Contreras, Óscar F.; Zanarone, Giorgio

Working Paper

Managing social comparison costs in organizations

Working Papers, No. 2018-25

Provided in Cooperation with:
Bank of Mexico, Mexico City

Suggested Citation: Contreras, Óscar F.; Zanarone, Giorgio (2018) : Managing social comparison costs in organizations, Working Papers, No. 2018-25, Banco de México, Ciudad de México

This Version is available at:
https://hdl.handle.net/10419/240666

Banco de México

Documentos de Investigación

Banco de México

Working Papers

N° 2018-25

Managing Social Comparison Costs in Organizations

Oscar F. Contreras
Banco de México

Giorgio Zanarone
CUNEF

December 2018

Documento de Investigación
2018-25

Working Paper
2018-25

Managing Social Comparison Costs in Organizations*

Oscar F. Contreras†
Banco de México

Giorgio Zanarone‡
CUNEF

Abstract: This paper studies how organizations manage the social comparisons that arise when their employees' pay and tasks, and hence their status vis-à-vis peers, differ. We show that under a "pay transparency policy", the organization may compress pay and distort the employees' tasks to minimize social comparison costs. We subsequently show that if the organization can credibly commit to informal agreements, it may remove social comparisons by implementing a "pay secrecy" policy. Under such a policy, the organization makes employees "officially equal" by granting them similar formal terms, while optimally differentiating their pay through self-enforcing informal adjustments.
Keywords: Social Comparisons, Organization Design, Informal Contracts, Formal Contracts.
JEL Classification: D03, D23, M52, M54.

Resumen: En este documento se estudia cómo las organizaciones gestionan las comparaciones sociales que surgen cuando el salario y las tareas de sus empleados, y por lo tanto su "estatus social" dentro de la organización, difieren. Se demuestra que bajo una "política de transparencia salarial", la organización puede verse obligada a comprimir los salarios y a distorsionar las tareas de sus empleados para minimizar los costos de comparación social. Posteriormente, se demuestra que si la organización puede comprometerse de forma creíble a acuerdos informales, esta podrá eliminar los efectos negativos generados por las comparaciones sociales implementando una política de compensación "privada". Bajo dicha política, la organización ofrece contratos formalmente idénticos a los empleados, a la vez que emplea transferencias privadas e informales para personalizar de manera óptima su compensación.
Palabras Clave: Comparaciones Sociales, Diseño Organizativo, Contratos Formales e Informales.

*We thank Benito Arruñada, Guillermo Caruana, Sylvain Chassang, Kenneth Corts, Florian Englmaier, Miguel Espinosa, Matthias Fahn, Guido Friebel, Willie Fuchs, Maija Halonen, Michael Kosfeld, Gerard Llobet, Bentley MacLeod, Marian Moszoro, Jens Prüfer, Rafael Repullo, Alan Schwartz, Brian Silverman, Pablo Spiller and Adam Szeidl for useful comments. This study received financial support from the Spanish Ministry of Economy and Competitiveness (ECO2017-85763-R).

† Dirección General de Investigación Económica. Email: oscar.contreras@banxico.org.mx.

‡ CUNEF. Email: gzanarone@cunef.edu.

1. Introduction

Organizations are riddled with social comparisons: employees dislike having a lower status than their peers, and press for disparities to be removed. Examples of social comparisons, and their effects on employee satisfaction and firm policies, abound in business history and practice. For instance, Nickerson and Zenger (2008) report that faculty at a North American business school opposed overload compensation for "star" professors, despite knowing that those professors received extra income from external activities.

In a recent string of field experiments, employees in groups with unequal pay significantly reduced productivity and reported lower levels of job satisfaction relative to similar employees in equal pay groups. These effects are strikingly similar across the diverse jobs, organizations and locations under study, which range from salesman positions at a German service firm (Cohn et al., 2014) to manufacturing worker positions at an Indian textile firm (Breza et al., 2018) to staff and faculty positions at a US state University (Card et al., 2012). The effects are also strikingly *asymmetric*: in all experiments, employees with *lower pay* than their peers exhibited reductions in productivity and job satisfaction, whereas no changes in productivity and job satisfaction were observed for employees with higher pay than their peers.[1]

Moving from employee behavior to managerial policies, multiple authors provide evidence consistent with the idea that social comparisons conduce to pay distortions. For instance, Baker, Gibbs and Holmstrom (1994) report that at a US service firm, managers in the higher wage quartiles had received systematically lower salary raises than their peers in lower quartiles. More recently, Ockenfels et al. (2015) study managerial compensation at a multinational firm and show that bonuses in its German division, where legal rules require pay transparency, are significantly more compressed than bonuses in the US division, where the pay transparency policy does not apply.

In this paper, we develop a formal model to study how organizations manage *social comparison costs*, such as those described above, through their choice of governance structure. Our analysis is based on three building blocks. First, we argue that social comparisons in an

[1] See Shaw (2014) for an interdisciplinary literature review on the effects of pay inequality on employee satisfaction and performance.

organization are especially acute when the members' relative status—in terms of pay and job position—is *publicly visible and explicit*. The importance of ostensible differences in status, as opposed to conjectural ones, is consistent with the psychological literature on self-serving beliefs (e.g., Kunda, 1990), according to which individuals may replace unpleasant beliefs (in our case, on having a lower status) with more favorable ones so long as there is hard evidence (in our case, public information on compensation and job position) supporting the substitution. Our view of social comparisons is also consistent with the aforementioned evidence in Card et al. (2012), showing that the satisfaction of lower-pay faculty and staff at University of California campuses (many of whom were long-term employees and hence were likely to have plausible conjectures on pay differences) decreased sharply once these were informed of the existence of a website publishing state employees' compensation.

As a second building block for our theory, we argue that an employee's relative status is more likely to be publicly visible and explicit when pay levels and job positions are formalized in a contract, compared to the case where they are implicitly agreed. Formal contract terms may be more easily disclosed or exchanged, and may also become public as a byproduct of litigation (e.g., Ben-Shahar and Bernstein, 2000) or mandatory disclosure rules such as those that apply to executive compensation in the U.S. (Gillan et al., 2009).

Third, and following a well-established literature, we argue that formal employment contracts are more credible than informal ones because they can be enforced by third parties, such as courts, whereas informal contracts must be self-enforcing—that is, they must rely on a credible threat of punishment by the employee in case the firm does not pay as promised (see MacLeod, 2007, Malcomson, 2013, and Gil and Zanarone, 2017, 2018, for up-to-date reviews of the theoretical and empirical literatures on informal contracting). When an employee has limited discretion (i.e., she cannot significantly lower future performance without being held accountable) and value to the firm, her punishment threat may not be credible, implying that an informal employment contract will not commit the firm and hence will not be accepted by the employee. In that case, the enhanced credibility of formal contracting is valuable to the firm. Therefore, organizations face a potential tradeoff between committing to formal terms of employment and avoiding social comparison costs by using less credible informal terms.

To analyze this trade-off, we model a simple organization that consists of a principal and two agents, each performing a partially contractible task in exchange for compensation. The agents are equally productive but have different outside options—for instance, because of varying personal constraints or firm-specific skills—and hence command different contract terms. However, explicit inequalities in compensation or task assignment may cause the low status agent to suffer disutility (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). Following Hart and Moore (2008), we assume that when the agents are frustrated, they retaliate against the organization by withdrawing the non-contractible part of their performance until either their frustration is fully compensated or their ability to withdraw performance without being caught by the principal is exhausted.

We begin by studying a "transparency" policy where the principal publicly announces the agents' compensation levels. We show that under transparency, the principal may assign the agent with low outside option an inefficiently easy task to reduce her ability to retaliate. At the same time, the principal may compress the agent's compensation upwards, relative to a setting with no social comparisons, in order to reduce the agent's frustration and the ensuing wage premium.

Next, we analyze a "privacy" policy where the principal grants identical "official" status to the two agents by equalizing their formal (and hence publicly observed) compensation levels, while adjusting the salary of the agent with high outside option upwards through an implicit, informally agreed payment that is not communicated to the other agent and hence does not generate social comparisons. We show that when privacy is considered, the possibility of retaliation acts as a double-edged sword. On one hand, when the agents' willingness and power to retaliate is too low, such that the firm cannot commit to honor informal compensation agreements, privacy is not sustainable. Under these circumstances, a marginal increase in retaliation power exacerbates the agents' reaction against negative social comparisons and the ensuing pay and task distortions. On the other hand, when the agents have high enough retaliation power, the principal can commit to pay informal salaries and hence to a privacy policy. In that case, a marginal increase in retaliation power reduces pay and task distortions, potentially allowing the firm to commit to a first best privacy policy in which social comparisons are fully removed.

Our model has counterintuitive implications for organizational design and policy. First, it predicts that as the organization moves from transparency (low retaliation power) towards privacy (high enough retaliation power), *formal* wage compression should *increase*, whereas *actual* wage compression should *decrease*. Second, the model predicts that as the organization moves from transparency to privacy, it is less likely to distort its internal task allocation, organizational architecture, and boundaries as means to reduce social comparisons. On one hand, we show that as argued by Nickerson and Zenger (2008), organizations that use pay transparency may end up separating employees, and hence reducing social comparisons among them, by inefficiently splitting them among different departments, or by outsourcing to external partners activities that would be optimally operated in-house. On the other hand, we show that organizations that rely on pay privacy do not need to resort to these distortionary policies, because they can use homogeneous formal pay levels to eliminate social comparisons, while relying on informal private agreements to optimally differentiate pay.

Finally, our model provides a rationale for pay secrecy norms. While at times criticized (e.g., Futrell, 1978; Burkus, 2016), pay secrecy appears resilient, especially in the U.S. (Edwards, 2005; Hill, 2016; Ockenfels et al., 2015). In our model, secrecy in informal compensation adjustments, combined with homogeneity in formal base salaries, allows an organization to optimally customize pay without triggering potentially disruptive social comparisons among its employees.[2]

The rest of the paper is organized as follows. Section 2 discusses our contributions to the literature on organizational design. Section 3 presents our baseline model of social comparison costs in organizations. Section 4 analyzes social comparison costs under transparency. Section 5 analyzes social comparison under privacy. Section 6 analyzes how social comparisons affect the organization's choice of compensation policy, firm boundaries, and internal architecture. Section 7 concludes.

[2] For an alternative explanation of pay secrecy as a device to reduce labor mobility, see Danziger and Katz (1997).

2. **Relation to the Literature**

Our paper belongs to a small but growing literature in organizational economics and strategy, which analyzes how non-standard and social preferences affect organizations. In a seminal paper, Hart and Moore (2008) argue that incomplete formal contracts serve as reference points for what the parties can expect to bargain, thus limiting frustration and conflict in the relationship. Hart and Moore (2008) use their model to analyze the tradeoff between rigid and flexible pricing terms and the optimal allocation of authority in employment contracts. In subsequent papers, Hart and coauthors build on the reference point idea to analyze asset ownership (Hart, 2009), firm scope (Hart and Holmstrom, 2010), and the optimal degree of contractual incompleteness (Halonen and Hart, 2013). Our model differs from this literature in two important ways. First, we study a different rationale for formal contracts—namely, homogenizing the perceived relative status of an organizations' members. Second, and most important, we explore the interaction between formal and informal contract terms in managing social comparison costs.[3]

In the strategic management literature, Zanarone et al. (2016) analyze a model where suppliers derive satisfaction from punishing uncompromising clients, and thus can credibly threaten to reveal confidential information on their clients to negotiate price increases. Zanarone et al. (2016) use their model to study how fixed price contracts and information disclosure policies may be used to discourage excess information acquisition. More related to our paper, Nickerson and Zenger (2008) argue that social comparison costs are more likely to arise within firms than between, and study how the formal governance and the boundaries of firms may be chosen to mitigate such costs. We contribute to their important insight by studying the complementary role of formal and informal contracts in managing social comparison costs, and methodologically, by embedding our analysis into a formal model that allows us to precisely identify the mechanisms underlying such complementarity.

A related literature analyzes incentive contracts in the presence of fairness concerns, both when performance is verifiable (e.g., Englmaier and Wambach, 2010; Englmaier and Leider,

[3] See Fehr et al. (2015) for some experimental evidence on the interaction between formal and informal contracts in shaping reference points.

2012) and when it is non-verifiable (e.g., Kragl and Schmid, 2009; Bartling and von Siemens, 2010; Kragl, 2015). One insight from this literature is that fairness concerns have an ambiguous effect on incentive compensation, as they may induce low performers to restore equity by increasing their effort. Consistent with this line of research, Bradler et al. (2016) show evidence that rewarding high performers increases the productivity of low performers. Unlike our paper, this literature does not explore the role of contracts as mechanisms to manage social comparison costs in organizations.

Our paper also relates to the literature on the interaction between formal and informal contracts (e.g., Klein, 2000; Poppo and Zenger, 2002; Baker, Gibbons, and Murphy, 1994, 2002; Battigalli and Maggi, 2008; Kvaløy and Olsen, 2009; Ryall and Sampson, 2009; Zanarone, 2013).[4] In this literature, the parties rely on informal agreements because writing, verifying or enforcing formal contracts is costly. At the same time, the parties may want to use some formal contract terms to realign incentives and reduce reneging temptations—that is, formal and informal contracts may be complements, rather than substitutes. We add two important insights to this literature. First, we explore a novel reason for the complementarity between formal and informal contracts—namely, managing social comparison costs—which *does not* depend on whether contract terms are verifiable. Second, and similarly to another recent paper (Fahn et al., 2017), we show how a firm's risk to dissipate the employees' intrinsic motivation to perform, as discussed in the reciprocity literature (Akerlof, 1982) and in the reference points literature (Hart and Moore, 2008), may enable enforcement of informal contracts even in settings where indefinitely repeated interaction is not feasible, and consequently the threat of relationship termination is not a credible deterrent against breach.

Finally, our paper relates to the literature on compensation policy and pay compression in organizations. Akerlof and Yellen (1990) analyze a model in which workers reduce effort when they perceive that they have received an unfair wage. MacLeod (2003) shows that when the performance of employees is subjectively evaluated, compensation is more compressed than when it is objectively evaluated because employees retaliate against the organization when they receive negative evaluations. Consistent with these theoretical works, empirical studies and

[4] See Gil and Zanarone (2018) for a discussion of the predictive power and empirical relevance of this literature.

case studies, such as Baker, Gibbs, and Holmstrom (1994), and Hall (2000), document wage compression policies in firms. We contribute to this literature by differentiating between formal and informal pay compression. Relatedly, we show that formal pay compression—and more generally, homogeneity in contract terms across the organization's members—may be used to decrease social comparison costs without necessarily translating into actual compression and homogeneity.

3. The Model

3.1. The Organization

We study an organization with one principal and two agents who interact for one period.[5] We may interpret the principal and the two agents as an employer and his employees, or alternatively, as a client firm and its distributors or suppliers.

We make a number of assumptions on the interaction between the principal and the two agents, their information and their preferences, some of which are non-standard. While we provide references, intuition and some empirical underpinnings for our assumptions as we state them, we defer a detailed discussion of these assumptions, and of the potential consequences of relaxing them, to section 5.3, till after the model is fully specified and analyzed.

Production

At the outset the principal may ask each agent to perform a task a_i, where $a_i \in \mathbb{R}^+$ defines both the task and its difficulty. For instance, if the two agents are managers, a_i may denote the size of the team manager i supervises, the number of products his division is responsible for, or the complexity of such products. Following Hart and Moore (2008), we assume that successful completion of the task requires two separate actions: "hard" performance, denoted as $d_i(a_i)$, and "soft" performance, denoted as $x_i(a_i)$. To simplify the analysis, we assume that while soft performance can take any value on the interval $[0,1]$, hard performance is restricted to the values $\{0,1\}$.

[5] See Contreras and Zanarone (2017) for an extension of this model to the case where the principal and the agents interact repeatedly.

We assume that hard performance is essential for value creation and that value increases in the extent of soft performance, such that the individual output generated by agent i for the principal is given by:

$$f_i \equiv d_i(a_i)x_i(a_i)y(a_i),$$

where the function $y(\cdot)$ is continuous, strictly increasing and concave, and satisfies $y(0) = 0$, $\lim_{a_i \to 0} y'(a_i) = \infty$, and $\lim_{a_i \to \infty} y'(a_i) = 0$.

Following Hart and Moore (2008), we assume that while providing hard performance is costly for the agents, soft performance does not impose additional costs and may even give the agents some satisfaction (which we assume for simplicity to be small and therefore ignore in the mathematical analysis). Formally, the performance cost of agent i is given by:

$$k_i \equiv d_i(a_i)C(a_i),$$

where we assume that $C(\cdot)$ is continuous, strictly increasing, strictly convex, and satisfies $C(0) = 0$.

To illustrate, if agent i is a manager and a_i denotes his managerial task, hard performance ($d_i(a_i) \in \{0,1\}$) may capture whether he shows up at work, is normally available to his team, and reports to his superiors as instructed. On the other hand, soft performance ($x_i(a_i) \in [0,1]$) may indicate the extent to which in addition to performing these basic duties, the manager inspires and motivates his team, convincingly defends his projects at meetings, or shares helpful ideas. In other words, $x_i(a_i)$ measures the kind of performance that a motivated employee is happy to provide even in the absence of explicit incentives, but may stop providing if he feels unfairly or poorly treated by the organization (see the discussion of our modeling of soft performance in section 5.3).

Differences among the agents

If agent i does not work for the principal, he receives outside option $\underline{u}_i$ and the principal receives zero. Importantly, we assume that while the agents are equally productive within the organization (in the sense that their output functions are identical), they have different outside options. As discussed below, this implies that the principal would like to pay the agents

differently and therefore creates the potential for social comparisons among the agents.[6] Without loss of generality, we assume agent 1 has a better outside option than agent 2.

Assumption 1. $\underline{u}_1 > \underline{u}_2$.

For instance, agent 2 may face higher relocation costs because of his personal situation (married, with children, etc.), or may be less productive than agent 1 in alternative jobs (think of the different non-academic options of an applied microeconomics professor as opposed to a corporate finance professor).

3.2. Contracting Environment

Information

We assume that the assigned tasks and whether agents provide hard performance are publicly observed – that is, they are observed by the principal, both agent, and any parties outside the organization, including courts. In contrast, and consistent with Hart and Moore (2008), we assume that soft performance is imperfectly observed. Formally, we assume the following:

Assumption 2. *For any given* $i \in \{1,2\}$*, the task's difficulty,* a_i*, and the agent's hard performance,* $d_i(a_i)$*, are publicly observed; the agent's soft performance,* $x_i(a_i)$*, and the output,* f_i*, are publicly observed for* $x_i(a_i) < 1 - \overline{\sigma}$*, whereas they are privately observed by agent* i *for* $x_i(a_i) \geq 1 - \overline{\sigma}$.

Assumption 2 implies that the principal can induce agent i to provide hard performance on the assigned task a_i, and possibly some degree of soft performance, by offering a contract that obliges the principal to pay an appropriate compensation if the agent chooses $d_i(a_i) = 1$ and (any) $x_i(a_i) \geq 1 - \overline{\sigma}$. At the same time, assumption 2 implies that even in the presence of such a contract, the agent has discretion on how much soft performance to provide above the threshold $1 - \overline{\sigma}$. Accordingly, we interpret $\overline{\sigma} \in [0,1]$ as an inverse measure of the principal's

[6] The model could easily be extended to the case where the agents have different productivities within the organization. Notice, however, that the case we consider here, where differences among agents solely come from the outside options and hence cannot be rationalized as "fair", is the one where frustration from negative social comparisons is most likely to arise. See Breza et al. (2018) for evidence consistent with this point.

monitoring capability, which may depend on the more or less creative and complex nature of the agents' job (e.g., managers vs. clerks), on the firm's size or, more generally, on monitoring costs.[7]

Assumption 2 also implies that we can restrict attention to agreements where the principal demands, and the two agents exert, hard performance. To see why, recall that agent *i*'s output and his performance cost are both zero when either the agent does not exert hard performance, or when he does but the assigned task is $a_i = 0$. This observation, together with the assumption that the task's, a_i, and the agent's hard performance, $d_i(a_i)$, are both publicly observed and therefore contractible, imply that if there is an agreement in which $d_i(a_i) = 0$, then there exists another agreement with $a_i = 0$ and $d_i(0) = 1$ that is equivalent in terms of output, performance cost, and incentives. Accordingly, hereafter we assume that the two agents exert hard performance (that is, $d_i(a_i) = 1$ for $i \in \{1,2\}$) and refer to soft performance simply as performance.

Formal and informal contracts

We consider two types of contract—formal and informal—that the principal may offer to agent i in order to induce the desired performance. As standard in the literature, we call a contract formal (informal) if its existence can (cannot) be verified by third parties. We denote the compensation stipulated in a formal and informal contract between the principal and agent i as w_i^F and w_i^I, respectively. At the same time, as discussed in the introduction, we emphasize an unremarked feature of formally contracted pay – namely, that being more conspicuously recorded and verifiable, it is more likely to "leak" to third parties in the organization. To capture this insight in a stark fashion, we assume for most of the paper that a formal contract is publicly observed. We relax this assumption in Appendix II, and show that the model's key results are preserved so long as formal contracts are sufficiently likely to leak out.

Assumption 3. *Let w_i^F and w_i^I be the formally and informally agreed compensation of agent i, respectively. Then, the existence of w_i^F is publicly observed (i.e., it is observed by*

[7] The model's results would be unchanged if we allowed contractibility of performance to be task-dependent and take the general form $\overline{\sigma}(a_i)$.

the principal, both agents, and courts). In contrast, w_i^I is observed only by the principal and agent i.

A well understood advantage of formal contracting is that it enables third-party (including judicial) enforcement, and therefore, given assumption 2, commits the principal to paying the promised compensation conditional on the agent's provision of contractible performance. On the other hand, Assumption 3 implies that formal contracting produces public information on the existence of an agreement between the principal and a given agent. As we shall see in a moment, this feature of formal contracting may generate costly social comparisons between the focal agent and his peer, thereby creating a tradeoff between the two contractual forms.

3.3. Preferences and Social Comparisons

The model's central feature is that the agents dislike having a lower *official status* in the organization than their peers (e.g., Adams, 1963; McAdams, 1992, 1995). We model this idea by assuming that each agent suffers a disutility when his "publicly observed payoff" (which given our assumption 1 is the difference between the formally specified pay and the cost of performing the assigned task, $w_i^F - C(a_i)$) is lower than that of the other agent. Thus, we define an agent's disutility from social comparisons as:

$$\alpha^d A_i \equiv \alpha^d max\{0, w_j^F - C(a_j) - [w_i^F - C(a_i)]\}, \quad (1)$$

where $\alpha^d \in [0,1]$ is a scaling parameter that measures how much the agents care about their relative status (e.g., Fehr and Schmidt, 1999).

Extending Hart and Moore (2008), we assume that an agent who suffers from negative social comparisons retaliates against the organization—that is, withdraws performance (if possible) until his frustration is fully compensated.[8] Formally, the agent's *net* disutility from negative social comparisons is given by:

$$max\{0, \alpha^d A_i - (1 - x_i(a_i))y(a_i)\}, \quad (2)$$

[8] Arguably, the agent with lower status may also retaliate against the agent with higher status, causing a utility loss to the latter. We abstract from this additional social comparison cost, since adding it to the analysis would not change any of our main results.

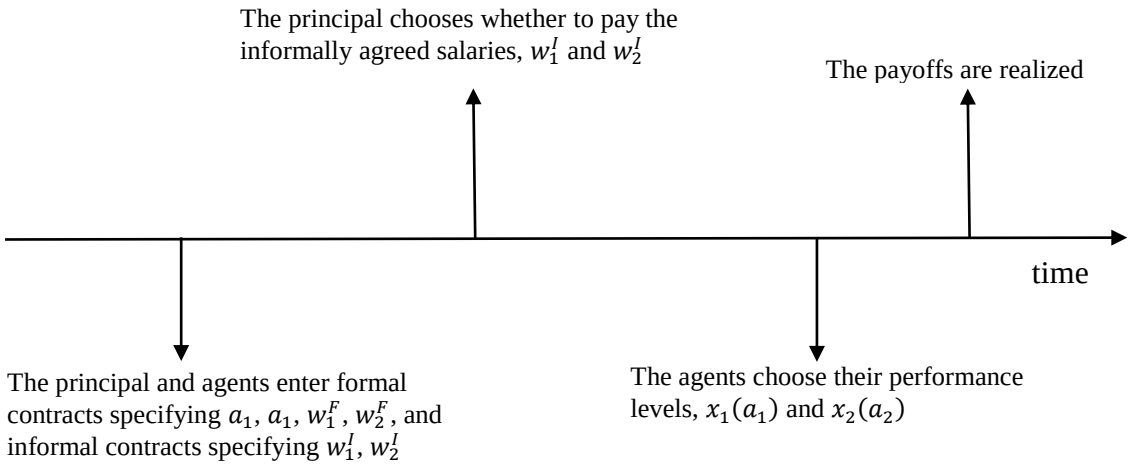
where the term $\big(1 - x_i(a_i)\big)y(a_i)$ captures how much of the agent's frustration is compensated by his retaliation against the organization.

Given that the agent's ability to retaliate is constrained by the principal's partial monitoring capability, $\overline{\sigma}$, the agent's optimal retaliation is given by $min\left\{\overline{\sigma}, \frac{\alpha^d A_i}{y(a_i)}\right\}$. Accordingly, the agent's level of performance equals:

$$x_i(a_i) = 1 - min\left\{\overline{\sigma}, \frac{\alpha^d A_i}{y(a_i)}\right\}. \qquad (3)$$

Given the above definitions and assumptions, we are now ready to study how the principal chooses governance (i.e., formal and informal contract terms) to manage social comparisons among the agents. Before we do so, it is useful to summarize the sequence of moves in the model through a timeline figure.

Figure 1. Timeline



4. **Managing Social Comparisons under Transparency**

We say that the organization adopts a "transparent" compensation policy if each agent's formal and informal pay levels coincide: $w_i^I = w_i^F = w_i$, for all i. Given the model's

definitions and assumptions, the agents' and principal's utilities under transparency are given, respectively, by:

$u_i \equiv w_i - C(a_i) - \left[\alpha^d A_i - min\left\{\overline{\sigma}, \frac{\alpha^d A_i}{y(a_i)}\right\} y(a_i)\right]$ for agent i, and

$u_P \equiv \left[1 - min\left\{\overline{\sigma}, \frac{\alpha^d A_1}{y(a_1)}\right\}\right] y(a_1) + \left[1 - min\left\{\overline{\sigma}, \frac{\alpha^d A_2}{y(a_2)}\right\}\right] y(a_2) - w_1 - w_2$ for the principal.

Then, the principal's optimal contract under transparency maximizes u_P, subject to the constraint that the agents be willing to work (participation constraints):

$max_{a_1, a_2, w_1, w_2}\{u_P\}$, subject to (4)

$u_i \geq \underline{u}_i$, for all i. (PC_i)

A useful, yet intuitive result is that it is optimal for the principal to make the two agents' participation constraints bind.

Lemma 1. *In the optimal contract under transparency, (PC_i) is binding for all i.*

Proof. Suppose to the contrary that there exists an optimal contract where (PC_i) holds with strict inequality for some agent $i \in \{1,2\}$. Then, decreasing w_i by a small $\varepsilon > 0$ would still satisfy (PC_i), would weakly relax (PC_j), $j \neq i$, and would increase the principal's utility by at least $\varepsilon(1 - \alpha^d)$, which is non-negative because $\alpha^d \in [0,1]$.[9] This, however, contradicts the optimality of the original agreement. ■

Lemma 1 implies that an agent's compensation is given by:

$$w_i = \underline{u}_i + C(a_i) + \left[\alpha^d A_i - min\left\{\overline{\sigma}, \frac{\alpha^d A_i}{y(a_i)}\right\} y(a_i)\right]. \quad (5)$$

The first two terms in (5) reflect the standard principle that an agent's compensation must cover his outside option, plus the cost of performance. The term in square brackets, however, is a specific feature of our model and measures the wage premium that the principal must pay in order to compensate the agent's frustration from negative social comparisons. The agent's

[9] To be precise, the proposed perturbation would increase the principal's utility by: (i) $\varepsilon(1 - \alpha^d)$ if $A_i \geq 0, A_j = 0$ and $\overline{\sigma} > \frac{\alpha^d A_i}{y(a_i)}$; (ii) $\varepsilon(1 + \alpha^d)$ if $A_j > 0$ and $\overline{\sigma} \geq \frac{\alpha^d A_j}{y(a_j)}$; and (iii) ε if $A_j > 0$ and $\overline{\sigma} < \frac{\alpha^d A_j}{y(a_j)}$, or if $A_i \geq 0, A_j = 0$ and $\overline{\sigma} \leq \frac{\alpha^d A_i}{y(a_i)}$.

premium increases in the degree of publicly observed disparity, A_i, and decreases in the agent's ability to retaliate against the principal, $\overline{\sigma}$.

4.1. Task Assignment, Performance and Compensation without Social Comparisons

As a benchmark, we begin by considering the case in which the agents do not engage in social comparisons: $\alpha^d = 0$. Define the first-best task, identical for both agents, as the one that maximizes joint surplus (that is, the difference between an agent's output and the cost of task completion, conditional on the agent providing soft performance):

$a^{FB} \equiv argmax_a\{y(a) - C(a)\}$.

The next proposition characterizes the principal's optimal contract in the absence of social comparisons.

Proposition 1. *Suppose that $\alpha^d = 0$. Then, under transparency, the principal's optimal contract has the following features:*

(i) *Each agent is assigned the first-best task and is paid his outside option plus the cost of performance: $a_i = a^{FB}$, and $w_i = \underline{u}_i + C(a^{FB})$ for $i = 1,2$.*

(ii) *There is no retaliation, and the principal's profit equals the first-best total surplus minus the agents' outside options: $u_P = 2[y(a^{FB}) - C(a^{FB})] - \underline{u}_1 - \underline{u}_2$.*

Proof. Lemma 1, together with the assumption that $\alpha^d = 0$, implies that there is no retaliation and that $w_i = \underline{u}_i + C(a_i)$. After substituting this into (4), the principal's problem simplifies to choosing a_1 and a_2 to maximize $u_P = y(a_1) + y(a_2) - \underline{u}_1 - C(a_i) - \underline{u}_2 + C(a_2)$, which is solved by setting $a_1 = a_2 = a^{FB}$. The principal's utility is therefore given by $u_P = 2[y(a^{FB}) - C(a^{FB})] - \underline{u}_1 - \underline{u}_2$. ■

The proposition shows that in the absence of social comparisons, and given that tasks and hard performance are observable and a satisfied agent is happy to provide soft performance, a simple work-for-wage contract between the principal and the two agents induces the first-best outcome. Because the agents have different outside options, they will receive different payments; but in the absence of social comparisons, these differences in formal contract terms

do not generate frustration and retaliation, and hence do not require that a wage premium be paid to the agents.

4.2. General Analysis

We now turn to the most interesting case of transparency and assume that the agents engage in social comparisons: $\alpha^d > 0$. Because we are interested in analyzing how social comparisons affect task assignment and compensation, in what follows we assume it is always profitable for the principal to employ the two agents.[10]

A first and useful result is that under transparency, the principal chooses task assignment and compensation such that only the agent with low outside option (agent 2) suffers from social comparisons.

Lemma 2. *Under transparency,* $A_2 > A_1 = 0$.

Proof. Since Lemma 1, together with Assumption 1, implies that $A_i > 0$ for some i, it suffices to prove that $A_1 = 0$. To this end, suppose to the contrary that $A_1 > 0$, which in turn implies that $A_2 = 0$. Note that $w_1 - C(a_1) \geq \underline{u}_1$ by (PC_1), and that $w_2 - C(a_2) = \underline{u}_2$ by eq. (5) in the text and the fact that $A_2 = 0$. Since $\underline{u}_1 > \underline{u}_2$ by assumption, it follows that $w_1 - C(a_1) > w_2 - C(a_2)$, or equivalently, that $A_1 = 0$, which is a contradiction. ■

This result is quite intuitive: since agent 1 has the higher outside option, the binding participation constraints require that he receive a higher payoff than agent 2. Importantly, the fact that agent 1 does not suffer from social comparisons implies that transparency does not distort his task, performance, or compensation, relative to the first-best.

Proposition 2. *Under transparency, the task, performance, and compensation of agent 1 are as in the first-best:* $a_1 = a^{FB}$, $x_1(a_1) = 1$, *and* $w_1 = \underline{u}_1 + C(a^{FB})$.

Proof. The result that $x_1(a_1) = 1$ follows directly from Lemma 2 and eq. (3) in the text. The result that agent 1 receives no wage premium (that is, $w_1 = C(a_1) + \underline{u}_1$) follows directly from Lemmas 1 and 2. To prove that $a_1 = a^{FB}$, suppose to the contrary that $a_1 \neq$

[10] Employing the two agents will be optimal if the first-best total surplus is "sufficiently" large.

a^{FB}. Consider now the following alternative contract for agent 1: $(\tilde{a}_1, \tilde{w}_1)$, where $\tilde{a}_1 = a^{FB}$ and $\tilde{w}_1 = C(a^{FB}) - C(a_1) + w_1$. It is straightforward to show that this alternative contract satisfies both participation constraints and increases the principal's payoff by $y(a^{FB}) - C(a^{FB}) - [y(a_1) - C(a_1)] > 0$, where the inequality follows from the assumption that $a_1 \neq a^{FB}$. But this contradicts the optimality of the original agreement, thereby proving our result. ■

We now turn to the agent with low outside option (agent 2). We show that because agent 2 suffers from negative social comparisons, his task assignment, compensation, and performance may be distorted relative to the first-best. Moreover, we show that these distortions depend on how much of his anger agent 2 can compensate by retaliating against the principal (formally, on whether $\overline{\sigma}y(a_2)$is greater or smaller than $\alpha^d A_i$). To build intuition on this last point, it is useful to separately analyze the two cases where the agent can (i.e., $\overline{\sigma}y(a_2) \geq \alpha^d A_i$) or cannot (i.e., $\overline{\sigma}y(a_2) < \alpha^d A_i$) fully retaliate. As a preliminary step, notice that given Lemma 1, the condition for full retaliation by agent 2, $\overline{\sigma}y(a_2) \geq \alpha^d A_i$, is satisfied if, and only if:

$$\overline{\sigma}y(a_2) \geq \alpha^d\left(\underline{u}_1 - \underline{u}_2\right). \tag{6}$$

Assume first that (6) holds. In that case, agent 2 fully compensates the anger from social comparisons and hence demands no wage premium. Lemma 1 and Proposition 2 then imply that with respect to this agent, the principal's problem, (4), simplifies to maximizing:

$$u_{P|\,\overline{\sigma}y(a_2)\geq\alpha^d(\underline{u}_1-\underline{u}_2)} \equiv y(a_2) - C(a_2) - \underline{u}_2 - \alpha^d\left(\underline{u}_1 - \underline{u}_2\right), \tag{7}$$

where the last term is the output reduction due to the agent retaliation. Clearly, the solution to this problem is for the principal to assign the first-best task, a^{FB}, and pay the first-best salary, $w_2 = \underline{u}_2 + C(a^{FB})$, to agent 2. At the same time, social comparisons and the ensuing retaliation imply that agent 2 performs sub-optimally: $x_2(a^{FB}) = 1 - \frac{\alpha^d(\underline{u}_1 - \underline{u}_2)}{y(a^{FB})} < 1$.

Assume now that (6) does not hold, so that full retaliation is not feasible. In this case, agent 2 demands a wage premium to compensate the frustration from social comparisons. Then, Lemma 1 and Proposition 2 imply that the agent's compensation is given by:

$$w_2 = \underline{u}_2 + C(a_2) + \frac{\alpha^d(\underline{u}_1 - \underline{u}_2) - \overline{\sigma}y(a_2)}{1+\alpha^d},$$

where the last term is the wage premium due to social comparisons. Notice that the wage premium is moderated, but not fully offset, by the retaliation term $\overline{\sigma}y(a_2)$. After substituting and rearranging, the principal's problem simplifies to maximizing:

$$u_{P|\,\overline{\sigma}y(a_2)<\alpha^d(\underline{u}_1-\underline{u}_2)} \equiv \left[1 - \frac{\alpha^d}{1+\alpha^d}\overline{\sigma}\right] y(a_2) - C(a_2) - \underline{u}_2 - \frac{\alpha^d(\underline{u}_1-\underline{u}_2)}{1+\alpha^d}, \qquad (8)$$

where $\frac{\alpha^d}{1+\alpha^d}\overline{\sigma}$ is the net output-reduction due to retaliation, while $\frac{\alpha^d(\underline{u}_1-\underline{u}_2)}{1+\alpha^d}$ is the remaining wage premium. Since retaliation now increases in output, the principal does not fully appropriate the marginal profit from the agent's performance, and therefore finds it optimal to distort the agent's task downwards, relative to the first-best. Specifically, the principal assigns to agent 2 the task $a_2 = a^*(\overline{\sigma})$, decreasing in $\overline{\sigma}$, which solves:

$$\left[1 - \frac{\alpha^d}{1+\alpha^d}\overline{\sigma}\right] y'\big(a^*(\overline{\sigma})\big) - C'\big(a^*(\overline{\sigma})\big) = 0. \qquad (9)$$

At the same time, the principal compresses the compensation of agent 2 upwards by paying a wage premium, $\frac{\alpha^d(\underline{u}_1-\underline{u}_2)-\overline{\sigma}y(a^*(\overline{\sigma}))}{1+\alpha^d} > 0$, on top of the agent's outside option and performance cost. Finally, retaliation implies that agent 2 underperforms relative to the first-best, albeit to a lesser degree than in the full retaliation case: $x_2(\mathrm{a}_2) = 1 - \frac{\alpha^d(\underline{u}_1-\underline{u}_2)}{y(a^*(\overline{\sigma}))} \in \left(1 - \frac{\alpha^d(\underline{u}_1-\underline{u}_2)}{y(a^{FB})}, 1\right)$.

Whether the agent's retaliation capacity constraint binds (condition (6)), and hence whether the principal is in the task distortion region, depends on the constraint's tightness, $\overline{\sigma}$, and on the principal's task choice, a_2. The lower a_2 and $\overline{\sigma}$, the more likely that (6) does not hold and hence that the principal is in the "distortion" region. This observation provides the intuition for our next result.

Proposition 3. *There exists $\sigma^* \in [0, \infty]$ such that under transparency, the principal's optimal policy towards the agent with low outside option (agent 2) has the following features:*

(i) If $\overline{\sigma} < \sigma^*$, the task and performance of agent 2 are distorted downwards, and his compensation is compressed upwards: $a_2 = a^*(\overline{\sigma})$, $x_2(a_2) = 1 - \overline{\sigma}$, and $w_2 = \underline{u}_2 + C\big(a^*(\overline{\sigma})\big) + \frac{\alpha^d(\underline{u}_1 - \underline{u}_2) - \overline{\sigma}y(a^*(\overline{\sigma}))}{1+\alpha^d} > \underline{u}_2 + C\big(a^*(\overline{\sigma})\big)$.

(ii) If $\overline{\sigma} > \sigma^*$, there are no task distortion or pay compression, but the performance of agent 2 is more severely distorted downwards: $a_2 = a^{FB}$, $x_2(a_2) = 1 - \frac{\alpha^d(\underline{u}_1 - \underline{u}_2)}{y(a^{FB})}$, and $w_2 = \underline{u}_2 + C(a^{FB})$.

(iii) If $\overline{\sigma} = \sigma^*$, then the principal is indifferent between the policies described in (i) and (ii) above.

Proof. See Appendix I.

A straightforward corollary of Proposition 3 is that the principal's profit is (weakly) decreasing in the agent's retaliation power, $\overline{\sigma}$, whereas task distortion in the organization, measured by the difference between the difficulty of the two agents' tasks, $a^{FB} - a^*(\overline{\sigma})$, is non-monotonic in $\overline{\sigma}$.

Corollary 1. *Under transparency, the principal's profit, u_P, and the task distortion of the agent with low outside option (agent 2), $a^{FB} - a^*(\overline{\sigma})$, exhibit the following features:*

(i) u_P decreases in $\overline{\sigma}$ for $\overline{\sigma} \in [0, \sigma^*]$, and stays constant for $\overline{\sigma} > \sigma^*$.

(ii) $a^{FB} - a^*(\overline{\sigma})$ increases in $\overline{\sigma}$ for $\overline{\sigma} \in [0, \sigma^*]$, and drops to zero for $\overline{\sigma} > \sigma^*$.

Proof. See Appendix I.

Proposition 3 and its corollary have interesting testable implications. First, if a firm makes its employees' compensation public, social comparisons and the ensuing employee frustration may prompt it to compress pay upwards—that is, to overpay certain employees relative to the compensation level they could command in the absence of social comparisons.

Second, and perhaps more surprisingly, social comparisons may force the firm to use its human capital inefficiently. In particular, the firm may assign the employees who suffer from social comparisons to simple/light tasks in order to constrain their retaliation capacity. This

result is consistent with Nickerson and Zenger's observation that firms respond to social comparison costs by "shaping a firm's production technology, broadly defined" (Nickerson and Zenger, 2008, p. 1437). However, while Nickerson and Zenger focus on distortions in the production technology that aim at reducing comparisons among employees (e.g., avoiding team production), the distortion emerging from our model aims at reducing retaliation holding the extent of social comparisons constant.

Finally, our analysis indicates that while increases in the employees' retaliation capability unambiguously reduce firm profits, they have a non-monotonic effect on task distortion. These comparative static predictions could be tested by relating firm-level or unit-level differentials in the task assignment of employees with similar skills ($a^{FB} - a^*(\overline{\sigma})$) to measures of the employees' discretion ($\overline{\sigma}$), such as effectiveness of the firm's monitoring technology (e.g., employees' distance from headquarters, as in Kosova and Sertsios, 2018, and the franchising literature reviewed by Lafontaine and Slade, 2013), or of the local contract enforcement institutions (see Johnson et al., 2002, and Antras and Foley, 2015, for empirical studies exploiting variations in the quality of contract enforcement institutions).

5. Managing Social Comparisons under Privacy

5.1. Enriching the Model: Enforcement of Informal Compensation

In this section we analyze the possibility that the principal may minimize differences in the two agents' formal compensation levels, so as to avoid social comparisons, while relying on private informal agreements to optimally differentiate pay. We call this type of compensation policy "privacy". Therefore, the distinctive feature of privacy, compared to transparency, is that informal compensation is used and may differ from formal compensation: $w_i^I \neq w_i^F$, for some i.

The key challenge of privacy arises from the fact that informal agreement on compensation cannot be verified by courts, and therefore from the requirement that the principal's promise to pay w_i^I be self-enforcing—that is, it must be backed by a credible threat of punishment by the

agent in case of non-payment. As discussed in the introduction, when this credibility problem is binding, it constrains the firm's ability to use informal compensation optimally.

Our modeling of self-enforcement and credibility in informal contracts differs from the standard relational contracting literature, and it is closer in spirit to recent models of relational contracting with reciprocity (Fahn et al., 2017). In most of the literature, informal employment agreements are modeled as equilibria of an infinitely repeated game sustained by the threat of termination (e.g., MacLeod and Malcomson, 1989; Baker et al., 1994; Levin, 2003). The reason for this approach is that in these models, employees would never perform in the last period of a finite horizon game, regardless of whether the employer has performed as promised. Infinite repetition is then needed to create a credible punishment threat for an employer who fails to pay the promised compensation. In our model, however, the agents are happy to provide soft performance when "well treated" by the principal, implying that they can credibly threaten to withdraw such performance if the principal reneges on the informal agreement, even in the one-shot game. In a companion working paper (Contreras and Zanarone, 2017), we analyze a repeated game version of the model, obtaining qualitatively similar results.

Formally, we model self-enforcement by slightly extending the retaliation mechanism introduced in section 3. We assume that if the principal promises informal compensation $w_i^I > w_i^F$ to agent i, and then deviates by paying only the formally agreed part, w_i^F, the agent feels "cheated" and suffers disutility $\alpha^l(w_i^I - w_i^F)$, where $\alpha^l \geq 0$ measures the extent to which agents are outraged by cheating. As before, the agent compensates his disutility (which may now come from both social comparisons *and* cheating) by withdrawing soft performance until either his frustration is fully offset or the retaliation capacity constraint is binding. Therefore, given any task a_i, any formal compensation w_i^F, and any informal compensation w_i^I, the agent's performance following a deviation by the principal is given by:

$$x_i(a_i) = 1 - min\left\{\overline{\sigma}, \frac{\alpha^d A_i + \alpha^l(w_i^I - w_i^F)}{y(a_i)}\right\}. \qquad (10)$$

Accordingly, we define privacy as a subgame perfect equilibrium where for any given agent i: (a) the principal assigns task a_i and offers formal compensation w_i^F and informal compensation w_i^I, (b) the agent accepts, (c) the principal pays w_i^I as promised, and (d) the agent

provides performance $x_i(a_i) = 1 - min\left\{\overline{\sigma}, \frac{\alpha^d A_i}{y(a_i)}\right\}$. In the event of a deviation by the principal (i.e., a failure to pay w_i^I as promised), the agent retaliates by lowering performance to the level described by (10).

As discussed above, parameter α^l measures the extent to which the principal's cheating frustrates the agent, and it may in principle differ from α^d, the agent's sensitivity to social comparisons. This parameter may be interpreted as a psychological feature of the two agents. Alternatively, and perhaps more interestingly, α^l may be interpreted as a characteristic of the organization that affects the agents' sensitivity to cheating. For instance, it may be argued that an agent who trusts the principal (perhaps due to the principal's organizational culture and reputation) feels more personally outraged by an outright violation of the informal agreement than an agent who has limited trust in the principal. Therefore, one may interpret α^l as the organization's perceived "integrity" (Guiso et al., 2015), or as the society's perceived level of trust (e.g., Guiso et al., 2008).

5.2. Task Assignment, Performance and Compensation under Privacy

Since the principal pays w_i^I in equilibrium, the agent's and principal's utilities under privacy are given by:

$u_i \equiv w_i^I - C(a_i) - \left[\alpha^d A_i - min\left\{\overline{\sigma}, \frac{\alpha^d A_i}{y(a_i)}\right\} y(a_i)\right]$ for agent i, and

$u_P \equiv \left[1 - min\left\{\overline{\sigma}, \frac{\alpha^d A_1}{y(a_1)}\right\}\right] y(a_1) + \left[1 - min\left\{\overline{\sigma}, \frac{\alpha^d A_2}{y(a_2)}\right\}\right] y(a_2) - w_1^I - w_2^I$ for the principal.

Then, the principal's optimal contract under privacy maximizes u_P, subject to the constraints that the agents be willing to work (participation constraints), that they be willing to take the informal compensation instead of demanding the formally contracted compensation (agents' incentive constraints), and that the principal be willing to pay the informally agreed compensation instead of the formally contracted one (principal's incentive constraint):

$max_{a_1, a_2, w_1^F, w_2^F, w_1^I, w_2^I}\{u_P\}$, subject to (11)

$u_i \geq \underline{u}_i$, for all i. (PC_i)

$w_i^I \geq w_i^F$, for all i. (IC_i)

$$w_i^I - w_i^F - \left[min\left\{\overline{\sigma}, \frac{\alpha^d A_i + \alpha^l (w_i^I - w_i^F)}{y(a_i)}\right\} - min\left\{\overline{\sigma}, \frac{\alpha^d A_i}{y(a_i)}\right\}\right] y(a_i) \leq 0, \text{ for all } i. \quad (IC_{Pi})$$

Condition (IC_i) has the intuitive implication that for the agent to accept the informally agreed compensation, and renounce on enforcing the formally contracted one, formal pay must be a baseline that the informal pay complements. In turn, condition (IC_{Pi}) implies that for the informally agreed compensation to be self-enforcing, the punishment the agent can inflict if the principal cheats (the term in square brackets) must be strong enough. Since the punishment increases in α^l, a first important implication of our model is that privacy cannot be sustained when α^l is small – that is, when the agents are relatively insensitive to, and hence have limited willingness to punish the principal's cheating.

Proposition 4. *Suppose the agents are not too sensitive to the principal's cheating ($\alpha^l < 1$). Then, privacy cannot be sustained in equilibrium, and therefore transparency is optimal.*

Proof. See Appendix I.

5.2.1. Privacy when the agents have high retaliation capability

Our next result shows that in contrast, when the agents are sensitive to cheating ($\alpha^l \geq 1$), a first-best privacy policy may be sustained, such that social comparisons between the two agents are completely eliminated by equalizing formal, publicly known compensation levels, while informal agreements ensure that actual pay levels are optimally differentiated. To rule out implausible multiple equilibria, we assume (realistically) that every time the principal drafts a formal contract term, he incurs a small but positive cost, such that if possible, the principal prefers the two formal payments to be equal: $w_1^F = w_2^F$.

Define $\sigma^{**} \equiv \frac{u_1 - u_2}{y(a^{FB})}$. We have the following result.

Proposition 5. *Suppose the agents are sensitive to cheating by the principal ($\alpha^l \geq 1$), and have high retaliation capability ($\overline{\sigma} \geq \sigma^{**}$). Then, under privacy, the principal's optimal contract has the following features:*

(i) *Each agent is assigned the first-best task and receives an informal pay equal to his outside option plus the cost of performance:* $a_i = a^{FB}$ *, and* $w_i^I = \underline{u}_i + C(a^{FB})$ $for\ i = 1,2.$

(ii) *The two agents receive the same formal pay, which equals the informal pay of the agent with low outside option (agent 2):* $w_1^F = w_2^F = w_2^I$.

(iii) *There is no retaliation:* $x_i(a^{FB}) = 1$, *for* $i = 1,2.$

(iv) *The principal's profit equals the first-best total surplus, minus the agents' outside options:* $u_P = 2[y(a^{FB}) - C(a^{FB})] - \underline{u}_1 - \underline{u}_2$.

Proof. See Appendix I.

Proposition 5 highlights the model's key insight: a privacy policy that includes uniform task assignments and formal pay to minimize social comparisons, while relying on informal pay to differentiate among employees, requires that the agents be willing to ($\alpha^l \geq 1$) and capable of ($\overline{\sigma} \geq \sigma^{**}$) severely punishing the principal for deviating on the informally agreed compensation.

5.2.2. Privacy when the agents have limited retaliation capability

We conclude our analysis by characterizing the optimal privacy policy when the agents are willing to punish the principal's deviations ($\alpha^l \geq 1$) but retaliation is limited by the principal's ability to monitor the agents' soft performance ($\overline{\sigma} < \sigma^{**}$). A key insight is that when agent retaliation is limited, a privacy policy may be feasible but cannot achieve the first best—that is, both optimal task assignment *and* pay differentiation.

To build intuition on the shape of an optimal privacy policy under limited retaliation capacity, we state three useful results. We begin by showing that under privacy, like under transparency, it is optimal for the principal to make the two agents' participation constraints bind.

Lemma 3. *Under the optimal privacy policy,* (PC_i) *is binding for all* i.

Proof. Suppose to the contrary that (PC_i) holds with strict inequality for some $i \in \{1,2\}$. Imagine first that $w_i^I > w_i^F$, and consider a perturbation whereby w_i^I is decreased by a small $\varepsilon > 0$. It is easy to check that after this perturbation the participation and incentive constraints are satisfied and the principal's utility increases by ε, contradicting the optimality of the original agreement. Imagine next that $w_i^I = w_i^F$, and consider a perturbation whereby both w_i^I and w_i^F are decreased by a small $\varepsilon > 0$. It is easy to check that after this perturbation the participation and incentive constraints are satisfied and the principal's utility increases by at least $\varepsilon(1 - \alpha^d) > 0$, where the last inequality holds since $\alpha^d < 1$ by assumption. This, however, contradicts the optimality of the original agreement. ■

Next, we show that under privacy, like under transparency, the agent with high outside option (agent 1) does not suffer from social comparisons.

Lemma 4. *Under privacy, agent 1 does not suffer from social comparisons:* $A_2 \geq A_1 = 0$.

Proof. Suppose to the contrary that $w_2^F - C(a_2) > w_1^F - C(a_1)$, so that $A_1 > A_2 = 0$. We first show that $w_1^I > w_1^F$. To this end, notice that $w_1^I - C(a_1) \geq \underline{u}_1$ by condition (PC_1), that $\underline{u}_1 > \underline{u}_2$ by assumption, and that $\underline{u}_2 = w_2^I - C(a_2)$ by Lemma 3 and the assumption that $A_2 = 0$. Consequently, it follows that $w_1^I - C(a_1) > w_2^I - C(a_2)$. Moreover, notice that $w_2^I \geq w_2^F$ by eq. (IC_2). Using the two previous observations, together with the original assumption that $w_2^F - C(a_2) > w_1^F - C(a_1)$, it follows that $w_1^I > w_1^F$, as desired. Consider now increasing w_1^F by an arbitrarily small $\varepsilon > 0$. Using the previous result that $w_1^I > w_1^F$ and the assumption that $w_2^F - C(a_2) > w_1^F - C(a_1)$, it is easy to verify that the perturbation induces a new equilibrium in which either the level of retaliation by agent 1 decreases or agent 1's participation constraint becomes slack. Since in the former case the principal's utility increases and in the second case we can use the same perturbations as in the proof of Lemma 3 to construct a new self-enforcing contract in which the principal is strictly better off, this contradicts the optimality of the original contract. ■

Finally, in the next two lemmas we show that due to limited retaliation ($\overline{\sigma} < \sigma^{**}$), the principal cannot set the agents' formal pay levels too low relative to the informal payments, or

else he would be tempted to stick to the formal contracts and renege on the informal ones, thereby undermining the privacy policy.

Lemma 5. *The formal compensation of agent 2 equals his informal one:* $w_2^I - w_2^F = 0$.

Proof. See Appendix I.

Lemma 6. *Incentive constraint* (IC_{P1}) *is binding and takes the form* $w_1^I - w_1^F = \overline{\sigma} y(a_1)$.

Proof. See Appendix I.

We are now ready to characterize the optimal privacy policy under limited retaliation. To this purpose, recall that the task $a^*(\overline{\sigma})$ is implicitly given by

$$\left[1 - \frac{\alpha^d}{1+\alpha^d}\overline{\sigma}\right] y'\left(a^*(\overline{\sigma})\right) - C'\left(a^*(\overline{\sigma})\right) = 0.$$

In addition, define the following critical agent tasks:

Task $a^{**}(\overline{\sigma}) > a^{FB}$ such that $y'\left(a^{**}(\overline{\sigma})\right)\left[1 + \frac{\alpha^d}{1+\alpha^d}\overline{\sigma}\right] - C'\left(a^{**}(\overline{\sigma})\right) = 0.$ (12)

Task $a^{***}(\overline{\sigma}) > a^{FB}$ such that $y'\left(a^{***}(\overline{\sigma})\right)[1 + \alpha^d\overline{\sigma}] - C'\left(a^{***}(\overline{\sigma})\right) = 0.$ (13)

Task $a^{****}(\overline{\sigma}) > a^{FB}$ such that $\overline{\sigma} y\left(a^{****}(\overline{\sigma})\right) = \underline{u}_1 - \underline{u}_2.$ (14)

Then, we can use Lemmas 3 through 6 to prove the following result.

Proposition 6. *Suppose the agents are sensitive to cheating by the principal* ($\alpha^l \geq 1$) *and have limited retaliation capability* ($\overline{\sigma} < \sigma^{**}$). *Then, there are* $\overline{\sigma}_1$ *and* $\overline{\sigma}_2$, *with* $0 < \overline{\sigma}_1 \leq \overline{\sigma}_2 < \sigma^{**}$, *such that the optimal privacy policy has the following features:*

(i) *Low retaliation capacity:* $\overline{\sigma} \in [0, \overline{\sigma}_1]$

- *(i) The informal pay of agent 1 is not distorted:* $w_1^I = \underline{u}_1 + C(a_1)$.
- *(ii) The informal pay of agent 2 is equal to the formal pay and is distorted:* $w_2^I = w_2^F = \underline{u}_2 + C(a_2) + \frac{\alpha^d[\underline{u}_1 - \underline{u}_2 - \overline{\sigma} y(a_1)] - \overline{\sigma} y(a_2)}{1+\alpha^d}$.

(iii)The formal pay of agent 1 differs from that of agent 2: $w_1^F = \underline{u}_1 + C(a_1) - \overline{\sigma}y(a_1)$.

(iv) The task of agent 1 is distorted upwards: $a_1 = a^{**}(\overline{\sigma}) > a^{FB}$.

(v) The task of agent 2 is distorted downwards: $a_2 = a^{*}(\overline{\sigma}) < a^{FB}$.

(ii) <u>*Medium-low retaliation capacity*</u>*:* $\overline{\sigma} \in [\overline{\sigma}_1, \overline{\sigma}_2]$

(i) The informal pay of agent 1 is not distorted: $w_1^I = \underline{u}_1 + C(a_1)$.

(ii) The informal pay of agent 2 is equal to the formal pay and is not distorted: $w_2^I = w_2^F = \underline{u}_2 + C(a_2)$.

(iii)The formal pay of agent 1 differs from that of agent 2: $w_1^F = \underline{u}_1 + C(a_1) - \overline{\sigma}y(a_1)$.

(iv) The task of agent 1 is distorted upwards: $a_1 = a^{***}(\overline{\sigma}) > a^{FB}$.

(v) The task of agent 2 is not distorted: $a_2 = a^{FB}$.

(iii)<u>*Medium-high retaliation capacity*</u>*:* $\overline{\sigma} \in [\overline{\sigma}_2, \sigma^{**})$

(i) The informal pay of agent 1 is not distorted: $w_1^I = \underline{u}_1 + C(a_1)$.

(ii) The informal pay of agent 2 is equal to the formal pay and is not distorted: $w_2^I = w_2^F = \underline{u}_2 + C(a_2)$.

(iii)The formal pay of agent 1 differs from that of agent 2: $w_1^F = \underline{u}_2 + C(a_1)$.

(iv) The task of agent 1 is distorted upwards: $a_1 = a^{****}(\overline{\sigma}) > a^{FB}$.

(v) The task of agent 2 is not distorted: $a_2 = a^{FB}$.

Proof. See Appendix I.

As in our former analysis of transparency and first-best privacy, since agent 1 does not suffer from social comparisons (Lemma 3), his compensation is not distorted—that is, he receives his outside option plus the cost of performance. Moreover, Lemma 6 implies that the

principal's ability to commit to a high gap between agent 1's informal and formal pay, thereby reducing the social comparisons suffered by agent 2, is constrained by the agent's ability to retaliate in case of cheating, $\overline{\sigma}y(a_1)$. This has two important consequences for the optimal privacy policy. First, agent 2 may suffer from some degree of social comparison in equilibrium and consequently, the principal may distort his task downwards to reduce output and retaliation, as in Proposition 3. Second, the principal may distort the task of agent 1 upwards, relative to the first-best, in order to increase potential output and through that channel, the agent's retaliation capacity, $\overline{\sigma}y(a_1)$ and his own commitment power.

A joint implication of Propositions 3, 5 and 6 is that under transparency the agents' retaliation power, $\overline{\sigma}$, increases social comparison costs, and hence is a net liability for the principal, whereas under privacy the threat of retaliation may increase the principal's commitment and hence reduce social comparison costs. Moreover, and similarly to transparency, Proposition 6 implies that task distortion, measured by the difference between the difficulty of the tasks of agent 1 and agent 2, is non-monotonic in the agents' retaliation power, $\overline{\sigma}$. Our next result makes these observations more precise.

Corollary 2. *Under privacy, the principal's profit, u_P, and the level of task distortion in the organization, $a_1 - a_2$, exhibit the following features:*

(i) u_P *increases in* $\overline{\sigma}$ *for* $\overline{\sigma} \in [0, \sigma^{**}]$ *and stays constant at the first-best level for* $\overline{\sigma} \geq \sigma^{**}$.

(ii) $a_1 - a_2$ *increases in* $\overline{\sigma}$ *for* $\overline{\sigma} \in [0, \overline{\sigma}_2]$, *decreases in* $\overline{\sigma}$ *for* $\overline{\sigma} \in [\overline{\sigma}_2, \sigma^{**}]$ *and stays constant at zero for* $\overline{\sigma} \geq \sigma^{**}$.

Proof. See Appendix I.

As in the case of transparency, under (imperfect) privacy social comparisons may force the firm to use its human capital inefficiently. Relative to transparency, however, task distortion under privacy is two-sided. On one hand, as before, the firm may assign the employees who suffer from social comparisons to simple/light tasks in order to constrain their retaliation capacity. On the other hand, the firm may assign employees who do not suffer from social comparisons to excessively difficult/burdensome tasks in order to increase potential output and hence expose itself to stronger employee retaliation if the informal contract is violated. In turn,

this stronger retaliation threat enhances the firm's power to commit itself to informal compensation agreements, and hence its ability to sustain privacy.

5.3. Discussion of the Assumptions

Since the key ingredients of this paper—social comparisons and retaliation—are not standard features of organizational design models, it is useful to discuss our assumptions, and how they may limit the scope of our model, in some more detail. First, we have assumed that the agents do not feel satisfaction (or to the contrary, compassion) when their status is *above* that of their peers. We believe the model's key results would extend to the case where the agents care about favorable social comparisons, although including such possibility may deliver additional interesting results. For instance, if an agent enjoys having higher status, he may "reciprocate" the principal by spontaneously performing a more demanding task than agreed. It is worth noticing however, that the more recent empirical literature discussed in the introduction supports our assumption that an organization's employees only care about negative social comparisons (Card et al., 2012; Cohn et al., 2014; Ockenfels et al., 2015; Breza et al., 2018). Moreover, while there seems to be a consensus in the literature on the fact that negative social comparisons generate frustration, it is less clear whether positive social comparisons should generate satisfaction and perhaps positive reciprocity towards the organization, or further frustration due to compassion towards the employees penalized by such comparisons (e.g., Ashraf, 2018). Thus, our assumption of negative social comparisons appears to be a natural starting point.

A second set of assumptions in our model that deserves closer scrutiny pertains to retaliation. We have closely followed the seminal model by Hart and Moore (2008) in assuming that soft performance (as opposed to hard performance) is costless and slightly pleasurable, such that the agents are happy to provide it if they feel "well treated" by the organization, but will stop providing if they feel mistreated. While assuming small pleasure and zero cost allows us to simplify the model's notation, it should be clear that our results would continue to hold in a model where both the agent's cost of and his intrinsic pleasure from providing soft performance are positive, so long as the cost is small enough relative to the pleasure. The idea that individuals are to some extent happy to perform if they feel well treated by the organization

they work for is well established in the behavioral economics literature on reciprocity (Akerlof, 1982; Fehr et al., 1993, 1998) and has received empirical and experimental support (see the review in Camerer and Weber, 2013).

At the same time, it should be noted that our model extends the notions of frustration and retaliation in Hart and Moore (2008). While they assume frustration and retaliation arise when an agent does not obtain what he expects from bargaining with the principal, we assume they arise when an agent: (i) feels that the principal treats him worse than another agent and/or (ii) feels that the principal fails to hold on to his promises. Extension (ii) seems natural because it pertains to the bilateral relationship between the principal and an agent. Extension (i) is less obvious and requires an implicit assumption that the agent blames the principal for observed status differences. At the same time, the empirical literature discussed in the introduction (e.g., Cohn et al., 2014; Breza et al., 2018), stressing how employees reduce productivity in response to observed pay inequality, encourages us to make this assumption.

Regarding the technology, our model assumes that the agents' contributions to firm output are separable. This formulation naturally captures a wide range of settings (for instance: teaching and research in a school, university or department; medical practice in a hospital; management of a division or project team in a firm; many textile manufacturing processes) where employees of similar rank conduct their work in relative isolation, and yet they socialize and therefore engage in social comparisons. At the same time, our model does not directly apply to team production settings, such as assembly lines, emergency and surgery rooms, and the like. As discussed in the Conclusion, extending our model to such settings may be interesting for at least two reasons. First, team production may increase socialization opportunities and therefore social comparisons and status concerns. Second, complementarities in production may result in parallel complementarities in retaliation. At the same time, we believe that the key qualitative insights from our model (namely, the presence of task and pay distortion under transparency, and the feasibility of privacy under high retaliation and commitment power) would continue to hold in a team production setting.

Another important feature of our model is that when computing their official status and level of frustration, the agents "believe what they see". In particular, we have defined privacy

as an equilibrium where social comparisons are driven by differences in formal pay while actual pay is contracted informally, implicitly assuming that social comparisons do not depend on the agents' conjectures about equilibrium payments. As discussed in the introduction, the psychology literature provides a micro-foundation for this assumption through the notion of "self-serving belief". According to such notion, individuals engage in beliefs that make them better off so long as there is evidence that they can use as an anchor to construct and justify such beliefs to themselves (e.g., Kunda, 1990; Dahl and Ransom, 1999; Haisley and Weber, 2010). Consistent with this notion, if in our model the agent with low outside option observes a lower public pay differential than the one that should arise in the optimal equilibrium, he will use such evidence as an anchor to convince himself that his status is equal to that of the other agent, thereby avoiding frustration. Investigating this application of the notion of self-serving beliefs experimentally will be an exciting topic for future research.

A last point pertains to our definition of informal contracts as agreements whose existence is *only* known by the parties. This definition rules out the possibility that not only third parties outside the organization, such as courts, but also third parties inside the organization—namely, agent j—may be aware of the informal contract between the principal and agent i. This modeling approach captures a range of important employment settings. In particular, the model naturally applies to firms and institutional environment where secrecy social norms (Edwards, 2005), and the fear of deteriorating personal relationships with colleagues, limit informational leakage on compensation via bragging and gossiping by the better paid employees. Nevertheless, there may also be firms where the existence of tacit and informal compensation agreements is more likely to "leak" out accidentally, as a byproduct of gossip and communication between employees, or intentionally, via bragging. In such settings, our baseline model does not directly apply. To address this potential limitation of our model's generality, in Appendix II we relax our definition of informal contracts to allow for the possibility that the two agents may observe each other's informally agreed compensation with positive probability. We show that an optimal privacy policy is still available when leakage is possible, and that it has similar features to those in the baseline model provided that the probability of leakage is not too large.

6. Testable Implications for Organization Design

In this section, we discuss testable predictions of our model on how compensation policy, within-firm task assignment, and organizational boundaries and architecture vary with our two key parameters—namely, the agents' willingness to retaliate if the principal cheats on compensation (α^l), and their ability to retaliate against the principal in case of cheating or negative social comparisons ($\overline{\sigma}$). A key insight from our analysis is that pay compression and distortions in firm boundaries, internal organization, and task assignments are primarily driven by the possibility to sustain privacy via self-enforcing informal agreements.

6.1. Pay Compression

We define formal pay compression as the pay difference between the two agents in the absence of social comparisons minus the formally agreed pay difference:

$$\Delta^F \equiv \left[\underline{u}_1 - \underline{u}_2 + C(a_1) - C(a_2)\right] - (w_1^F - w_2^F).$$

Likewise, we define informal pay compression as the pay difference in the absence of social comparisons minus the informally agreed pay difference:

$$\Delta^I \equiv \left[\underline{u}_1 - \underline{u}_2 + C(a_1) - C(a_2)\right] - (w_1^I - w_2^I).$$

A testable implication of our model is that privacy exhibits higher formal pay compression, but lower informal compression, than transparency. Intuitively, this occurs because under privacy, the principal can decrease social comparisons by reducing the formal pay gap between agents, while increasing the informal pay gap towards the optimal level. Given proposition 4, this implies, in turn, that as the agents' willingness to punish the principal for cheating on informal compensation, α^l, increases, formal pay compression also increases, while informal pay compression decreases.

Organizational Implication 1. *In an organization where the agents are highly sensitive to cheating ($\alpha^l \geq 1$) and hence privacy is potentially feasible, formal pay compression, Δ^F, is higher, and informal pay compression, Δ^I, is lower than in an organization where the agents are relatively insensitive to cheating ($\alpha^l < 1$) and hence transparency must be used.*

Proof. In Appendix I.

As discussed before, if α^l increases in organizations with higher perceived integrity, or in social environments with higher perceived trust, one can use the measurement techniques developed by Guiso et al. (2008) for trust, and by Guiso et al. (2015) for organizational integrity, to test Organizational Implication 1. Alternatively, one could compare pay compression levels in organizations where transparency is exogenously given (for instance, due to legal constraints or social norms; see Gillan et al., 2009, and Ockenfels et al., 2015) to the levels in organizations where privacy is allowed and used.

6.2. Firm Boundaries and Organizational Architecture

Although social comparisons may occur both between and within organizations, the relative strength of such comparisons is likely to vary with organizational boundaries and internal architecture. For instance, Nickerson and Zenger (2008) argue that while employees within a firm compare each other's status, they are less likely to engage in such comparisons with employees outside the firm, due to the lack of physical proximity, common identity, and social interaction. A similar, albeit perhaps smoother reduction in social comparisons may arise when employees belong to different departments or divisions within the same organization.

To incorporate the choice of firm boundaries and organizational architecture into our model in the simplest possible way, let $U_P^I(\alpha^l, \overline{\sigma})$ be the principal's utility under integration, as defined by the analysis in former sections, and let $U_P^O - k$ be the principal's utility under outsourcing, where k is a random variable with cdf $F(k)$ that captures the costs of outsourcing.[11] Note that U_P^O does not depend on $(\alpha^l, \overline{\sigma})$ because of the absence of social comparisons, and hence of informal contracting, under outsourcing.[12]

[11] The literature has pointed out multiple reasons for why the reduced form loss k may arise. For instance, outsourcing may bias the partner towards profit maximization at the expense of unobservable activities that are valuable to the principal (Holmstrom and Milgrom, 1991), or it may reduce the principal's incentive to undertake specific investments (Williamson, 1975; Grossman and Hart, 1986). Alternatively, both outsourcing and the complex organizational architecture may reduce the two agents' ability to develop communication routines that help them coordinate production (Cremer et al., 2007).

[12] The absence of informal contracting under outsourcing follows from our assumption that the agents are willing to provide soft performance without incentives if they are not aggrieved by social comparisons or cheating by the principal. Clearly, informal contracting would become relevant in a richer model of outsourcing with moral hazard, as shown by Baker et al. (2002). Enriching the model in that direction would not alter the main point we make here.

The above analysis implies that the probability that the principal prefers integration over outsourcing as a governance mode is given by:

$$Pr(I) \equiv Pr\Big(k > U_P^O - U_P^I(\alpha^l, \overline{\sigma})\Big) = 1 - F\Big(U_P^O - U_P^I(\alpha^l, \overline{\sigma})\Big).$$

It immediately follows from Proposition 4 that $Pr(I)$ is non-increasing in α^l, the agents' sensitivity to cheating. Moreover, it follows from Corollary 1 and Proposition 4 that conditional on privacy not being feasible ($\alpha^l < 1$), $Pr(I)$ is decreasing in the agent's retaliation power, $\overline{\sigma}$, for $\overline{\sigma} \in [0, \sigma^*]$, and stays constant for $\overline{\sigma} \geq \sigma^*$. In contrast, it follows from Corollary 2 that conditional on privacy being feasible ($\alpha^l \geq 1$), $Pr(I)$ increases in $\overline{\sigma}$ for $\overline{\sigma} \in [0, \sigma^{**}]$, and stays constant for $\overline{\sigma} \geq \sigma^{**}$. Note that the same argument applies if we assume that social comparisons disappear when the two agents are assigned to different units within the same firm, and we replace "integration" with "simple organizational structure" (i.e., one unit), and "outsourcing" with "complex organizational structure" (i.e., two units). We summarize these findings below.

Organizational Implication 2. *The probability that integration (simple organizational structure) is preferred to outsourcing (complex organizational structure):*

(i) *Is non-decreasing in the agents' sensitivity to cheating (α^l);*

(ii) *Conditional on privacy not being feasible ($\alpha^l < 1$), decreases in the agents' retaliation power, $\overline{\sigma}$, for $\overline{\sigma} \in [0, \sigma^*]$, and stays constant for $\overline{\sigma} \geq \sigma^*$.*

(iii) *Conditional on privacy being feasible ($\alpha^l \geq 1$), increases in the agents' retaliation power, $\overline{\sigma}$, for $\overline{\sigma} \in [0, \sigma^{**}]$, and stays constant for $\overline{\sigma} \geq \sigma^{**}$.*

The above result highlights novel determinants of firm boundaries and architecture. In particular, it shows that parameters that affect the effectiveness of within-firm privacy policies—namely, the agents' sensitivity to, and hence willingness to retaliate against cheating (α^l), as well as their power to retaliate ($\overline{\sigma}$)—importantly affect distortions in firm boundaries and organizational architecture. The easier it is to sustain privacy, the lower the need to distort boundaries and architecture to reduce social comparisons, and hence the more likely that a firm will be integrated and/or will have a simpler organizational structure.

Another novel implication of our analysis is that organizational distortions co-vary with the agents' retaliation power in a non-monotonic way. The intuition for this result goes back to our earlier discussion of the distinct roles of retaliation under transparency (output destruction) and privacy (increased credibility of the principal, and hence of informal contracts). This implies that when privacy is not feasible, high retaliation power imposes social comparison costs on the organization, and therefore encourages suboptimal organizational policies (outsourcing, complex architecture) as means to reduce social comparisons. In contrast, when privacy is potentially feasible, high retaliation power may increase the principal's ability to commit to informal contracts, thereby facilitating privacy and rendering distortions in firm boundaries and architecture unnecessary.

7. Conclusion

This paper has developed a formal model to analyze how organizations manage social comparisons among their members. We have argued that an organization's members are less likely to compare each other's status when status differences are not explicit and formalized. In such a context, the organization may eliminate social comparisons by offering formally homogeneous contract terms to its employees while relying on private, informal agreements to optimally differentiate their compensation. We have also shown that when the organization does not have sufficient credibility to commit to informal agreements, it may be forced to distort the employees' task assignment, as well as its boundaries and internal architecture in order to minimize social comparison costs. The model provides a rich set of predictions that link the strength of informal relationships inside the firm to observable policy outcomes such as pay compression, firm boundaries, task assignment and internal organization.

The model presented in this paper provides a tractable framework to analyze the role of social comparisons in organization, which may be extended in several directions. First, one could extend our model to settings where the employees engage in team production, such that their efforts are complementary, and explore how such complementarity may affect both the employees' sensitivity to social comparisons (for instance, due to enhanced proximity and social interaction) and their retaliation power. Second, and related, one may incorporate firm

size into the model, and explore how it may affect social comparisons and the way the organization optimally responds to them. Third, the model could be extended to settings where employees' moral hazard and/or relationship-specific investments are relevant. These extensions may create potential beneficial effects of social comparisons. For instance, in a moral hazard model with incentive pay, the agents may increase their efforts to reduce their own frustration from social comparisons, as in Kragl and Schmid (2009). On the other hand, by committing the firm to raise the compensation of the agent with lower outside option, social comparisons may provide the latter with an incentive to undertake specific investments in the firm. Once benefits of social comparisons are introduced, the desirability and specific features of a privacy policy may change in interesting ways. Finally, it may be interesting, and empirically relevant, to incorporate non-monetary compensation into our framework. Hard to observe, non-monetary aspects of compensation may enable firms to reduce social comparisons via private informal agreements with employees even in settings where keeping compensation secret is difficult due to legal constraints or social norms.

We hope our analytical framework for analyzing the governance of social comparisons in organizations will serve as a basis for future theoretical and empirical research on this fundamental dimension of firm governance.

References

Adams, J. S., (1963), "Toward an Understanding of Inequity," *Journal of Abnormal and Social Psychology* 67(5): 422-436.

Akerlof, G., (1982), "Labor Contracts as Partial Gift Exchange," *Quarterly Journal of Economics* 97: 543-569.

Akerlof, G., and J. Yellen, (1990), "The Fair Wage-Effort Hypothesis and Unemployment," *Quarterly Journal of Economics* 105: 255-283.

Antras, P., and C. Foley, (2015), "Poultry in Motion: A Study of International Trade Finance Practices," *Journal of Political Economy* 123: 853-901.

Ashraf, A., (2018), "Do Performance Ranks Increase Productivity? Evidence from a Field Experiment," Warwick University Working Paper.

Baker, G., R. Gibbons, and K. J. Murphy, (1994), "Subjective Performance Measures in Optimal Incentive Contracts," *Quarterly Journal of Economics* 109(4): 1125-1156.

Baker, G., R. Gibbons, and K. J. Murphy (2002), "Relational Contracts and the Theory of the Firm," *Quarterly Journal of Economics* 117: 39-84.

Bartling, B., and F. von Siemens, (2010), "The Intensity of Incentives in Firms and Markets: Moral Hazard with Envious Agents," *Labour Economics* 17: 598-607.

Battigalli, P., and G. Maggi, (2008), "Costly contracting in a long-term relationship," *The RAND Journal of Economics* 39(2): 352-377.

Ben-Shahar, O., and L. Bernstein, (2000), "The Secrecy Interest in Contract Law," *Yale Law Journal* 109: 1885-1925.

Bolton, P., and A. Ockenfels, (2000), "A Theory of Equity, Reciprocity and Competition," *American Economic Review* 90: 166-193.

Bradler, C., R. Dur, S. Neckermann, and A. Non, (2016), "Employee Recognition and Performance: A Field Experiment," *Management Science* 62: 3085-99.

Breza, E., S. Kaur, and Y. Shamdasani, (2018), "The Morale Effects of Pay Inequality," *Quarterly Journal of Economics* 133: 611-633.

Burkus, D., (2016), "Why Keeping Salaries a Secret May Hurt your Company," *Harvard Business Review*, March 2016.

Camerer, C., and R. Weber, (2013), "Experimental Organizational Economics" in Gibbons, R., and J. Roberts (eds.), *The Handbook of Organizational Economics* (pp. 213-262), Princeton: Princeton University Press.

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.

Cohn, A., E. Fehr, B. Hermann, and F. Schneider, (2014), "Social Comparison and Effort Provision: Evidence from a Field Experiment," *Journal of the European Economic Association* 12: 877-898.

Contreras, O., and G. Zanarone (2017), "Managing Social Comparison Costs in Organizations," Working Paper.

Cremer, J., L. Garicano, and A. Prat, (2007), "Language and the Theory of the Firm," *Quarterly Journal of Economics* 122: 373-407.

Dahl, G., and M. Ransom, (1999), "Does Where You Stand Depend on Where You Sit? Tithing Donations and Self-Serving Beliefs," *American Economic Review* 89: 703-727.

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

Edwards, M., (2005), "The Law and Social Norms of Pay Secrecy," *Journal of Employment and Labor Law* 26: 41-63.

Englmaier, F., and S. Leider, (2012), "Contractual and Organizational Structure with Reciprocal Agents," *American Economic Journal: Microeconomics* 4(2): 146-83.

Englmaier, F., and A. Wambach, (2010), "Optimal Incentive Contracts under Inequity Aversion," *Games and Economic Behavior* 69: 312-328.

Fahn, M., A. Schade, and K. Schüßler, (2017), "What Drives Reciprocal Behavior? The Optimal Provision of Incentives over the Course of Careers," CESIFO Working Paper 6635.

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

Fehr, E., G. Kirchsteiger, and A. Riedl, (1998), "Gift Exchange and Reciprocity in Competitive Experimental Markets," *European Economic Review* 42: 1-34.

Fehr, E., and K. Schmidt, (1999), "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics* 114: 817-868.

Fehr, E., O. Hart, and C. Zender, (2015), "How do Informal Agreements and Revision Shape Contractual Reference Points?" *Journal of the European Economic Association* 13: 1-28.

Futrell, C., (1978), "Effects of Pay Disclosure on Satisfaction for Sales Managers," *Academy of Management Journal* 21: 140-144.

Gil, R., and G. Zanarone, (2017), "Formal and Informal Contracting: Theory and Evidence," *Annual Review of Law and Social Science* 13: 141-159.

Gil, R., and G. Zanarone, (2018), "On the Determinants and Consequences of Informal Contracting," *Journal of Economics and Management Strategy* (forthcoming, DOI: 10.1111/jems.12246).

Gillan, S., J. Hatzell, and R. Parrino, (2009), "Explicit Versus Implicit Contracts: Evidence from CEO Employment Agreements," *Journal of Finance* 64: 1629-55.

Grossman, S., and O. Hart, (1986), "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy* 94: 2691-2719.

Guiso, L., P. Sapienza, and L. Zingales, (2008), "Social Capital as Good Culture," *Journal of the European Economic Association* 6: 295-320.

Guiso, L., P. Sapienza, and L. Zingales, (2015), "The Value of Corporate Culture," *Journal of Financial Economics* 117: 60-76.

Haisley, E., and R. Weber, (2010), "Self-Serving Interpretations of Ambiguity in Other-Regarding Behavior," *Games and Economic Behavior* 68: 614-625.

Hall, B., (2000), "Compensation and Performance Evaluation at Arrow Electronics," Harvard Business School Case 9-800-290.

Halonen, M., and O. Hart, (2013), "More is Less: Why Parties May Deliberately Write Incomplete Contracts," Harvard University Working Paper.

Hart, O., (2009), "Hold-Up, Asset Ownership, and Reference Points," *Quarterly Journal of Economics* 124: 267-300.

Hart, O., and B. Holmstrom, (2010), "A Theory of Firm Scope," *Quarterly Journal of Economics* 125: 483-513.

Hart, O., and J. Moore, (2008), "Contracts as Reference Points," *Quarterly Journal of Economics* 123(1): 1-48.

Hill, A., (2016), "Pay Transparency is the Last Taboo in Business," *Financial Times*.

Holmstrom, B., and P. Milgrom, (1991), "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics and Organization* 7: 24-52.

Johnson, S., J. McMillan, and C. Woodruff, (2002), "Courts and Relational Contracts," *Journal of Law, Economics, and Organization* 18: 221-77.

Klein, B., (2000), "The Role of Incomplete Contracts in Self-Enforcing Relationships," *Revue d'Économie Industrielle* 92: 67-80.

Kragl, J., (2015), "Group versus Individual Performance Pay in Relational Employment Contracts when Workers are Envious," *Journal of Economics and Management Strategy* 24(1): 131-150.

Kragl, J., and J. Schmid, (2009), "The Impact of Envy on Relational Employment Contracts," *Journal of Economic Behavior and Organization* 72(2): 766-779.

Kvaløy, O., and T. E. Olsen, (2009), "Endogenous Verifiability and Relational Contracting," *American Economic Review* 99(5): 2193-2208.

Kosova, R., and G. Sertsios, (2018), "An Empirical Analysis of Self-Enforcement Mechanisms: Evidence from Hotel Franchising," *Management Science* 64: 43-63.

Kunda, Z., (1990), "The Case for Motivated Reasoning," *Psychological Bulletin* 108: 480-498.

Lafontaine, F., and M. Slade, (2013), "Inter-Firm Contracts: The Evidence," in Gibbons, R., and J. Roberts (editors), *The Handbook of Organizational Economics*, Princeton: Princeton University Press.

Levin, J., (2003), "Relational Incentive Contracts," *American Economic Review* 93(3): 835-857.

MacLeod, W. B., (2003), "Optimal Contracting with Subjective Evaluation," *American Economic Review* 93(1): 216-240.

MacLeod, W. B., (2007), "Reputations, Relationships, and Contract Enforcement," *Journal of Economic Literature* 45: 595-628.

MacLeod, W. B., and J. M. Malcomson, (1989), "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica* 57(2): 447-80.

Malcomson, J. M., (2013), "Relational Incentive Contracts," In Gibbons, R., and J. Roberts (editors), *The Handbook of Organizational Economics*, Princeton: Princeton University Press.

McAdams, R., (1992), "Relative Preferences," *Yale Law Journal* 9: 1-104.

McAdams, R., (1995), "Cooperation and Conflict: The Economics of Group Status Production and Race Discrimination," *Harvard Law Review* 108: 1008-1083.

Nickerson, J. A., and T. R. Zenger, (2008), "Envy, Comparison Costs, and the Economic Theory of the Firm," *Strategic Management Journal* 29(13): 1429-1449.

Ockenfels, A., D. Sliwka, and P. Werner, (2015), "Bonus Payments and Reference Point Violations," *Management Science* 61: 1496-1513.

Shaw, J., (2014), "Pay Dispersion," *Annual Review of Organizational Psychology and Organizational Behavior* 1: 521-44.

Williamson, O., (1975), *Markets and Hierarchies: Antitrust Implications*, New York: New York Free Press.

Zanarone, G., (2013), "Contract Adaptation under Legal Constraints," *Journal of Law, Economics and Organization* 29: 799-834.

Zanarone, G., D. Lo, and T. Madsen, (2016), "The Double-Edged Effect of Knowledge Acquisition: How Contracts Safeguard Pre-Existing Resources," *Strategic Management Journal* 37: 2104-2120.

Appendix I: Proofs

Proof of Proposition 3

Let $\bar{a}(\overline{\sigma})$ be implicitly defined by $\overline{\sigma} = \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(\bar{a}(\overline{\sigma}))}$. Note that $\bar{a}(\overline{\sigma}) \in (0, a^{FB}]$ by assumption 3 and by the fact that $y(\bullet)$ is strictly increasing with $y(0) = 0$. Let the principal's utility in the $[0, \bar{a}]$ interval be $u_P^{[0,\bar{a}(\overline{\sigma})]}(a_2) \equiv \left[1 - \frac{\alpha^d}{1+\alpha^d}\overline{\sigma}\right] y(a_2) - C(a_2) - \underline{u}_2 - \frac{\alpha^d(\underline{u}_1 - \underline{u}_2)}{1+\alpha^d}$. Since the previous function is strictly concave, and since $a^*(\overline{\sigma}) > 0$ by the assumption that $\lim_{a \to 0} y'(a) = \infty$, the optimal value of a_2 on the interval $[0, \bar{a}(\overline{\sigma})]$ equals $min\{a^*(\overline{\sigma}), \bar{a}(\overline{\sigma})\}$. Also, let the principal's utility in the $[\bar{a}(\overline{\sigma}), \infty]$ interval be $u_P^{[\bar{a}(\overline{\sigma}),\infty]}(a_2) \equiv y(a_2) - C(a_2) - \underline{u}_2 - \alpha^d[\underline{u}_1 - \underline{u}_2]$. Since the previous function is strictly concave and maximized at a^{FB}, the optimal value of a_2 on the interval $[\bar{a}(\overline{\sigma}), \infty]$ equals $max\{\bar{a}(\overline{\sigma}), a^{FB}\}$. Finally, let σ^* be the (unique) positive real number k that solves $u_P^{[\bar{a}(\overline{\sigma}),\infty]}(a^{FB}) = u_P^{[0,\bar{a}(\overline{\sigma})]}\left(a^*(k)\right)$.

The proof then consists of two steps.

Step 1. (i) If $\overline{\sigma} < \sigma^*$, then $u_P^{[\bar{a}(\overline{\sigma}),\infty]}(a^{FB}) < u_P^{[0,\bar{a}(\overline{\sigma})]}\left(a^*(\overline{\sigma})\right)$. (ii) If $\overline{\sigma} = \sigma^*$, then $u_P^{[\bar{a}(\overline{\sigma}),\infty]}(a^{FB}) = u_P^{[0,\bar{a}(\overline{\sigma})]}\left(a^*(\overline{\sigma})\right)$. (ii) If $\overline{\sigma} > \sigma^*$ then $u_P^{[\bar{a}(\overline{\sigma}),\infty]}(a^{FB}) > u_P^{[0,\bar{a}(\overline{\sigma})]}\left(a^*(\overline{\sigma})\right)$.

Proof of Step 1. By definition, σ^* solves $u_P^{[\bar{a}(\overline{\sigma}),\infty]}(a^{FB}) = u_P^{[0,\bar{a}(\overline{\sigma})]}\left(a^*(\overline{\sigma})\right)$. The result then follows since $u_P^{[\bar{a}(\overline{\sigma}),\infty]}(a^{FB})$ is independent of $\overline{\sigma}$, whereas $u_P^{[0,\bar{a}(\overline{\sigma})]}\left(a^*(\overline{\sigma})\right)$ is strictly decreasing in $\overline{\sigma}$. ■

Given step 1 above, if we can prove that $a^*(\overline{\sigma}) \leq \bar{a}(\overline{\sigma})$ for $\overline{\sigma} < \sigma^*$, it would follow that if $\overline{\sigma} < \sigma^*$, then: (i) the optimal value of a_2 on the interval $[0, \bar{a}(\overline{\sigma})]$ equals $a^*(\overline{\sigma})$; and (ii) the principal's utility under $a^*(\overline{\sigma})$ is higher than $u_P^{[\bar{a}(\overline{\sigma}),\infty]}(a^{FB})$, and thus higher than his utility under the optimal value of a_2 on the interval $[\bar{a}(\overline{\sigma}), \infty]$—i.e., $max\{\bar{a}(\overline{\sigma}), a^{FB}\}$. Therefore, it would follow that $a^*(\overline{\sigma})$ is the global optimum, proving part (i) of Proposition 3.

Similarly, if we can prove that $a^{FB} \geq \bar{a}(\overline{\sigma})$ for $\overline{\sigma} > \sigma^*$, it would follow that if $\overline{\sigma} > \sigma^*$, then: (i) the optimal value of a_2 on the interval $[\bar{a}(\overline{\sigma}), \infty]$ equals a^{FB}; and (ii) the principal's utility

under a^{FB} is higher than $u_P^{[0,\bar{a}(\overline{\sigma})]}\big(a^*(\overline{\sigma})\big)$, and thus higher than his utility under the optimal value of a_2 on the interval $[0, \bar{a}(\overline{\sigma})]$—i.e., *min*$\{a^*(\overline{\sigma}), \bar{a}(\overline{\sigma})\}$. Therefore, it would follow that a^{FB} is the global optimum, proving part (ii) of Proposition 3.

Finally, if we can prove that both $a^*(\overline{\sigma}) \leq \bar{a}(\overline{\sigma})$ and $a^{FB} \geq \bar{a}(\overline{\sigma})$ for $\overline{\sigma} = \sigma^*$, it would follow that if $\overline{\sigma} = \sigma^*$, then: (i) the optimal value of a_2 on the interval $[0, \bar{a}(\overline{\sigma})]$ equals $a^*(\overline{\sigma})$, whereas the optimal value of a_2 on the interval $[\bar{a}(\overline{\sigma}), \infty]$ equals a^{FB}; and (ii) the principal's utility under a^{FB} is identical to his utility under $a^*(\overline{\sigma})$. Therefore, it would follow that both a^{FB} and $a^*(\overline{\sigma})$ are optimal, which would prove part (iii) of Proposition 3.

Taken together, the three previous paragraphs imply that to complete the proof it suffices to show that $a^*(\overline{\sigma}) \leq \bar{a}(\overline{\sigma})$ for $\overline{\sigma} \leq \sigma^*$ and $a^{FB} \geq \bar{a}(\overline{\sigma})$ for $\overline{\sigma} \geq \sigma^*$. This is done in the next step.

Step 2. (i) $a^*(\overline{\sigma}) \leq \bar{a}(\overline{\sigma})$ for $\overline{\sigma} \leq \sigma^*$, and (ii) $a^{FB} \geq \bar{a}(\overline{\sigma})$ for $\overline{\sigma} \geq \sigma^*$.

Proof of Step 2. Let $\sigma_1 = \frac{a^d[\underline{u}_1 - \underline{u}_2]}{y(a^{FB})}$ be the (unique) value of $\overline{\sigma}$ such that $\bar{a}(\overline{\sigma}) = a^{FB}$. Let $\sigma_2 = \frac{a^d[\underline{u}_1 - \underline{u}_2]}{y(a^*(\overline{\sigma}))}$ be the smallest value of $\overline{\sigma}$ such that $\bar{a}(\overline{\sigma}) = a^*(\overline{\sigma})$. Since $\bar{a}(\overline{\sigma})$ and $a^*(\overline{\sigma})$ are both continuous and strictly decreasing, $\lim_{\overline{\sigma} \to 0} \bar{a}(\overline{\sigma}) = \infty$, and $a^*(0) = a^{FB}$, it follows that $\sigma_1 < \sigma_2$.

Notice that if we can prove that $\sigma^* \geq \sigma_1$, then the definition of σ_1, together with the fact that $\bar{a}(\overline{\sigma})$ is strictly decreasing, would imply that $a^{FB} \geq \bar{a}(\overline{\sigma})$ for $\overline{\sigma} \geq \sigma^*$, as desired. Similarly, if we can prove that $\sigma^* \leq \sigma_2$, then the definition of σ_2, together with the fact that $\bar{a}(\overline{\sigma})$ and $a^*(\overline{\sigma})$ are both strictly decreasing, would imply that $a^*(\overline{\sigma}) \leq \bar{a}(\overline{\sigma})$ for $\overline{\sigma} \leq \sigma^*$, as desired.

To prove that $\sigma^* \geq \sigma_1$, notice that $u_P^{[0,\bar{a}(\overline{\sigma})]}(a^{FB}) > u_P^{[\bar{a}(\overline{\sigma}),\infty]}(a^{FB})$ at σ_1. Since $u_P^{[0,\bar{a}(\overline{\sigma})]}\big(a^*(\sigma_1)\big) \geq u_P^{[0,\bar{a}(\overline{\sigma})]}(a^{FB})$ at σ_1 by the definition of $a^*(\overline{\sigma})$, it follows that $u_P^{[0,\bar{a}(\overline{\sigma})]}\big(a^*(\sigma_1)\big) > u_P^{[\bar{a}(\overline{\sigma}),\infty]}(a^{FB})$.

But since σ^* is implicitly defined by $u_P^{[0,\bar{a}(\overline{\sigma})]}\left(a^*(\overline{\sigma})\right) = u_P^{[\bar{a}(\overline{\sigma}),\infty]}(a^{FB})$, and since $u_P^{[0,\bar{a}(\overline{\sigma})]}\left(a^*(\overline{\sigma})\right)$ is strictly decreasing in $\overline{\sigma}$, it then follows that $\sigma^* > \sigma_1$, as desired.

Finally, to prove that $\sigma^* \leq \sigma_2$, notice that $u_P^{[0,\bar{a}(\overline{\sigma})]}(a^*(\sigma_2)) < u_P^{[\bar{a}(\overline{\sigma}),\infty]}(a^{FB})$. But since σ^* is implicitly defined by $u_P^{[0,\bar{a}(\overline{\sigma})]}(a^*(\sigma^*)) = u_P^{[\bar{a}(\overline{\sigma}),\infty]}(a^{FB})$, and since $u_P^{[0,\bar{a}(\overline{\sigma})]}(a^*(\overline{\sigma}))$ is strictly decreasing in $\overline{\sigma}$,it then follows that $\sigma^* < \sigma_2$, as desired. ■

Proof of Corollary 1

By Proposition 3, we know that if $\overline{\sigma} \geq \sigma^*$, then $a_2 = a^{FB}$ and $u_P = 2[y(a^{FB}) - C(a^{FB})] - \alpha^d\left[\underline{u}_1 - \underline{u}_2\right] - \underline{u}_1 - \underline{u}_2$, both of which are independent of $\overline{\sigma}$. This proves that a_2 and u_P are both constant in $\overline{\sigma}$ on the interval [σ^*,1], with $a_2 = a^{FB}$. To prove that a_2 and u_P are both strictly decreasing on the interval [0,σ^*], note that, by Proposition 3, we know that if $\overline{\sigma} \leq \sigma^*$, then $a_2 = a^*(\overline{\sigma})$ and $u_P = y(a^{FB}) + \left[1 + \frac{\alpha^d}{1+\alpha^d}\overline{\sigma}\right] y\left(a^*(\overline{\sigma})\right) - C(a^{FB}) - C(a_2^*) - \underline{u}_1 - \frac{\underline{u}_2 + \alpha^d \underline{u}_1}{1+\alpha^d}$. Moreover, recall that $a^*(\overline{\sigma})$ is implicitly defined by $\left[1 - \frac{\alpha^d}{1+\alpha^d}\overline{\sigma}\right] y'\left(a^*(\overline{\sigma})\right) - C'\left(a^*(\overline{\sigma})\right) = 0$, from which it is straightforward to verify that both $a^*(\overline{\sigma})$ and $\left[1 - \frac{\alpha^d}{1+\alpha^d}\overline{\sigma}\right] y\left(a^*(\overline{\sigma})\right) - C\left(a^*(\overline{\sigma})\right)$ are strictly decreasing in $\overline{\sigma}$. ■

Proof of Proposition 4

To prove the proposition we must show that (IC_i) binds at $\alpha^l < 1$, that is: $w_i^I = w_i^F$ for $i \in \{1,2\}$. Since $w_i^I \geq w_i^F$ by condition (IC_i), it suffices to show that $w_i^I \leq w_i^F$. To this end, consider condition (IC_P) and suppose that $\overline{\sigma} \leq \frac{\alpha^d A_i(\bullet)}{y(a_i)}$. Since $w_i^I \geq w_i^F$, it must then be that $\overline{\sigma} \leq \frac{\alpha^d A_i + \alpha^l(w_i^I - w_i^F)}{y(a_i)}$. Thus, condition (IC_P) reduces to $w_i^I \leq w_i^F$, as desired. Suppose next that $\overline{\sigma} > \frac{\alpha^d A_i(\bullet)}{y(a_i)}$, so that (IC_P) becomes $w_i^I - w_i^F \leq min\left\{\overline{\sigma}, \frac{\alpha^d A_i + \alpha^l(w_i^I - w_i^F)}{y(a_i)}\right\} y(a_i) - \alpha^d A_i$. Using the

fact that $min\left\{\overline{\sigma}, \frac{\alpha^d A_i + \alpha^l (w_i^I - w_i^F)}{y(a_i)}\right\} \leq \frac{\alpha^d A_i + \alpha^l (w_i^I - w_i^F)}{y(a_i)}$ and doing some calculation, (IC_P) can be written as $(1 - \alpha^l)(w_i^I - w_i^F) \leq 0$. The result that $w_i^I \leq w_i^F$ then follows directly from the assumption that $\alpha^l < 1$. ■

Proof of Proposition 5

Note first that if the proposed contract is self-enforcing, it must be optimal since it completely eliminates social comparison costs and extracts all rents from the two agents. By substituting the proposed contract into the self-enforcement constraints, it is straightforward to verify that conditions (PC_i) and (IC_i) are all satisfied, whereas condition (IC_P) will be satisfied as long as $\underline{u}_1 - \underline{u}_2 - min\left\{\overline{\sigma}, \frac{\alpha^l[\underline{u}_1 - \underline{u}_2]}{y(a^{FB})}\right\} y(a^{FB}) \leq 0$. If $\overline{\sigma} \geq \frac{\alpha^l[\underline{u}_1 - \underline{u}_2]}{y(a^{FB})}$, then (IC_P) becomes $\underline{u}_1 - \underline{u}_2 - \alpha^l[\underline{u}_1 - \underline{u}_2] \leq 0$, which is satisfied since $\alpha^l \geq 1$ and $\underline{u}_1 - \underline{u}_2 > 0$. Alternatively, if $\overline{\sigma} < \frac{\alpha^l[\underline{u}_1 - \underline{u}_2]}{y(a^{FB})}$, then (IC_P) becomes $\underline{u}_1 - \underline{u}_2 - \overline{\sigma} y(a^{FB}) \leq 0$, or equivalently, $\overline{\sigma} \geq \frac{\underline{u}_1 - \underline{u}_2}{y(a^{FB})}$, which is true by assumption. Taken together the two previous cases imply that (IC_P) is satisfied under the proposed contract, as desired. ■

Proof of Lemma 5

Suppose to the contrary that $w_2^I > w_2^F$. By Lemma 4, we know that $w_1^F - C(a_1) \geq w_2^F - C(a_1)$. Suppose first that $w_1^F - C(a_1) > w_2^F - C(a_1)$. Consider now increasing w_2^F by an arbitrarily small $\varepsilon > 0$. Using the assumption that $w_2^I > w_2^F$, we can easily verify that the perturbation induces a new self-enforcing contract in which either retaliation by agent 2 decreases or agent 2's participation constraint becomes slack. Since in the former case the principal's utility increases and in the second case we can use the same perturbations as in the proof of Lemma 3 to construct a new self-enforcing contract in which the principal is strictly better off, this contradicts the optimality of the original contract.

Suppose now that $w_1^F - C(a_1) = w_2^F - C(a_1)$. We first prove that $w_1^I > w_1^F$. To see this, notice that since (PC_1) and (PC_2) are both binding by Lemma 3 and $\underline{u}_1 > \underline{u}_2$ by assumption, it follows that $w_1^I - w_2^I > C(a_1) - C(a_2)$. Moreover, since $w_1^F - C(a_1) = w_2^F - C(a_1)$ by assumption, it follows that $C(a_1) - C(a_2) = w_1^F - w_2^F$. Combining the two previous observations implies that $w_1^I - w_1^F > w_2^I - w_2^F$, which, since $w_2^I - w_2^F > 0$ by assumption, requires $w_1^I - w_1^F > 0$, as desired.

Consider now increasing both w_1^F and w_2^F by an arbitrarily small $\varepsilon > 0$. Using the assumption that $w_2^I > w_2^F$ and the previous result that $w_1^I > w_1^F$, we can verify that the perturbation induces a new equilibrium in which the principal's utility remains unchanged and the principal's incentive constraint (IC_{Pi}) becomes slack, for all i. Notice that if $a_1 = a_2 = a^{FB}$, Lemma 3, and the assumption that $w_1^F - C(a_1) = w_2^F - C(a_1)$, would imply that $w_1^I = \underline{u}_1 + C(a^{FB})$ and $p_1^F = \underline{u}_2 + C(a^{FB})$, which substituted into (IC_{P1}) would imply that $\overline{\sigma} \geq [\underline{u}_1 - \underline{u}_2]/y(a^{FB})$, a contradiction. Thus, to complete our proof it suffices to show that if (IC_{Pi}) is slack, then $a_1 = a_2 = a^{FB}$. To see this, suppose first that (IC_{Pi}) is slack and $a_i > a^{FB}$ for some $i \in \{1,2\}$. Then, it is easy to verify that decreasing a_i by an arbitrarily small $\varepsilon > 0$, and decreasing w_i^F by $C'(a_i)\varepsilon$, induces a new self-enforcing contract with higher principal's utility, which contradicts optimality of the initial contract. Suppose next that (IC_{Pi}) is slack and $a_i < a^{FB}$ for some $i \in \{1,2\}$. Then, it is easy to verify that increasing a_i by an arbitrarily small $\varepsilon > 0$, and w_i^I and w_i^F by $C'(a_i)\varepsilon$, induces a new self-enforcing contract with higher principal's utility, which again contradicts optimality of the initial contract. Therefore, it must be that if (IC_{Pi}) is slack, $a_i = a^{FB}$ for all $i \in \{1,2\}$, as desired. ■

Proof of Lemma 6

Proof. We first show that (IC_{P1}) is binding, and then we show that $\overline{\sigma} \leq \frac{\alpha^l[w_1^I - w_1^F]}{y(a_1)}$, which implies that the last term in the left-hand side of (IC_{P1}) is equal to $\overline{\sigma}y(a_1^l)$. Proof of statement I: (IC_{P1}) is binding.

Suppose to the contrary that (IC_{P1}) is slack. There are two cases to be considered: $A_2 = A_1 = 0$, and $A_2 > 0$. If $A_2 = 0$, it must be that $w_1^F - C(a_1) = w_2^F - C(a_2)$, which by Lemmas 3 and 5 can be rewritten as $w_1^F = \underline{u}_2 + C(a_1)$. Since $w_1^I = \underline{u}_1 + C(a_1)$ by Lemmas 3 and 4, substituting w_1^F and w_1^I into (IC_{P1}) yields $\underline{u}_1 - \underline{u}_2 \leq min\left\{\overline{\sigma}, \frac{\alpha^l[\underline{u}_1 - \underline{u}_2]}{y(a_1)}\right\} y(a_1)$. This implies that since $\overline{\sigma} < \frac{\alpha^l[\underline{u}_1 - \underline{u}_2]}{y(a^{FB})}$ by assumption, (IC_{P1}) can only hold if $a_1 > a^{FB}$. But then, if (IC_{P1}) is slack the principal can increase his utility, without affecting A_2 and the self-enforcement conditions, by reducing a_1 by an arbitrarily small amount $\varepsilon > 0$ and both w_1^F and w_1^I by $C'(a_1)\varepsilon$. This contradicts the initial assumption that (IC_{P1}) is slack. Consider next the case where $A_2 > 0$. Consider a perturbation to the initial contract whereby w_1^F is reduced by an arbitrarily small amount $\varepsilon > 0$. This perturbation reduces A_2 by $\alpha^d\varepsilon$, and consequently increases the principal's utility by $\alpha^d\varepsilon$ (either through a reduction in the wage premium of agent 2 or through a reduction in retaliation). Moreover, since (IC_{P1}) was initially slack, (IC_{P1}) still holds after the perturbation, while all the other constraints are relaxed. Thus, the new contract is self-enforcing, which contradicts the optimality of the initial contract with slack (IC_{P1}). ■

Proof of statement II: $\overline{\sigma} \leq \frac{\alpha^l[w_1^I - w_1^F]}{y(a_1)}$.

Suppose to the contrary that $\overline{\sigma} > \frac{\alpha^l[w_1^I - w_1^F]}{y(a_1)}$. Then, given statement I above, (IC_{P1}) is given by: $w_1^I - w_1^F = \alpha^l[w_1^I - w_1^F]$. If $\alpha^l > 1$, this is a contradiction. If $\alpha^l = 1$, the binding (IC_{P1}) is satisfied for any pair of w_1^I, w_1^F, so P can maximize his utility, while satisfying all the self-enforcement constraints, by choosing $a_i = a^{FB}$ for all i, and $w_1^F = w_2^F = w_2^I = \underline{u}_2 + C(a^{FB})$ (with $w_2^F = w_2^I$ coming from Lemma 5 and $w_2^I = \underline{u}_2 + C(a^{FB})$ coming from Lemma 3 and from the fact that $A_2 = 0$ under this optimal contract). Moreover, Lemmas 3 and 4 imply that $w_1^I = \underline{u}_1 + C(a^{FB})$. Combining all these facts, our assumption that $\overline{\sigma} > \frac{\alpha^l[w_1^I - w_1^F]}{y(a_1)}$ can be rewritten as $\overline{\sigma} > \frac{[\underline{u}_1 - \underline{u}_2]}{y(a^{FB})}$, which is a contradiction. ■

Proof of Proposition 6

To solve the principal's program, (11), under $\overline{\sigma} < \sigma^{**}$, note that by Lemma 5, we can write $w_2^I = w_2^F = w_2$ and ignore constraints (IC_2) and (IC_{P2}). Moreover, by Lemma 6 we can ignore constraint (IC_1). Then, applying lemmas 3 through 5, the principal's program can be rewritten as follows:

$$\max_{a_1,a_2,w_1^I,w_1^F,w_2} \quad y(a_1) + \left(1 - min\left\{\overline{\sigma}, \frac{\alpha^d[w_1^F - C(a_1) - w_2 + C(a_2)]}{y(a_2)}\right\}\right) y(a_2) - w_1^I - w_2$$

subject to

$$w_1^I = C(a_1) + \underline{u}_1, \qquad (PC_1)$$

$$w_2 - C(a_2) - \alpha^d\left[\underline{u}_1 - \overline{\sigma}y(a_1) - w_2 + C(a_2)\right] + min\left\{\overline{\sigma}, \frac{\alpha^d[\underline{u}_1 - \overline{\sigma}y(a_1) - w_2 + C(a_2)]}{y(a_2)}\right\} y(a_2) = \underline{u}_2, \qquad (PC_2)$$

$$w_1^F = \underline{u}_1 + C(a_1) - \overline{\sigma}y(a_1), \text{ and} \qquad (IC_{P1})$$

$$\underline{u}_1 - \overline{\sigma}y(a_1) - w_2 + C(a_2) \geq 0, \qquad (A1)$$

where constraint $(A1)$ requires that agent 1 not suffer from social comparisons, as prescribed by Lemma 4.

Since we have 3 equations that hold as equalities and 5 unknowns, to fully characterize the optimal contract it suffices to find values for a_1 and a_2. Consider the following two cases depending on the value of $min\left\{\overline{\sigma}, \frac{\alpha^d[\underline{u}_1 - \overline{\sigma}y(a_1) - w_2 + C(a_2)]}{y(a_2)}\right\}$.

Case 1. If $\overline{\sigma} \leq \frac{\alpha^d[\underline{u}_1 - \overline{\sigma}y(a_1) - w_2 + C(a_2)]}{y(a_2)}$, then (PC_2) becomes $w_2 = C(a_2) + \frac{\alpha^d[\underline{u}_1 - \overline{\sigma}y(a_1)] + \underline{u}_2 - \overline{\sigma}y(a_2)}{1+\alpha^d}$. Note that the original assumption can be written as $\overline{\sigma} \leq \frac{\alpha^d[\underline{u}_1 - \overline{\sigma}y(a_1) - \underline{u}_2]}{y(a_2)}$, or equivalently, as $\overline{\sigma} \leq \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2) + \alpha^d y(a_1)}$. Notice also that we can ignore constraint $(A1)$ since $\overline{\sigma} \leq \frac{\alpha^d[\underline{u}_1 - \overline{\sigma}y(a_1) - w_2 + C(a_2)]}{y(a_2)}$ directly implies that $\underline{u}_1 - \overline{\sigma}y(a_1) - w_2 + C(a_2) \geq 0$. Substituting (PC_1), (PC_2), and (IC_{P1}) into the objective function we obtain:

$$U_P^1(a_1, a_2) \equiv y(a_1) + (1-\overline{\sigma})y(a_2) - C(a_1) - \underline{u}_1 - C(a_2) - \frac{\alpha^d[\underline{u}_1 - \overline{\sigma}y(a_1)] + \underline{u}_2 - \overline{\sigma}y(a_2)}{1+\alpha^d}.$$

Since the previous expression is strictly concave in a_1 and a_2, the optimal values of a_1 and a_2, assuming that they satisfy $\overline{\sigma} \leq \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2)+\alpha^d y(a_1)}$, are given, respectively, by the following first order conditions:

$y'(a_1)\left[1 + \frac{\alpha^d}{1+\alpha^d}\overline{\sigma}\right] - C'(a_1) = 0$, which implies that $a_1 = a^{**}(\overline{\sigma})$, as defined by (12), and

$\left(1 - \frac{\alpha^d}{1+\alpha^d}\overline{\sigma}\right)y'(a_2) - C'(a_2) = 0$, which implies that $a_2 = a^{*}(\overline{\sigma})$, as defined by (9).

Case 2. If $\overline{\sigma} \geq \frac{\alpha^d[\underline{u}_1 - \overline{\sigma}y(a_1) - w_2 + C(a_2)]}{y(a_2)}$, then (PC_2) becomes $w_2 = \underline{u}_2 + C(a_2)$. Note that the original assumption can be written as $\overline{\sigma} \geq \frac{\alpha^d[\underline{u}_1 - \overline{\sigma}y(a_1) - \underline{u}_2]}{y(a_2)}$, or equivalently, as $\overline{\sigma} \geq \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2)+\alpha^d y(a_1)}$. Substituting (PC_2), (PC_2), and (IC_{P1}) into the objective function we obtain:

$$U_P^2(a_1, a_2) \equiv y(a_1) + y(a_2) - \alpha^d\left[\underline{u}_1 - \overline{\sigma}y(a_1) - \underline{u}_2\right] - C(a_1) - \underline{u}_1 - C(a_2) - \underline{u}_2.$$

The optimal contract then maximizes the previous expression subject to $(A1)$, which can be rewritten as: $\underline{u}_1 - \overline{\sigma}y(a_1) - \underline{u}_2 \geq 0$. Since the objective function is strictly concave in a_1 and a_2, if the optimal contract satisfies $\overline{\sigma} \geq \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2)+\alpha^d y(a_1)}$, the optimal value of a_2 is given by the first order condition $y'(a_2) - C'(a_2) = 0$, which implies that $a_2 = a^{FB}$, and the optimal value of a_1 is given by $\min\{a^{***}(\overline{\sigma}), a^{****}(\overline{\sigma})\} > a^{FB}$, as defined by (13) and (14), respectively.

Given any $\overline{\sigma}$, let $\{a_1^1(\overline{\sigma}), a_2^1(\overline{\sigma})\} = \underset{a_1,a_2}{\arg\max}\, U_P^1(a_1, a_2|\overline{\sigma})$ and let $U_P^1(\overline{\sigma}) \equiv U_P^1[a_1^1(\overline{\sigma}), a_2^1(\overline{\sigma})|\overline{\sigma}]$, i.e., the principal's utility under the optimal contract assuming that his utility is as in case 1, regardless of whether the condition $\overline{\sigma} \leq \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2)+\alpha^d y(a_1)}$ is actually satisfied. Similarly, given any $\overline{\sigma}$, let $\{a_1^2(\overline{\sigma}), a_2^2(\overline{\sigma})\} = \underset{a_1,a_2}{\arg\max}\{U_P^2(a_1, a_2|\overline{\sigma})$ subject to $\underline{u}_1 - \overline{\sigma}y(a_1) - \underline{u}_2 \geq 0\}$ and let $U_P^2(\overline{\sigma}) \equiv U_P^2[a_1^2(\overline{\sigma}), a_2^2(\overline{\sigma})|\overline{\sigma}]$, i.e., the principal's utility under the optimal contract assuming that his utility is as in case 2, regardless of whether the condition

$\overline{\sigma} \geq \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2)+\alpha^d y(a_1)}$ is actually satisfied. It should be clear that $a_j^i(\overline{\sigma}), i, j = 1,2$ is continuous in $\overline{\sigma}$.

The rest of the proof is divided in 5 steps.

Step 1. $a_1^1(\overline{\sigma}) < a_1^2(\overline{\sigma})$.

Proof of Step 1. It immediately follows from the first order conditions (12) and (13) that $a_1^1(\overline{\sigma}) = a^{**}(\overline{\sigma}) < a^{***}(\overline{\sigma})$. Since $a_1^2(\overline{\sigma}) = \min\{a^{***}(\overline{\sigma}), a^{****}(\overline{\sigma})\}$, to prove step 1 it suffices to show that $a^{**}(\overline{\sigma}) < a^{****}(\overline{\sigma})$, or equivalently, that $y(a^{**}(\overline{\sigma})) < y(a^{****}(\overline{\sigma}))$. To this end, note that condition $(A1)$ implies that $\overline{\sigma} y\left(a^{**}(\overline{\sigma})\right) \leq \underline{u}_1 - w_2 + C\left(a^{**}(\overline{\sigma})\right)$, whereas the definition of $a^{****}(\overline{\sigma})$ implies that $\overline{\sigma} y\left(a^{****}(\overline{\sigma})\right) = \underline{u}_1 - \underline{u}_2$. The result that $y(a^{**}(\overline{\sigma})) < y(a^{****}(\overline{\sigma}))$ is then equivalent to $\underline{u}_1 - \underline{u}_2 > \underline{u}_1 - w_2 + C\left(a^{**}(\overline{\sigma})\right)$, or after some calculations, to $w_2 - C\left(a^{**}(\overline{\sigma})\right) - \underline{u}_2 > 0$. But the previous expression follows directly from condition (PC_2), proving the result. ■

Step 2. $U_P^1(\overline{\sigma}) - U_P^2(\overline{\sigma})$ is strictly decreasing.

Proof of Step 2. Suppose first that condition $(A1)$ is slack ($\underline{u}_1 - \overline{\sigma} y\left(a_1^2(\overline{\sigma})\right) - \underline{u}_2 > 0$). Then, using the definitions of $U_P^1(\overline{\sigma})$ and $U_P^2(\overline{\sigma})$ and applying the envelope theorem, we have:

$$\frac{\partial U_P^1(\overline{\sigma})}{\partial \overline{\sigma}} - \frac{\partial U_P^2(\overline{\sigma})}{\partial \overline{\sigma}} = \frac{\alpha^d}{1+\alpha^d}[y(a_1^1(\overline{\sigma})) - y(a_2^1(\overline{\sigma}))] - \alpha^d y(a_1^2(\overline{\sigma})). \quad (A4)$$

The expression in $(A4)$ is strictly negative $a_1^2(\overline{\sigma}) > a_1^1(\overline{\sigma})$ by step 1, and because $\alpha^d > 0$ by assumption.

Suppose next that condition $(A1)$ is binding ($\underline{u}_1 - \overline{\sigma} y\left(a_1^2(\overline{\sigma})\right) - \underline{u}_2 = 0$). Then, using the definitions of $U_P^1(\overline{\sigma})$ and $U_P^2(\overline{\sigma})$ and after some calculations, we can write:

$$\frac{\partial U_P^1(\overline{\sigma})}{\partial \overline{\sigma}} - \frac{\partial U_P^2(\overline{\sigma})}{\partial \overline{\sigma}} = \frac{\alpha^d}{1+\alpha^d} y(a_1^1(\overline{\sigma})) - \frac{C'\left(a_1^2(\overline{\sigma})\right) - y'\left(a_1^2(\overline{\sigma})\right)}{\overline{\sigma} y'\left(a_1^2(\overline{\sigma})\right)} y\left(a_1^2(\overline{\sigma})\right) - \frac{\alpha^d}{1+\alpha^d} y(a_2^1(\overline{\sigma})). \quad (A5)$$

It is easy to verify that the expression in $(A5)$ is negative because $a_1^1(\overline{\sigma}) < a_1^2(\overline{\sigma})$ by step 1 and $a_1^2(\overline{\sigma}) > a^{FB}$ by first order condition (13). ■

Step 3. There exists a value of $\overline{\sigma}$, say $\overline{\sigma}_1$, such that $\overline{\sigma}_1 \in \left(0, [\underline{u}_1 - \underline{u}_2]/y(a^{FB})\right)$ and $U_P^1(\overline{\sigma}_1) = U_P^2(\overline{\sigma}_1)$.

Proof of Step 3. Let $\overline{\sigma}_L$ be implicitly given by $\overline{\sigma}_L = \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2^2(\overline{\sigma}_L)) + \alpha^d y(a_1^2(\overline{\sigma}_L))}$ and let $\overline{\sigma}_H$ be the minimum value of $\overline{\sigma}$ such that $\overline{\sigma}_H = \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2^1(\overline{\sigma}_H)) + \alpha^d y(a_1^1(\overline{\sigma}_H))}$. Using the facts that $a_2^2(\overline{\sigma}) > a_2^1(\overline{\sigma})$ and $a_1^2(\overline{\sigma}) > a_1^1(\overline{\sigma}) > a^{FB}$ for any $\overline{\sigma} \geq 0$, it is easy to check that $0 < \overline{\sigma}_L < \overline{\sigma}_H < \frac{[\underline{u}_1 - \underline{u}_2]}{y(a^{FB})}$. Moreover, using the fact that $U_P^1(a_1, a_2|\overline{\sigma}) = U_P^2(a_1, a_2|\overline{\sigma})$ when $\overline{\sigma} = \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2) + \alpha^d y(a_1)}$, it is easy to check that $U_P^1(\overline{\sigma}_L) > U_P^2(\overline{\sigma}_L)$ and $U_P^1(\overline{\sigma}_H) < U_P^2(\overline{\sigma}_H)$. Since $U_P^1(\overline{\sigma}) - U_P^2(\overline{\sigma})$ is strictly decreasing by step 2 and continuous by the Theorem of the maximum, it then follows that there exist a $\overline{\sigma}_1 \in (\overline{\sigma}_L, \overline{\sigma}_H) \subset \left(0, [\underline{u}_1 - \underline{u}_2]/y(a^{FB})\right)$ such that $U_P^1(\overline{\sigma}_1) = U_P^2(\overline{\sigma}_1)$, as desired. ■

Step 4. We prove the following:

(i) If $\overline{\sigma} \in [0, \overline{\sigma}_1]$, then $\{a_1^1(\overline{\sigma}), a_2^1(\overline{\sigma})\}$ satisfy $\overline{\sigma} \leq \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2^1(\overline{\sigma})) + \alpha^d y(a_1^1(\overline{\sigma}))}$.

(ii) If $\overline{\sigma} \in \left[\overline{\sigma}_1, [\underline{u}_1 - \underline{u}_2]/y(a^{FB})\right]$, then $\{a_1^2(\overline{\sigma}), a_2^2(\overline{\sigma})\}$ satisfy $\overline{\sigma} \geq \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2^2(\overline{\sigma})) + \alpha^d y(a_1^2(\overline{\sigma}))}$.

Proof of Step 4. To prove part (i), note that from the definition of $\overline{\sigma}_H$ and the fact that $0 < \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2^1(0)) + \alpha^d y(a_1^1(0))}$, we know that $\overline{\sigma} \leq \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2^1(\overline{\sigma})) + \alpha^d y(a_1^1(\overline{\sigma}))}$ for $\overline{\sigma} \in [0, \overline{\sigma}_H]$. The result then follows since $\overline{\sigma}_1 < \overline{\sigma}_H$, as shown in the proof of step 3 above.

To prove part (ii), recall that $\overline{\sigma}_L = \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2^2(\overline{\sigma}_L)) + \alpha^d y(a_1^2(\overline{\sigma}_L))}$ by the definition of $\overline{\sigma}_L$. Using the definitions of $a_2^1(\overline{\sigma})$ and $a_2^2(\overline{\sigma})$, we can easily verify that if $\overline{\sigma} \geq \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2^2(\overline{\sigma})) + \alpha^d y(a_1^2(\overline{\sigma}))}$ holds for a given $\overline{\sigma}$, say $\tilde{\sigma}$, it must also holds for any $\overline{\sigma} > \tilde{\sigma}$. The result then follows since $\overline{\sigma}_1 > \overline{\sigma}_L$, as shown in the proof of step 3 above. ■

Step 5. We are now ready to prove parts A, B, and C of Proposition 6.

To prove part A, note that by steps 2, 3 and 4 above, it follows that if $\overline{\sigma} \in [0, \overline{\sigma}_1]$, then (i) $U_P^1(a_1^1(\overline{\sigma}), a_2^1(\overline{\sigma})) \geq U_P^2(a_1^2(\overline{\sigma}), a_2^2(\overline{\sigma}))$, where the inequality is strict unless $\overline{\sigma} = \overline{\sigma}_1$, and (ii) $\{a_1^1(\overline{\sigma}), a_2^1(\overline{\sigma})\}$ satisfies $\overline{\sigma} \leq \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2^1(\overline{\sigma})) + \alpha^d y(a_1^1(\overline{\sigma}))}$. Thus, it follows that the optimal value of w_2 equals $w_2 = C(a_2) + \frac{\alpha^d[\underline{u}_1 - \overline{\sigma}y(a_1)] + \underline{u}_2 - \overline{\sigma}y(a_2)}{1+\alpha^d}$ and that the optimal values of a_1 and a_2 are given by $a_1^1(\overline{\sigma})$ and $a_2^1(\overline{\sigma})$, i.e., the optimal values of a_1 and a_2 solve $y'(a_1)\left[1 + \frac{\alpha^d}{1+\alpha^d}\overline{\sigma}\right] - C'(a_1) = 0$ and $\left(1 - \frac{\alpha^d}{1+\alpha^d}\overline{\sigma}\right)y'(a_2) - C'(a_2) = 0$, as desired.

Similarly, to prove parts B and C, note that by steps 2, 3 and 4 above, it follows that if $\overline{\sigma} \in [\overline{\sigma}_1, (\underline{u}_1 - \underline{u}_2)/y(a^{FB}))$, then (i) $U_P^1(a_1^1(\overline{\sigma}), a_2^1(\overline{\sigma})) \leq U_P^2(a_1^2(\overline{\sigma}), a_2^2(\overline{\sigma}))$, where the inequality is strict unless $\overline{\sigma} = \overline{\sigma}_1$, and (ii) $\{a_1^2(\overline{\sigma}), a_2^2(\overline{\sigma})\}$ satisfies $\overline{\sigma} \geq \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a_2^2(\overline{\sigma})) + \alpha^d y(a_1^2(\overline{\sigma}))}$. Thus, it follows that the optimal value of w_2 equals $w_2 = \underline{u}_2 + C(a_2)$ and that the optimal values of a_1 and a_2 are given by $a_1^2(\overline{\sigma})$ and $a_2^2(\overline{\sigma})$, i.e., the optimal value of a_2 solves $y'(a_2) - C'(a_2) = 0$ (which implies that $a_2 = a^{FB}$) and the optimal value of a_1 is given by $\min\{a^{***}(\overline{\sigma}), a^{****}(\overline{\sigma})\} > a^{FB}$, as defined by (13) and (14), respectively.

It remains only to prove that there exists a $\overline{\sigma}_2 < (\underline{u}_1 - \underline{u}_2)/y(a^{FB})$ such that $a_1 = a^{***}(\overline{\sigma})$ for $\overline{\sigma} \leq \overline{\sigma}_2$ and $a_1 = a^{****}(\overline{\sigma})$ for $\overline{\sigma} > \overline{\sigma}_2$. Note that increasing $\overline{\sigma}$ strictly increases $a^{***}(\overline{\sigma})$ and strictly decreases $a^{****}(\overline{\sigma})$; accordingly, it suffices to prove that there exists a $\overline{\sigma}_2 < (\underline{u}_1 - \underline{u}_2)/y(a^{FB})$ such $a^{***}(\overline{\sigma}) \geq a^{****}(\overline{\sigma})$ and therefore $a_1 = a^{****}(\overline{\sigma})$. To this end, suppose to the contrary that $a^{***}(\overline{\sigma}) < a^{****}(\overline{\sigma})$ for all $\overline{\sigma} < (\underline{u}_1 - \underline{u}_2)/y(a^{FB})$. In this case, $a_1 = a^{***}(\overline{\sigma})$ for all $\overline{\sigma} < (\underline{u}_1 - \underline{u}_2)/y(a^{FB})$. However, since $a^{***}(\overline{\sigma})$ is strictly increasing in $\overline{\sigma}$ and satisfies $a^{***}(\overline{\sigma}) > a^{FB}$, it follows that $\overline{\sigma}y(a^{***}(\overline{\sigma})) > \underline{u}_1 - \underline{u}_2$ for $\overline{\sigma}$ sufficiently close to $(\underline{u}_1 - \underline{u}_2)/y(a^{FB})$, which is a contradiction. ■

Proof of Corollary 2

Part (i)

We know from Proposition 6 that for any $\overline{\sigma} \in [0, \overline{\sigma}_1]$, the principal's utility is given by $U_P^1(a^{**}(\overline{\sigma}), a^*(\overline{\sigma}))$. Applying the envelope theorem we have $\frac{\partial U_P^1(a^{**}(\overline{\sigma}), a^*(\overline{\sigma}))}{\partial \overline{\sigma}} = \frac{\alpha^d [y(a^{**}(\overline{\sigma})) - y(a^*(\overline{\sigma}))]}{1+\alpha^d} > 0$, where the last inequality holds since $a^{**}(\overline{\sigma}) > a^*(\overline{\sigma})$ and $y'(\bullet) > 0$.

For $\overline{\sigma} \in [\overline{\sigma}_1, \overline{\sigma}_2]$, we know that the principal's utility is given by $U_P^2(a^{***}(\overline{\sigma}), a^{FB})$. Applying the envelope theorem we have $\frac{\partial U_P^2(a^{***}(\overline{\sigma}), a^{FB})}{\partial \overline{\sigma}} = \alpha^d y(a^{***}(\overline{\sigma})) > 0$.

Finally, for $\overline{\sigma} \in [\overline{\sigma}_2, \sigma^{**}]$, we know that the principal's utility is given by $U_P^2(a^{****}(\overline{\sigma}), a^{FB})$, where it should be noted that $a^{****}(\overline{\sigma})$ solves a binding-constrained maximization problem and therefore the envelope theorem does not apply. After straightforward calculations, however, it can be checked that $\frac{\partial U_P^2(a^{****}(\overline{\sigma}), a^{FB})}{\partial \overline{\sigma}} = [y'(a^{****}(\overline{\sigma})) - C'(a^{****}(\overline{\sigma}))] \frac{\partial a^{****}(\overline{\sigma})}{\partial \overline{\sigma}}$, which is strictly positive because $a^{****}(\overline{\sigma}) > a^{FB}$ and $\frac{\partial a^{****}(\overline{\sigma})}{\partial \overline{\sigma}} < 0$.

The above findings imply that the principal's utility increases in $\overline{\sigma}$ on the interval $[0, \overline{\sigma}_2]$. Moreover, it follows directly from Proposition 5 that the principal's utility is equal to the first-best level for any $\overline{\sigma} \geq \sigma^{**}$, which concludes the proof. ■

Part (ii)

We know from Propositions 5 and 6, and from the properties of $a^*(\overline{\sigma})$, $a^{**}(\overline{\sigma})$, $a^{***}(\overline{\sigma})$ and $a^{****}(\overline{\sigma})$, that task distortion is given by:

- $a^{**}(\overline{\sigma}) - a^*(\overline{\sigma})$ for $\overline{\sigma} \in [0, \overline{\sigma}_1]$, which is increasing in $\overline{\sigma}$,
- $a^{***}(\overline{\sigma}) - a^{FB}$ for $\overline{\sigma} \in [\overline{\sigma}_1, \overline{\sigma}_2]$, which is increasing in $\overline{\sigma}$,
- $a^{****}(\overline{\sigma}) - a^{FB}$ for $\overline{\sigma} \in [\overline{\sigma}_2, \sigma^{**}]$, which is decreasing in $\overline{\sigma}$,
- Zero for $\overline{\sigma} \geq \sigma^{**}$, and thus constant in $\overline{\sigma}$. ■

Proof of Organizational Implication 1

Propositions 4, 5 and 6 imply that the optimal policy is transparency when $\alpha^l < 1$, and privacy when $\alpha^l \geq 1$. Therefore, it suffices to show that under transparency Δ^F is lower, and Δ^I higher, than under privacy. To this end, note that, from Proposition 3, it immediately follows that under transparency, formal and informal pay compression coincide and are equal to $\Delta^F_{\overline{\sigma}\in[0,\sigma^*]} = \Delta^I_{\overline{\sigma}\in[0,\sigma^*]} = \Delta_{\overline{\sigma}\in[0,\sigma^*]} = \frac{\alpha^d[\underline{u}_1 - \underline{u}_2] - \overline{\sigma}y(a^*(\overline{\sigma}))}{1+\alpha^d} > 0$ for $\overline{\sigma} \in [0, \sigma^*]$ and $\Delta^F_{\overline{\sigma}\in[\sigma^*,1]} = \Delta^I_{\overline{\sigma}\in[\sigma^*,1]} = \Delta_{\overline{\sigma}\in[\sigma^*,1]} = 0$ for $\overline{\sigma} \in [\sigma^*, 1]$. At the same time, Propositions 5 and 6 imply that under privacy, formal pay compression is $\Delta^F_{\overline{\sigma}\in[0,\overline{\sigma}_1]} \equiv \frac{\alpha^d[\underline{u}_1 - \underline{u}_2] + \overline{\sigma}[y(a^{**}(\overline{\sigma})) - y(a^*(\overline{\sigma}))]}{1+\alpha^d}$ for $\overline{\sigma} \in [0, \overline{\sigma}_1]$, $\Delta^F_{\overline{\sigma}\in[\overline{\sigma}_1,\overline{\sigma}_2]} \equiv \overline{\sigma}y\left(a^{***}(\overline{\sigma})\right)$ for $\overline{\sigma} \in [\overline{\sigma}_1, \overline{\sigma}_2]$, and $\Delta^F_{\overline{\sigma}\in[\overline{\sigma}_2,1]} \equiv \underline{u}_1 - \underline{u}_2$ for $\overline{\sigma} \in [\overline{\sigma}_2, 1]$. Finally, Propositions 5 and 6 imply that under privacy, informal pay compression is $\Delta^I_{\overline{\sigma}\in[0,\overline{\sigma}_1]} \equiv \frac{\alpha^d[\underline{u}_1 - \underline{u}_2 - \overline{\sigma}y(a^{**}(\overline{\sigma}))] - \overline{\sigma}y(a^*(\overline{\sigma}))}{1+\alpha^d}$ for $\overline{\sigma} \in [0, \overline{\sigma}_1]$, and $\Delta^I_{\overline{\sigma}\in[\overline{\sigma}_1,1]} \equiv 0$ for $\overline{\sigma} \in [\overline{\sigma}_1, 1]$.

To prove that Δ^F is lower under transparency than under privacy, we can use the facts that $a^{**}(\overline{\sigma}) > a^{FB}$ (from eq. (12) in the main text), that $y(\bullet)$ is non-negative and strictly increasing (by assumption) and that $\overline{\sigma}\left[y(a^{FB}) + \alpha^d y\left(a^{***}(\overline{\sigma})\right)\right] \geq \alpha^d\left[\underline{u}_1 - \underline{u}_2\right]$(by step 5 in the proof of Proposition 6), to verify that $min\left\{\Delta^F_{\overline{\sigma}\in[0,\overline{\sigma}_1]}, \Delta^F_{\overline{\sigma}\in[\overline{\sigma}_1,\overline{\sigma}_2]}, \Delta^F_{\overline{\sigma}\in[\overline{\sigma}_2,1]}\right\} > max\left\{\Delta_{\overline{\sigma}\in[0,\sigma^*]}, \Delta_{\overline{\sigma}\in[\sigma^*,1]}\right\}$, as desired.

To prove that Δ^I is higher under transparency than under privacy, there are two relevant cases to consider. If $\overline{\sigma} \leq \sigma^*$, we can use the fact that $\overline{\sigma}y\left(a^{**}(\overline{\sigma})\right) > 0$ to easily show that $\Delta_{\overline{\sigma}\in[0,\sigma^*]} > max\left\{\Delta^I_{\overline{\sigma}\in[0,\overline{\sigma}_1]}, \Delta^I_{\overline{\sigma}\in[\overline{\sigma}_1,1]}\right\}$, proving the result. Alternatively, if $\overline{\sigma} \geq \sigma^*$, our previous discussion implies that the value of Δ^I under transparency equals $\Delta_{\overline{\sigma}\in[\sigma^*,1]} = 0$. Because $\Delta^I_{\overline{\sigma}\in[\overline{\sigma}_1,1]} = 0$ for $\overline{\sigma} \in [\overline{\sigma}_1, 1]$, it suffices to prove that $\overline{\sigma} \geq \sigma^*$ implies $\overline{\sigma} \geq \overline{\sigma}_1$, which, in turn, will follow if we can prove that $\sigma^* > \overline{\sigma}_1$. To this end, define

$\Omega(\overline{\sigma}) \equiv \left[1 - \frac{\alpha^d}{1+\alpha^d}\overline{\sigma}\right] y\left(a^*(\overline{\sigma})\right) - C\left(a^*(\overline{\sigma})\right) - \frac{\alpha^d}{1+\alpha^d}\left[\underline{u}_1 - \underline{u}_2\right]$, where

$\Upsilon \equiv y(a^{FB}) - C(a^{FB}) - \alpha^d\left[\underline{u}_1 - \underline{u}_2\right]$,

$\Sigma(\overline{\sigma}) \equiv \left[1 + \frac{\alpha^d}{1+\alpha^d}\overline{\sigma}\right] y\left(a^{**}(\overline{\sigma})\right) - C\left(a^{**}(\overline{\sigma})\right)$, and

$\zeta(\overline{\sigma}) \equiv [1 + \alpha^d\overline{\sigma}] y\left(a^{***}(\overline{\sigma})\right) - C\left(a^{***}(\overline{\sigma})\right)$.

We know from the proof of Proposition 6 that $\overline{\sigma}_1$ is implicitly defined by $\Omega(\overline{\sigma}_1) - \Upsilon + \Sigma(\overline{\sigma}_1) - \zeta(\overline{\sigma}_1) = 0$. Using Equations (12) and (13), it is possible to verify that $\Sigma(\overline{\sigma}) < \zeta(\overline{\sigma})$ for all $\overline{\sigma}$, so that $\Omega(\overline{\sigma}_1) - \Upsilon > 0$. But we also know from the proof of Proposition 3 that σ^* is implicitly defined by $\Omega(\sigma^*) - \Upsilon = 0$. The result that $\sigma^* > \overline{\sigma}_1$ then follows because $\Omega'(\overline{\sigma}) < 0$, as one can easily check. ■

Appendix II: Imperfect Privacy

We have so far assumed that by informally contracting with the agents, the principal can maintain their actual compensation levels perfectly private throughout the employment relationship. This is a restrictive assumption because depending on the firm's institutional setting and social norms, even if privately agreed with the principal, there is some chance that the compensation of agent 1 may "leak" to agent 2, and vice versa. For instance, the agents may accidentally exchange comments or take actions, such as driving to work on cars with very different values, that could potentially reveal their pay gap. Alternatively, exogenous changes in the law or in the organization's culture may force the principal to suddenly make both formal and actual compensation levels public.

In this section we relax the "perfect privacy" assumption by allowing for potential leakage of informal compensation agreements. In particular, we continue to assume that the agents observe each other's formal compensation with certainty, but now we also assume that they may observe each other's informal compensation with probability $1-\theta$, so that $\theta \in [0,1]$ measures the privacy of informal agreements. Naturally, we also assume that when the agents observe each other's informal compensation, social comparisons are driven by the gap in informal (and hence actual) pay levels, rather than by the gap in formal pay levels. Accordingly, an agent's frustration from social comparisons when informal pay levels are leaked, and when they are not, are respectively given by:

$A_i^I \equiv max\{0, w_j^I - C(a_j) - w_i^I + C(a_i)\}$, and

$A_i^F \equiv max\{0, w_j^F - C(a_j) - w_i^F + C(a_i)\}$.

We show below that the key insight from the model with perfect privacy ($\theta = 1$)—namely, that the principal may combine formal and informal compensation to optimally manage social comparisons—continues to hold under imperfect privacy ($0 < \theta < 1$). Through this Appendix we assume that the agents are sensitive to cheating by the principal (that is, $\alpha^l \geq 1$), for otherwise transparency would be optimal and the results would be identical to those of the baseline model. To keep the analysis simple, we also assume that the principal and the agents are risk-neutral.

Before proceeding, it is useful to define an agent's frustration in the case where informal pay levels leak but the principal (or the agent) deviates by paying (or demanding) the formal salary:

$$A_i^{dev} \equiv max\{0, w_j^I - C(a_j) - w_i^F + C(a_i)\}.$$

Since deviations do not occur in equilibrium, under imperfect privacy the expected utilities of the principal and agent i are given, respectively, by:

$$u_P \equiv \sum_i [y(a_i) - w_i^I] - \theta \sum_i min\left\{\overline{\sigma}, \frac{\alpha^d A_i^F}{y(a_i)}\right\} y(a_i) - (1-\theta) \sum_i min\left\{\overline{\sigma}, \frac{\alpha^d A_i^I}{y(a_i)}\right\} y(a_i), \text{ and}$$

$$u_i \equiv w_i^I - C(a_i) - \theta\left[\alpha^d A_i^F - min\left\{\overline{\sigma}, \frac{\alpha^d A_i^F}{y(a_i)}\right\} y(a_i)\right] - (1-\theta)\left[\alpha^d A_i^I - min\left\{\overline{\sigma}, \frac{\alpha^d A_i^I}{y(a_i)}\right\} y(a_i)\right].$$

The principal's problem can then be written as follows:

$$\max_{a_1, a_2, w_1^F, w_2^F, w_1^I, w_2^I} \{u_P\}, \text{ subject to}$$

$$u_i \geq \underline{u}_i, \text{ for all } i. \qquad (PC_i)$$

$$w_i^I - w_i^F \geq (1-\theta)\left[\alpha^d A_i^I - min\left\{\overline{\sigma}, \frac{\alpha^d A_i^I}{y(a_i)}\right\} y(a_i)\right] - (1-\theta)\left[\alpha^d A_i^{dev} - min\left\{\overline{\sigma}, \frac{\alpha^d A_i^{dev}}{y(a_i)}\right\} y(a_i)\right], \text{ for all } i. \qquad (IC_i)$$

$$w_i^I - w_i^F \leq \theta\left[min\left\{\overline{\sigma}, \frac{\alpha^d A_i^F + \alpha^l (w_i^I - w_i^F)}{y(a_i)}\right\} - min\left\{\overline{\sigma}, \frac{\alpha^d A_i^F}{y(a_i)}\right\}\right] y(a_i) + (1-\theta)\left[min\left\{\overline{\sigma}, \frac{\alpha^d A_i^{dev} + \alpha^l (w_i^I - w_i^F)}{y(a_i)}\right\} - min\left\{\overline{\sigma}, \frac{\alpha^d A_i^I}{y(a_i)}\right\}\right] y(a_i) \text{ for all } i. \qquad (IC_{Pi})$$

Moreover, note that (IC_i) can be written as $g(w_i^I) \geq g(w_i^F)$, where

$$g(w) = w - (1-\theta)\left[\alpha^d A_i^g - min\left\{\overline{\sigma}, \frac{\alpha^d A_i^g}{y(a_i)}\right\} y(a_i)\right] \text{ and}$$

$$A_i^g = max\{0, w_j^I - C(a_j) - w + C(a_i)\}.$$

Since $g'(w) > 0$, it follows that $g(w_i^I) \geq g(w_i^F)$ if and only if $w_i^I \geq w_i^F$, so we can rewrite (IC_i) as in the baseline model:

$w_i^I \geq w_i^F$, for all i. (IC_i)

Optimal privacy contract with high retaliation capability

We begin by characterizing the optimal (imperfect) privacy policy when the agents have high retaliation capability (that is, when $\overline{\sigma}$ is large). The analysis is divided in two steps. First, we study the agents' optimal tasks and informal compensation levels ignoring the incentive compatibility constraints, in which case it is optimal for the principal to set formal compensation so as to minimize social comparison costs (i.e., $A_1^F = A_2^F = 0$). Second, and consistent with Proposition 5 in the baseline model, we prove that the incentive compatibility constraints will be satisfied so long as $\overline{\sigma}$ is large enough, and therefore that the agents' tasks and informal compensation levels identified in part one are indeed optimal. In addition, we characterize the formal compensation levels that allow the principal to implement the optimal contract.

Step 1: Optimal contract ignoring the incentive compatibility constraints

As mentioned in the previous paragraph, without the incentive compatibility constraints, it is optimal for the principal to set formal compensation so as to minimize social comparison costs (that is, so that $A_1^F = A_2^F = 0$).[13] Accordingly, the principal's problem can be written as follows:

$$\max_{a_1,a_2,w_1^I,w_2^I} \left\{ u_P \equiv \sum_i [y(a_i) - w_i^I] - (1-\theta) \sum_i min\left\{\overline{\sigma}, \frac{\alpha^d A_i^I}{y(a_i)}\right\} y(a_i) \right\}, \text{ subject to}$$

$$u_i \equiv w_i^I - C(a_i) - (1-\theta)\left[\alpha^d A_i^I - min\left\{\overline{\sigma}, \frac{\alpha^d A_i^I}{y(a_i)}\right\} y(a_i)\right] \geq \underline{u}_i, \text{ for all } i.^{14} \quad (PC_i)$$

[13] If one of the agents were suffering from social comparisons as a result of formal pay differences, the principal could decrease the formal pay of the other agent. This change would not violate the agents' participation constraints and would increase the principal's utility, either directly by reducing retaliation by the agent who was originally suffering from social comparisons or indirectly by allowing the principal to pay a lower informal compensation to the agent.

[14] Formally, the principal must also choose w_1^F and w_2^F subject to $w_1^F - C(a_1) = w_2^F - C(a_2)$ (which guarantees that $A_1^F = A_2^F = 0$). However, since, conditional on $A_1^F = A_2^F = 0$, formal payments do not affect the principal's nor the agents' utilities, and there are no restrictions on the values that they can take (except for the aforementioned constraint), in this step we can ignore both the selection of w_1^F and w_2^F and the constraint $w_1^F - C(a_1) = w_2^F - C(a_2)$.

Applying the same arguments as in the proofs of Lemmas 2 and 3 and Proposition 2, we can show that for any given θ: *(a)* constraint (PC_i) binds for all i, *(b)* agent 1 does not suffer from social comparisons even in the event of leakage ($A_1^I = 0$), and *(c)* there is no distortion in the task ($a_1 = a^{FB}$) or informal compensation ($w_1^I = \underline{u}_1 + C(a^{FB})$) of agent 1. Using these results, the principal's problem can be simplified to choosing the task and informal compensation of agent 2 subject to the latter's binding participation constraint:

$$\max_{a_2, w_2^I} \left\{ u_P \equiv y(a_2) - w_2^I - (1-\theta) min \left\{ \overline{\sigma}, \frac{\alpha^d [\underline{u}_1 + C(a_2) - w_2^I]}{y(a_2)} \right\} y(a_2) \right\}, \text{ subject to}$$

$$w_2^I = \underline{u}_2 + C(a_2) + (1-\theta) \left[\alpha^d \left[\underline{u}_1 + C(a_2) - w_2^I \right] - min \left\{ \overline{\sigma}, \frac{\alpha^d [\underline{u}_1 + C(a_2) - w_2^I]}{y(a_2)} \right\} y(a_2) \right] . \quad (PC_2)$$

Let $a^*(\theta, \overline{\sigma})$ be implicitly defined by:

$$\left[1 - \frac{(1-\theta)^2 \alpha^d}{1+(1-\theta)\alpha^d} \overline{\sigma} \right] y'\left(a^*(\theta, \overline{\sigma})\right) - C'\left(a^*(\theta, \overline{\sigma})\right) = 0. \quad (A6)$$

Note that $a^*(\theta, \overline{\sigma})$ increases in θ and decreases in $\overline{\sigma}$, with $a^*(0, \overline{\sigma}) = a^*(\overline{\sigma})$ and $a^*(1, \overline{\sigma}) = a^{FB}$.

Using (PC_2) and solving for the optimal task, the principal's utility when $\overline{\sigma} \geq \frac{\alpha^d [\underline{u}_1 + C(a_2) - w_2^I]}{y(a_2)}$ can be written as:

$$u_P^1(\theta) \equiv y(a^{FB}) - C(a^{FB}) - \underline{u}_2 - (1-\theta)\alpha^d \left[\underline{u}_1 - \underline{u}_2 \right].$$

Similarly, using (PC_2) and $(A6)$ and solving for the optimal task, the principal's utility when $\overline{\sigma} \leq \frac{\alpha^d [\underline{u}_1 + C(a_2) - w_2^I]}{y(a_2)}$ can be written as:

$$u_P^2(\theta, \overline{\sigma}) \equiv \left[1 - \frac{(1-\theta)^2 \alpha^d}{1+(1-\theta)\alpha^d} \overline{\sigma} \right] y\left(a^*(\theta, \overline{\sigma})\right) - C\left(a^*(\theta, \overline{\sigma})\right) - \underline{u}_2 - \frac{(1-\theta)\alpha^d [\underline{u}_1 - \underline{u}_2]}{1+(1-\theta)\alpha^d}.$$

Finally, let $\sigma^*(\theta)$ be implicitly defined by:

$$u_P^1(\theta) = u_P^2\left(\theta, \sigma^*(\theta)\right).$$

Then, we can prove the following result, which closely mirrors Proposition 3 in the baseline model.

Lemma A1. *Assume that the incentive compatibility constraints, (IC_i) and (IC_{Pi}), are satisfied. Then, the principal's optimal policy towards agent 2 has the following features:*

(i) If $\overline{\sigma} < \sigma^*(\theta)$, the task of agent 2 is distorted downwards and his compensation is compressed upwards, with both distortions decreasing in θ and tending to zero as $\theta \to 1$: $a_2 = a^*(\theta, \overline{\sigma})$, and $w_2^I = \underline{u}_2 + C\left(a^*(\theta, \overline{\sigma})\right) + \frac{(1-\theta)\{\alpha^d[\underline{u}_1 - \underline{u}_2] - \overline{\sigma}y(a^*(\theta,\overline{\sigma}))\}}{1+(1-\theta)\alpha^d}$.

(ii) If $\overline{\sigma} > \sigma^*(\theta)$, there are no task distortion or pay compression: $a_2 = a^{FB}$, and $w_2^I = \underline{u}_2 + C(a^{FB})$.

(iii)If $\overline{\sigma} = \sigma^*(\theta)$, then the principal is indifferent between the policies described in (i) and (ii) above.

Proof. The proofs of parts (i), (ii) and (ii) are almost identical to those in Proposition 3, and are therefore omitted. The result that the distortions in (i) decrease in θ and tend to zero as $\theta \to 1$ follows directly from the previous observation that $a^*(\theta, \overline{\sigma})$ increases in θ and the fact that $\alpha^d[\underline{u}_1 - \underline{u}_2] - \overline{\sigma}y\left(a^*(\theta, \overline{\sigma})\right) > 0$ for $\overline{\sigma} < \sigma^*(\theta)$. ■

Lemma A1 is quite intuitive, and implies that even if the principal can eliminate the social comparisons driven by differences in formal pay, the possibility of leakage creates a residual distortion in the task and informal compensation of agent 2. This distortion tends to disappear as the probability of leakage tends to zero (that is, as θ tends to one), such that the optimal contract under imperfect privacy described by Lemma A1 tends to the one under perfect privacy, as described by Proposition 5.

Step 2: Implementing the optimal contract with minimum social comparisons

The (imperfect) privacy policy described in Lemma A1 is optimal conditional on the incentive compatibility constraints being satisfied. We now provide conditions under which constraints (IC_i) and (IC_{Pi}) are both satisfied, and therefore under which the proposed policy is

indeed optimal. To obtain these conditions we must distinguish between two cases: $\overline{\sigma} > \sigma^*(\theta)$ and $\overline{\sigma} < \sigma^*(\theta)$.

Case 1: $\boldsymbol{\overline{\sigma} > \sigma^*(\theta)}$

In this case, the optimal policy from Lemma A1 entails $a_1 = a_2 = a^{FB}$ and $w_1^I = \underline{u}_1 + C(a^{FB}) > \underline{u}_2 + C(a^{FB}) = w_2^I$. Therefore, implementing $A_1^F = A_2^F = 0$ requires that the two agents' formal salaries be equal: $w_1^F = w_2^F$. The equal formal salaries that relax the (IC_{Pi}) constraints as much as possible without violating the (IC_i) constraints are given by $w_1^F = w_2^F = w_2^I$. Substituting these payments into the incentive constraint (IC_{P1}), using the fact that they imply that $A_1^{dev} = A_1^I$, and using the fact that $A_1^I = 0$ by an argument similar to that of Lemma 2, we obtain that the optimal policy is self-enforcing as long as $\overline{\sigma} \geq \sigma^{**} \equiv \frac{\underline{u}_1 - \underline{u}_1}{y(a^{FB})}$, irrespective of θ. This result mirrors Proposition 5, except that now the optimal privacy policy does not achieve the first-best for $\theta < 1$, because agent 2 experiences social comparisons and hence retaliates with positive probability.

Case 2: $\boldsymbol{\overline{\sigma} < \sigma^*(\theta)}$

In this case, the optimal policy from Lemma A1 entails $a_1 = a^{FB}$, $a_2 = a^*(\theta, \overline{\sigma})$, and $w_1^I = \underline{u}_1 + C(a^{FB}) > \underline{u}_2 + C\big(a^*(\theta, \overline{\sigma})\big) + \frac{(1-\theta)\{\alpha^d[\underline{u}_1 - \underline{u}_1] - \overline{\sigma} y(a^*(\theta,\overline{\sigma}))\}}{1+(1-\theta)\alpha^d} = w_2^I$. Thus, implementing $A_1^F = A_2^F = 0$ requires that the formal salary of agent 1 be higher than that of agent 2:

$$w_1^F = w_2^F + C(a^{FB}) - C\big(a^*(\theta, \overline{\sigma})\big). \qquad (A7)$$

Given $(A7)$, the formal salaries that relax the (IC_{Pi}) constraints as much as possible without violating the (IC_i) constraints are given by:

$w_2^F = w_2^I$, and

$w_1^F = w_2^I + C(a^{FB}) - C\big(a^*(\theta, \overline{\sigma})\big)$.

Substituting these payments into the incentive constraint (IC_{P1}), using the fact that they imply that $A_1^{dev} = 0$, and using the fact that $A_1^I = 0$ by an argument similar to that of Lemma 2, we obtain that the optimal policy is self-enforcing as long as $\overline{\sigma}$ satisfies:

$$\overline{\sigma} \geq \frac{\sigma^{**}}{1+(1-\theta)\alpha^d} + \frac{(1-\theta)\overline{\sigma}y\left(a^*(\theta,\overline{\sigma})\right)}{\left[1+(1-\theta)\alpha^d\right]y(a^{FB})}. \qquad (A8)$$

Defining $\sigma^{**}(\theta)$ as the smallest value of $\overline{\sigma}$ such that $(A8)$ holds with strict equality, it is easy to verify that $\overline{\sigma}$ satisfies $(A8)$ if and only if $\overline{\sigma} \geq \sigma^{**}(\theta)$. Moreover, given our initial assumption that $\overline{\sigma} < \sigma^*$, it must be that $\overline{\sigma}y\left(a^*(\theta,\overline{\sigma})\right) < \alpha^d\left[\underline{u}_1 - \underline{u}_1\right]$, which, after a few calculations, implies that $\sigma^{**}(\theta) \leq \sigma^{**}$, with the inequality holding strictly for $\theta < 1$.

We can now summarize steps 1 and 2 above in the following result.

Proposition A2. *Assume that the agents are sensitive to cheating by the principal (i.e., $\alpha^l \geq 1$). Then there are values $\sigma^*(\theta)$ and $\sigma^{**}(\theta)$, with $\sigma^{**}(\theta) \leq \sigma^{**} \equiv \frac{\underline{u}_1 - \underline{u}_1}{y(a^{FB})} \leq \sigma^*(\theta)$, such that the optimal policy under imperfect privacy has the following features:*

(i) If $\overline{\sigma} > \sigma^*(\theta)$, then:

a) Each agent is assigned the first-best task and receives an informal pay equal to his outside option plus the cost of performance: $a_i = a^{FB}$ and $w_i^I = \underline{u}_i + C(a^{FB})$ for $i = 1,2$.

b) The two agents receive the same formal pay, which equals the informal pay of the agent with low outside option (agent 2): $w_1^F = w_2^F = w_2^I$.

(ii) If $\overline{\sigma} \in \left[\sigma^{**}(\theta), \sigma^*(\theta)\right)$, then:

a) There is no distortion in the task ($a_1 = a^{FB}$) or informal compensation ($w_1^I = \underline{u}_1 + C(a^{FB})$) of agent 1.

b) The task of agent 2 is distorted downwards, and his informal compensation is compressed upwards, with both distortions decreasing in θ and tending to zero as $\theta \to 1$: $a_2 = a^*(\theta,\overline{\sigma})$, and $w_2^I = \underline{u}_2 + C\left(a^*(\theta,\overline{\sigma})\right) + \frac{(1-\theta)\{\alpha^d[\underline{u}_1 - \underline{u}_2] - \overline{\sigma}y(a^*(\theta,\overline{\sigma}))\}}{1+(1-\theta)\alpha^d}$.

c) The formal pay of agent 2 is equal to his informal pay, whereas the formal pay of agent 1 equals $w_1^F = \underline{u}_2 + C(a^{FB}) + \frac{(1-\theta)\{\alpha^d[\underline{u}_1 - \underline{u}_2] - \overline{\sigma}y(a^*(\theta,\overline{\sigma}))\}}{1+(1-\theta)\alpha^d}$.

(i) If $\overline{\sigma} = \sigma^*(\theta)$, then the principal is indifferent between the policies described in (i) and (ii) above.

Proof. Parts (i), (ii) and (iii), as well as the result that $\sigma^{**}(\theta) \leq \sigma^{**} \equiv \frac{\underline{u}_1 - \underline{u}_1}{y(a^{FB})}$, all follow directly from steps 1 and 2 above. To prove that $\frac{\underline{u}_1 - \underline{u}_1}{y(a^{FB})} \leq \sigma^*(\theta)$, note that by Lemma A1 we know that for any $\overline{\sigma} \geq \sigma^*(\theta)$: (a) it is optimal to set $a_2 = a^{FB}$ and $w_2^I = \underline{u}_2 + C(a^{FB})$, and that (b) such values satisfy $\overline{\sigma} \geq \frac{\alpha^d[\underline{u}_1 + C(a_2) - w_2^I]}{y(a_2)}$. Accordingly, we know that $\sigma^*(\theta) \geq \frac{\alpha^d[\underline{u}_1 - \underline{u}_2]}{y(a^{FB})}$, as desired. ■

These results on imperfect privacy deviate from those on perfect privacy in two interesting ways. First, while the principal can still use formal contracts to minimize social comparisons while paying the agents optimally, he may compress formal pay levels less than under perfect privacy. This occurs because the possibility of informal pay leakage forces the principal to give a lighter task to agent 2, so setting the formal pay levels equal would cause agent 1 to suffer from unnecessary social comparisons in the event of no leakage.

Second, and related, while the possibility of leakage under imperfect privacy reduces the principal's profit, it may facilitate the use of formal pay to manage social comparisons ($\sigma^{**}(\theta) \leq \sigma^{**}$). This occurs because compared to the case of perfect privacy, the principal now needs a lower gap between the informal and formal pay of agent 1 to minimize social comparisons (that is, to ensure $A_1^F = A_2^F = 0$), as shown by condition $(A7)$. Therefore, the principal's temptation to deviate on the informal salaries is smaller than in the case of pure privacy.

Optimal privacy contract with limited retaliation capacity

We conclude our analysis by characterizing the optimal (imperfect) privacy policy when the agents have limited retaliation capability (that is, when $\overline{\sigma} < \sigma^{**}(\theta)$).

Applying the same arguments as in the proofs of Lemmas 3 and 4, we can show that for any given θ: *(a)* constraint (PC_i) binds for all i, and *(b)* agent 1 does not suffer from social comparisons regardless of whether a leakage occurs ($A_1^F = A_1^I = 0$). Following the proofs of

Lemmas 5 and 6, we can also show that: *(c)* $w_2^I = w_2^F = w_2$—which allows us to ignore constraints (IC_2) and (IC_{P2}) —and *(d)* the incentive constraint (IC_{P1}) is binding and takes the form $w_1^I - w_1^F = \overline{\sigma} y(a_1)$—which allows us to ignore constraint (IC_1).

Using results *(a)* to *(d)*, the principal's problem can be written as follows:

$$\max_{a_1,a_2,w_1^I,w_1^F,w_2} \sum_i \left[y(a_i) - w_i^I - \theta min\left\{\overline{\sigma}, \frac{\alpha^d A_i^F}{y(a_i)}\right\} y(a_i) - (1-\theta) min\left\{\overline{\sigma}, \frac{\alpha^d A_i^I}{y(a_i)}\right\} y(a_i)\right],$$

subject to

$$w_1^I - C(a_1) = \underline{u}_1, \qquad (PC_1)$$

$$w_2 - C(a_2) - \theta\left[\alpha^d A_2^F - min\left\{\overline{\sigma}, \frac{\alpha^d A_2^F}{y(a_2)}\right\} y(a_2)\right] - (1-\theta)\left[\alpha^d A_2^I - min\left\{\overline{\sigma}, \frac{\alpha^d A_2^I}{y(a_2)}\right\} y(a_2)\right] = \underline{u}_2, \qquad (PC_2)$$

$$w_1^I - w_1^F = \overline{\sigma} y(a_1), \qquad (IC_{P1})$$

$$\underline{u}_1 - \overline{\sigma} y(a_1) - w_2 + C(a_2) \geq 0, \text{ and} \qquad (B1)$$

$$\underline{u}_1 - w_2 + C(a_2) \geq 0, \qquad (B2)$$

where constraints $(B1)$ and $(B2)$ have been added to ensure that agent 1 does not suffer from social comparisons regardless of whether a leakage occurs.

To characterize the solution to the principal's problem, define the following critical tasks:

$a^{*1}(\overline{\sigma}, \theta) > a^{FB}$, such that $\left[1 + \theta \frac{\alpha^d}{1+\alpha^d}\overline{\sigma}\right] y'\left(a^{*1}(\overline{\sigma}, \theta)\right) - C'\left(a^{*1}(\overline{\sigma}, \theta)\right) = 0$,

$a^{*2}(\overline{\sigma}, \theta) < a^{FB}$, such that $\left[1 - \frac{\alpha^d}{1+\alpha^d}\overline{\sigma}\right] y'\left(a^{*2}(\overline{\sigma}, \theta)\right) - C'\left(a^{*2}(\overline{\sigma}, \theta)\right) = 0$,

$a^{*3}(\overline{\sigma}, \theta) > a^{FB}$, such that $[1 + \theta\alpha^d\overline{\sigma}] y'\left(a^{*3}(\overline{\sigma}, \theta)\right) - C'\left(a^{*3}(\overline{\sigma}, \theta)\right) = 0$,

$a^{*4}(\overline{\sigma}, \theta) < a^{FB}$, such that $\left[1 - \frac{(1-\theta)\alpha^d}{1+(1-\theta)\alpha^d}\overline{\sigma}\right] y'\left(a^{*4}(\overline{\sigma}, \theta)\right) - C'\left(a^{*4}(\overline{\sigma}, \theta)\right) = 0$, and

$a^{*5}(\overline{\sigma}, \theta) > a^{FB}$, such that $\underline{u}_1 - \underline{u}_2 = \overline{\sigma} y\left(a^{*5}(\overline{\sigma}, \theta)\right)$.

We have the following result, which closely mirrors Proposition 6 in the baseline model.

Proposition A3. *Suppose the agents are sensitive to cheating by the principal ($\alpha^l \geq 1$). Then, there are values $\overline{\sigma}_1(\theta)$, $\overline{\sigma}_2(\theta)$ and $\overline{\sigma}_3(\theta)$, with $0 < \overline{\sigma}_1(\theta) \leq \overline{\sigma}_2(\theta) \leq \overline{\sigma}_3(\theta) < \sigma^{**}(\theta)$, such that the optimal (imperfect) privacy policy has the following features:*

(i) If $\overline{\sigma} \in [0, \overline{\sigma}_1(\theta)]$:

a) The informal pay of agent 1 is not distorted: $w_1^I = \underline{u}_1 + C(a_1)$.

b) The informal pay of agent 2 is equal to the formal pay and is distorted: $w_2^I = w_2^F = \underline{u}_2 + C(a_2) + \frac{\alpha^d[\underline{u}_1 - \underline{u}_2 - \theta\overline{\sigma}y(a_1)] - \overline{\sigma}y(a_2)}{1+\alpha^d}$.

c) The formal pay of agent 1 differs from that of agent 2: $w_1^F = \underline{u}_1 + C(a_1) - \overline{\sigma}y(a_1)$.

d) The task of agent 1 is distorted upwards: $a_1 = a^{*1}(\overline{\sigma}, \theta) > a^{FB}$.

e) The task of agent 2 is distorted downwards: $a_2 = a^{*2}(\overline{\sigma}, \theta) < a^{FB}$.

(ii) If $\overline{\sigma} \in [\overline{\sigma}_1(\theta), \overline{\sigma}_2(\theta)]$:

a) The informal pay of agent 1 is not distorted: $w_1^I = \underline{u}_1 + C(a_1)$.

b) The informal pay of agent 2 is equal to the formal pay and is distorted: $w_2^I = w_2^F = \underline{u}_2 + C(a_2) + \frac{(1-\theta)[\alpha^d(\underline{u}_1 - \underline{u}_2) - \overline{\sigma}y(a_2)]}{1+(1-\theta)\alpha^d}$.

c) The formal pay of agent 1 differs from that of agent 2: $w_1^F = \underline{u}_1 + C(a_1) - \overline{\sigma}y(a_1)$.

d) The task of agent 1 is distorted upwards: $a_1 = a^{*3}(\overline{\sigma}, \theta) > a^{FB}$.

e) The task of agent 2 is distorted downwards: $a_2 = a^{*4}(\overline{\sigma}, \theta) < a^{FB}$.

(iii)If $\overline{\sigma} \in [\overline{\sigma}_2(\theta), \overline{\sigma}_3(\theta)]$:

a) The informal pay of agent 1 is not distorted: $w_1^I = \underline{u}_1 + C(a_1)$.

b) The informal pay of agent 2 is equal to the formal pay and is not distorted: $w_2^I = w_2^F = \underline{u}_2 + C(a_2)$.

c) The formal pay of agent 1 differs from that of agent 2: $w_1^F = \underline{u}_1 + C(a_1) - \overline{\sigma} y(a_1)$.

d) The task of agent 1 is distorted upwards: $a_1 = a^{*3}(\overline{\sigma}, \theta) > a^{FB}$.

e) The task of agent 2 is not distorted: $a_2 = a^{FB}$.

(iv) If $\overline{\sigma} \in [\overline{\sigma}_3(\theta), \sigma^{**}(\theta))$:

a) The informal pay of agent 1 is not distorted: $w_1^I = \underline{u}_1 + C(a_1)$.

b) The informal pay of agent 2 is equal to the formal pay and is not distorted: $w_2^I = w_2^F = \underline{u}_2 + C(a_2)$.

c) The formal pay of agent 1 differs from that of agent 2: $w_1^F = \underline{u}_2 + C(a_1)$.

d) The task of agent 1 is distorted upwards: $a_1 = a^{*5}(\overline{\sigma}, \theta) > a^{FB}$.

e) The task of agent 2 is not distorted: $a_2 = a^{FB}$.

Proof. Note first that we have 3 equations that hold as equality and 5 unknowns. Accordingly, to fully characterize the optimal contract it suffices to find values for a_1 and a_2. Note also that $(B1)$ implies $(B2)$, so hereafter we can ignore the latter.

Consider the following three cases:

Case 1: $\overline{\sigma} \leq \frac{\alpha^d[\underline{u}_1 - \overline{\sigma} y(a_1) - w_2 + C(a_2)]}{y(a_2)}$.

Using (PC_1), (PC_2) and (IC_{P1}), we can verify that tasks $a^{*1}(\overline{\sigma}, \theta)$ and $a^{*2}(\overline{\sigma}, \theta)$ satisfy $(B1)$ and maximize the objective function. The principal's utility can then be written as:

$$u_P^{*1}(\theta, \overline{\sigma}) \equiv \left[1 + \theta \frac{\alpha^d}{1+\alpha^d} \overline{\sigma}\right] y\left(a^{*1}(\overline{\sigma}, \theta)\right) - \underline{u}_1 - C\left(a^{*1}(\overline{\sigma}, \theta)\right) + \left[1 - \frac{\alpha^d}{1+\alpha^d} \overline{\sigma}\right] y\left(a^{*2}(\overline{\sigma}, \theta)\right) - \underline{u}_2 - C\left(a^{*2}(\overline{\sigma}, \theta)\right) - \frac{\alpha^d(\underline{u}_1 - \underline{u}_2)}{1+\alpha^d},$$

Case 2: $\frac{\alpha^d[\underline{u}_1-\overline{\sigma}y(a_1)-w_2+C(a_2)]}{y(a_2)} \leq \overline{\sigma} \leq \frac{\alpha^d[\underline{u}_1-w_2+C(a_2)]}{y(a_2)}$.

Using (PC_1), (PC_2) and (IC_{P1}), we can verify that tasks $a^{*3}(\overline{\sigma},\theta)$ and $a^{*4}(\overline{\sigma},\theta)$ satisfy $(B1)$ and maximize the objective function. The principal's utility can then be written as:

$$u_P^{*2}(\theta,\overline{\sigma}) \equiv (1+\theta\alpha^d\overline{\sigma})y\big(a^{*3}(\overline{\sigma},\theta)\big) + \Big[1-\tfrac{(1-\theta)\alpha^d}{1+(1-\theta)\alpha^d}\overline{\sigma}\Big]y\big(a^{*4}(\overline{\sigma},\theta)\big) - (1+\theta\alpha^d)\underline{u}_1 - C\big(a^{*3}(\overline{\sigma},\theta)\big) - (1-\theta\alpha^d)\underline{u}_2 - C\big(a^{*4}(\overline{\sigma},\theta)\big) - \Big[\tfrac{(1-\theta)(1-\theta\alpha^d)\alpha^d(\underline{u}_1-\underline{u}_2)}{1+(1-\theta)\alpha^d}\Big].$$

Case 3: $\overline{\sigma} \geq \frac{\alpha^d[\underline{u}_1-w_2+C(a_2)]}{y(a_2)}$.

Using (PC_1), (PC_2) and (IC_{P1}), we can verify that tasks $a^{*6}(\overline{\sigma},\theta) \equiv min\{a^{*3}(\overline{\sigma},\theta), a^{*5}(\overline{\sigma},\theta)\}$ and a^{FB} satisfy $(B1)$ and maximize the objective function. The principal's utility can then be written as:

$$u_P^{*3}(\theta,\overline{\sigma}) \equiv (1+\theta\alpha^d\overline{\sigma})y\big(a^{*6}(\overline{\sigma},\theta)\big) - \underline{u}_1 - C(a_1) + y(a^{FB}) - \underline{u}_2 - C(a^{FB}) - \alpha^d\big(\underline{u}_1 - \underline{u}_2\big).$$

Next, let the critical values $\overline{\sigma}_1(\theta)$, $\overline{\sigma}_2(\theta)$ and $\overline{\sigma}_3(\theta)$ be implicitly defined by:

$u_P^{*1}(\theta,\overline{\sigma}_1(\theta)) = u_P^{*2}(\theta,\overline{\sigma}_1(\theta))$,

$u_P^{*2}(\theta,\overline{\sigma}_2(\theta)) = u_P^{*3}(\theta,\overline{\sigma}_3(\theta))$, and

$\underline{u}_1 - \underline{u}_2 = \overline{\sigma}y\big(a^{*3}(\overline{\sigma},\theta)\big)$.

Using the same arguments as in steps 1 through 5 of the proof of Proposition 6, we can then prove the following:

- If $\overline{\sigma} \in [0,\overline{\sigma}_1(\theta)]$, then tasks $a^{*1}(\overline{\sigma},\theta)$ and $a^{*2}(\overline{\sigma},\theta)$ satisfy the condition for case 1 and are indeed optimal.
- If $\overline{\sigma} \in [\overline{\sigma}_1(\theta),\overline{\sigma}_2(\theta)]$, then tasks $a^{*3}(\overline{\sigma},\theta)$ and $a^{*4}(\overline{\sigma},\theta)$ satisfy the condition for case 2 and are indeed optimal.
- If $\overline{\sigma} \in [\overline{\sigma}_2(\theta),\overline{\sigma}_3(\theta)]$, then: (i) $a^{*3}(\overline{\sigma},\theta) \leq a^{*5}(\overline{\sigma},\theta)$, and (iii) tasks $a^{*3}(\overline{\sigma},\theta)$ and a^{FB} satisfy the condition for case 3 and are indeed optimal.

- If $\overline{\sigma} \in [\overline{\sigma}_3(\theta), \sigma^{**}(\theta)]$, then: (i) $a^{*3}(\overline{\sigma}, \theta) \geq a^{*5}(\overline{\sigma}, \theta)$, and (ii) tasks $a^{*5}(\overline{\sigma}, \theta)$ and a^{FB} satisfy the condition for case 3 and are indeed optimal.

Given the optimal tasks, we can then use constraints (PC_1), (PC_2) and (IC_{P1}), together with the result that $w_2^I = w_2^F = w_2$, to recover the optimal values of the agents' formal and informal compensations, which completes the proof. ■