frac{-4e(w, w) + 4e(0, w) + 4e(w, w) - 4e(0, w)}{e_w(w, w) + e_{w_p}(w, w)} = 0,
\end{aligned}$$

where the inequality follows from the conditions over e. Note that the denominator is strictly positive by the assumption that $e(w, w_p)$ is increasing, subject to equal wages. So, the expected cost is increasing at $p = 1$, and the result follows. ■

F Proof of Theorem 4

Proof. We begin our proof with the first part of the theorem. Fix $T > 0$ and a symmetric profile of OP policies where workers' expected utility is $\tilde{U}$. As mentioned, wage compression is evident under OP policies, which guarantees ex-post stability and eliminates the second constraint of the optimization problem. Moreover, the first constraint must be binding as one would increase w to maximize the goal function. Thus, we can extract $w = \frac{1-\gamma}{4}$ and plug into the expected utility to get

$$\mathbf{E}[U(e, w; w_p)] = \frac{1 - \gamma}{4} + \frac{(1 - \gamma)^4}{32}.$$

The latter is a non-negative decreasing function w.r.t. γ, such that $\lim_{\gamma \to 1} \mathbf{E}[U(e, w; w_p)] = 0$.

We shall prove that any firm has a profitable deviation by constructing a specific CP policy. For that purpose, consider a policy of two wage levels, 0 and w, w.p. $1 - p$ and p, respectively. The stability problem, subject to the mentioned profile, is

$$\begin{aligned} \max_{w>0, p\in[0,1]} \quad & wp + p\frac{(w-\gamma wp)^2}{2}, \\ \text{s.t.} \quad 0 & \geqslant 2wp - \left[p^2(w-\gamma wp)\right]^{1/2}, \\ T & \geqslant \tilde{U}. \end{aligned}$$

Henceforth assume that the first constraint in binding. Notice that this problem incorporates a (limited) CP regime and, without loss of generality, any optimal OP policy. Again, we can extract $w = \frac{1-\gamma p}{4}$ and plug into the expected utility to get

$$\mathbf{E}[U(e, w; w_p)] = \frac{p(1-\gamma p)}{4} + \frac{p(1-\gamma p)^4}{32}.$$

Denote the last function by $G(p, \gamma)$, such that $G(1, \gamma) = \tilde{U}$. Note that $G(p, \gamma)$ is a polynomial in p and in γ on a compact set $[0, 1]^2$, so it is uniformly bounded and a solution exists.

If $G(1, \gamma) > T$, then the second constraint of the limited CP problem cannot hold. Therefore, a necessary condition for deviating is $G(1, \gamma) \leqslant T$. Moreover, if $G(1, \gamma) < T$ and the problem above dictates an optimal probability $p_\gamma < 1$, then there exists a CP policy that dominates the optimal OP policy. A sufficient condition for $p_\gamma < 1$ is $\frac{\partial G(1,\gamma)}{\partial p} < 0$. Evidently, $\frac{\partial G(p,\gamma)}{\partial p} = \frac{1-2\gamma p}{4} + (1-\gamma p)^3 \left[\frac{1-5\gamma p}{32}\right]$, and the condition holds for $\gamma \geqslant 0.5$.

We are left with the problem of estimating γ such that $G(1, \gamma) < T$. The function $G(1, \gamma)$ is decreasing in γ, and $G(1, 1) = 0$. Hence, for every $T > 0$, we can define a value $\gamma_T < 1$ such that the inequality $G(1, \gamma) < T$ holds if and only if $\gamma > \gamma_T$. The monotonicity of $G(1, \gamma)$ suggests that γ_T also decreases in T, until T is sufficiently large such that $\gamma_T \leqslant 0$. Define the function $\gamma(T) = \max\{0.5, \gamma_T\}$. For every $\gamma > \gamma(T)$, the optimal probability is bounded away from 1 and the optimal OP policy falls short of the upper bound T, proving the existence of a CP policy superior to the optimal OP policy, as needed.

We turn to the second part of the theorem, proving the existence of a symmetric CP policy equilibrium, given a fixed finite number of bounded wage levels. We do so by solving the problem while (momentarily) ignoring the second constraint of ex-post stability. The assumptions over wages impose a compact set in the euclidean space, while the objective function is continuous, and a solution exists for every $\gamma \geqslant 0$. Denote the workers' maximal expected utility (uniformly for all possible values of γ) by $\tilde{U}$, and consider a profile of policies where all firms follow the same optimal policy. Fix $T^* = \tilde{U}$ such that the second constraint holds uniformly for every γ. Clearly, all firms are optimizing w.r.t. the given problem, and no firm has an incentive to deviate, establishing an equilibrium. In addition, the first part of the theorem shows that for $\gamma > \gamma(T^*)$ the optimal policy is not OP policy, and $\gamma(T)$ is weakly decreasing in T, so the statement holds for every $\gamma > \gamma(T^*)$ and every $T > T^*$, as needed. ■