Supervise me if you can. Relational feelings, incentive pays and supervisory violations

Stefano Dughera1,2 and Alain Marciano3

Received: 6 May 2021 / Accepted: 30 September 2021 / Published online: 22 October 2021 © The Author(s) 2021

Abstract
We study a bonus pay setting where a principal hires a supervisor to evaluate a group of potentially shirking workers. The supervisor and her workers develop relational feelings (either positive or negative) after interacting with each other. We analyze a novel class of organizational infractions where the supervisor provides false evaluations of the workers’ performance to accommodate her other-regarding preferences. We show that the workers’ reaction to these supervisory violations and their decision to provide effort depend on their relational concerns. The implication is that incentive pays can be ineffective due to the distortion of relational concerns.

Keywords Other-regarding preferences · Supervision · Incentive pay · Organizational infractions · Shirking

JEL Classification J41 · M54 · D64
1 Introduction

Motivating workers is a major problem in organizations, that can be solved by using incentive pays. The latter are studied under two forms in the literature. First, bonus payments, that correspond to a premium that is paid in case of good performance. Second, efficiency wages, that consist in a dismissal or wage reduction that comes along with bad performance (Shapiro and Stiglitz 1984). Both mechanisms have respective advantages and limits, that have been discussed in an abundant literature (see e.g., MacLeod and Malcomson 1989). They also share a common difficulty: workers’ performances may not be evaluated truthfully. Since the objective evaluation of workers’ output is in all likelihood impossible – for instance, because of multitasking concerns, measurement cost and flexibility issues – performances cannot but be evaluated subjectively by a supervisor (Bull 1987; Gibbons, 1998; Prendergast 1999; Baker et al. 1994; Levin 2003; Gibbs et al. 2004). Indeed, “thinking of any job in which subjective evaluation or supervisor discretion does not play some role in incentives is difficult” (Gibbs 2013: 411). The key implication is that supervisors may have reasons to be dishonest and provide false evaluations of their supervisees’ performances. This occurs, it must be noted, whether or not hard evidence about performances exists. As a consequence, the way incentive payments are used (as well as their effectiveness) depend on how supervisors appraise the workers’ performance.

Such subjective evaluations can be performed either by self-interested or, one should rather say, egoist supervisors (see, among many others, Tirole 1986; Laffont and Martimort 1997; Chang and Lai 2002, Bac and Kucuksenel 2006; or, in a different context, Prendergast and Topel 1996) or by supervisors who have relational feelings toward their employees (e.g., Kucuksenel and Saygi, 2019). The key implication of allowing for other-regarding preferences is that supervisors may decide to lie about their supervisees’ performances either because they care about them they are altruistic or because they want to harm them they are spiteful. This is the question we study in this paper: how altruism and spite affect the way

---

1 Millward et al. (1992: 388) assert that 34% of employees in the industrial sector in the United Kingdom received some form of merit pay “which depended on a subjective judgement by a supervisor or manager of the individual’s performance”. For similar evidence, see MacLeod and Parent (2000) and Gibbs et al. (2004).

2 That supervisors could lie about the performances of the agents they supervise is a rather standard assumption in the literature to which this paper relates (from Tirole 1986, to Saygili and Kucucksenel 2019). It is true that experiments have shown that lie aversion can be rather strong (see Cartwright et al. 2021). Yet, Cartwright et al. (2021) also point out that the evidence on lie aversion is yet to be conclusive, yielding very different predictions depending on the type of lie (white, black, altruistic, selfish and so on) and the type of experimental design. In addition, even if almost a half of the experimental subjects surveyed by Cartwright et al. (2021) were reluctant to tell even a Pareto white lie (i.e., a lie improving the wellbeing of all the agents involved), the share of individuals having no problem in telling lies range from 80 to 40% depending on the type of the latter.

3 An altruist is someone whose utility depends on her own gains and on the gains of the other individual with whom she interacts. Spitefulness is defined here as “an action in which an actor incurs some form of self-harm or cost in order to inflict harm or cost on another, even if the act will not result in any benefits for the actor” (Southard, 2017).
supervisors evaluate their supervisees’ performance, and how it affects the use of bonus payments as a worker discipline device.

Previous research has analyzed the effects of other-regarding preferences on the design of self-enforcing relational contracts between principals and agents (e.g., Prendergast and Topel, 1996; Brown et al., 2004, 2012; Casadesus-Masanell, 2004; Sliwka, 2007; Dur and Glazer, 2008; Kragl and Schmid, 2009; Shchetinin, 2010; Lee and Persson, 2011; Cordero Salas and Roe, 2012; Non 2012; Dur and Tichem, 2015). The closest works to the topic of the present paper are obviously those that study the effect of altruism and spite in organizations. Brown (2004) and Brown et al. (2012) theorize and experimentally prove that reciprocity leads to the emergence of long-term relational contracts and high effort provision. Cordero Salas and Roe (2012), in turn, analyze the interaction between a selfish agent and an altruistic principal and find that the credibility of the bonus is increasing in the principal’s altruism. The result is confirmed by Dur and Tichem (2015), who additionally show that when also the agent is altruistic or spiteful towards the principal, the optimal bonus depends on her other-regarding preferences as well. Thus, these articles show that both positive and negative relational feelings affect effort and performance.

The analyses developed in these works, however, remain in a two-tier structure. For obvious reasons, it is less reductive and more realistic to use a three-tier structure, in which a principal hires a supervisor to report about the performance of a group of agents. Indeed, real-world corporations normally have complex pyramidal structures, where each hierarchical level is headed by an employee-supervisor who, as already recalled by Marshall in his Principles (1980: 259), “has still to obey as well as to command”. Hence, “once the layer of supervisors is integrated into a firm’s hierarchy, a host of new problems will be spawned under incentive theories” (Chang and Lai, 2002: 28). Therefore, our contribution complements the literature on relational contracts by showing under what circumstances the possibility of supervisory violations annihilate the effort-inducing function of incentive pays.

The key intuitions are as follows: while a caring manager may find it hard to be harsh on loafing employees, spiteful supervisors may be unjust towards those who work hard to satisfy their psychological inclination. In both cases, the effort-enhancing function of bonus payments may be impotent due to the distortion of these supervisory infractions. The actual effect of these violations on effort and performance, however, also depends on the workers’ response to the possibility of receiving a wrongful evaluation. Indeed, our results suggest that altruistic (resp., spiteful) workers interacting with an altruistic (resp., spiteful) supervisor have greater incentives to provide effort because they want to avoid (resp., induce) their supervisor to commit a violation and run the risk of being punished by the principal.

Three-tier hierarchies were introduced in agency theory by Tirole (1986), who first considered an organization composed by principal, a supervisor and an agent to study collusive behaviors between supervisors and workers. Ever since then, this three-layered structure has become quite common in the analysis of organizational corruption (e.g., Laffont and Martimort, 1997; Chang and Lai, 2002; Bac and Kucuksenel, 2006) and not only (e.g., Prendergast, 2007), although, to the best of
our knowledge, few have used it to study the impact of other-regarding preferences in organizations (Kucuksenel and Saygi, 2019).

There are important similarities between our study and the literature on organizational corruption to which our paper relates tangentially—but also, some important differences (e.g., Tirole, 1986; Laffont and Martimort, 1997; Chang and Lai, 2002; Bac and Kucuksenel, 2006; Saygili and Kucuksenel, 2019). Indeed, our approach is close to the one developed in Saygili and Kucuksenel (2019). The latter are to the best of our knowledge the only ones who use a three-tier hierarchy to analyze the impact of relational feelings on the relationship between agents and supervisors. Even if they study a different type of other-regarding preferences—inequity aversion—they nonetheless derive conclusions in terms of effort and performance that are interesting for us. They show that the more inequity-averse agents are, the more effort they exert they do not want to receive a lower wage than the supervisor does. This compares to our result, according to which altruistic (resp., spiteful) workers tend to exert more effort than spiteful (resp., altruistic) workers when they are coupled with an altruistic (resp., spiteful) supervisors.

However, their purpose differs from ours. Indeed, while they study how other-regarding preferences affect the side-contract between corrupt supervisors and agents, we focus on the relationship between relational feelings and incentive pays. As a further development, it would be interesting to analyze how relational feelings affect the supervisors’ incentive to accept a bribe, and, more generally, to incorporate altruism and spite in the literature on organizational corruption. Are altruistic supervisors less open to bribery than spiteful ones? Or the reverse? And moreover: would hard-workers accepting their supervisors to be biased against them may want to send a signal to the principal or agree to be extorted and to pay a bribe to avoid the forthcoming untruthful evaluation? All the conclusions we could suggest in the present paper, however, would remain tentative, since (let us insist) our paper does not discuss collusion between supervisors and agents, but rather, analyze organizational infractions that take place before the flourishing of corrupt activities in firms. In this sense, our analysis can be seen as a “side investigation” of the determinants of organizational corruption, thus shedding light on the antecedents of collusion in firms. We leave the possibility of studying the effect of other-regarding preferences on organizational corruption to future research.

Finally, our contribution extends the discussion that analyzes the inefficiency of the efficiency wage (Chang and Lai 2002) to bonus-pay settings. Indeed, our findings suggest that some of the inconsistencies that have been found in efficiency wage theory i.e., that salaries are not always effective in promoting effort (Chang and Lai 2002) and that the wage and the supervision incentives cannot always be used as substitutes (Groshen and Krueger 1990; Chang and Lai 1999) carry over to bonus payments as well. In addition, our paper provides support to the findings of Campbell and Kamlani (1997), who argue that good-quality supervisor–employee relationships are more important in determining effort than good working conditions, high wages, and monitoring.

The remainder of the paper is organized as follows. Section 2 briefly reviews the evidence on relational feelings in organizations. Sections 3 and 4, respectively, present and analyze the model. Section 5 concludes.
2 Related evidence

The available evidence suggests that social relations in firms are a complex phenomenon, involving both positive and negative feelings, and also, that the organizational effects of the latter are mixed. Social exchange theory, for instance, has theoretically claimed and then empirically proved that positive relational outcomes at the workplace engender effective work behavior, favorable employee attitudes, and organizational commitment (see, e.g., Rhoades and Eisenberger 2002; Cropanzano and Mitchell 2005). In the same vein, both Griffeth and Riordan (1995) and Wagner and Harter (2006) show that good employers-employees relationships increase job-satisfaction, reduce labor turnover and boost work effort, while the evidence in Bandiera et al. (2010) suggest that positive relations among cowers increase labor productivity. Berman et al. (2002) corroborate this view by showing that friendships between managers and employees is a recurring phenomenon.

The finding of Cortina, et al. (2001), Moerbeek and Need (2003), and Morrison (2008), however, are less encouraging, showing that almost 71 percent of workers have been mistreated by both their co-workers and superiors, and that 8% perceive their relationship with their supervisor as either “bad” or “very bad”. These negative perceptions seem to hinder productivity and increase the workers’ intention to quit. Along similar lines, recent Gallup data reveal that almost a half of all US employees have quit their organization because of (perceived) negative relationships with their boss (Harter and Adkins 2015). This astonishingly large figure is consistent with what has become known as the “Peter Principle” (Peter and Hull 1969), according to which, undue promotions of incompetent or even abusive supervisors are a recurring phenomenon in organizations (Lazear 2004; Barmby et al. 2012). In addition, cultural beliefs and media representations continue to support the idea that firms are largely populated by bad bosses. These concerns seem also confirmed by recent evidence presented in Artz et al. (2020), who find that one in eight bosses can be classified as bad. The figure, although less striking that that put forward by Harter and Adkins (2015), is non-negligible. These pieces of evidence seem loosely in line with those suggesting that spitefulness is a robust phenomenon in human behavior (Falk et al. 2005; Herrmann et al. 2008; Norris et al. 2014). Indeed, a sizeable amount of experimental research has shown that spitefulness is more widespread than one would imagine, as many participants in experiments are willing to punish free-riders at own costs. 4,5

Although the stylized facts reported in the above cannot be taken as a direct indicator of the relevance of the behavioral mechanisms analyzed by our theoretical model, they suggest that both positive and negative relational feelings are pervasive

---

4 For an introduction to the vast experimental research on other-regarding preferences, see, for instance, Gätcher and Herrmann (2006) and the references therein.

5 This evidence provides support to the reasonable idea that individuals may be willing to sacrifice a share of their utility in order to punish deviating conducts. By contrast, we assume that some agents may be willing to forego their wellbeing damage those towards whom they have an antipathy. Despite we are aware that this is a strong behavioral assumption, we also believe that it may prove itself fructuous if empirically corroborated. Hence, we hope that our contribution will stimulate future experimental research in this direction.
at the workplace, and thus, that none of their theoretical implications should be neglected by the theory of organizations. The key finding of this paper, in this respect, is that negative relational feelings are not necessarily bad for effort and performance, at least as long as the interacting agents (i.e., workers and supervisors) share a common perception of their relationship. When both workers and supervisor are resentful toward each other, in fact, we find that other-regarding preferences boost work effort, which incidentally, is loosely in line the findings of Kragl and Schmid (2009), who suggest that there exist situations in which envy may increase organizational performance. Similarly, when both workers and supervisor care about each other, we find that other-regarding preferences boost work effort, which is consistent with the idea put forward by Brown (2004) and Brown et al. (2012) whereby reciprocity allows the emergence of long-term relational contracts and high effort provision.

3 The model

3.1 Timing and setup

We study a two-stage game where a team composed of \( n > 0 \) workers interact with a supervisor (she) through a three-layer hierarchy dominated by a principal (he). Our modelling strategy is borrowed from Chang and Lai (1999 2002), with the significant exception that we do not have network effects creating the possibility of multiple equilibria. To focus on the supervisor’s monitoring role, we assume that she performs no work of coordination or decision-making – i.e., she is “unproductive”.\(^6\) We assume that workers receive a baseline compensation \( w_l \geq 0 \) that is increased by a bonus \( w > 0 \) if the supervisor’s evaluation of their performance is positive. For future reference, define \( w_h = w_l + w \). Similarly, the supervisor is also paid a baseline wage \( s_l \geq 0 \) that is raised to \( s_h = s_l + s \) if the principal’s evaluation of her performance is positive. We assume that the bonuses are credible, i.e., that the principal finds it rational not to renege the latter after the workers’ effort choice. In addition, we assume that the baseline compensations \( w_l \) and \( s_l \) satisfy both the workers’ and the supervisor’s participation constraints, so that both type of agents find it rational to participate in the firm when given the opportunity to do so. Finally, we normalize both the workers’ and the supervisor’s outside option to zero.

The timing of the game is as follows. At stage 1 (the design stage), the principal chooses the optimal bonus for both workers and the supervisor and the optimal team size. By doing so, he also defines the supervisor-worker ratio, \( 1/n \), that determines the probability \( 1<p<0 \) that a worker’s performance will be evaluated by the supervisor during the productive period. Indeed, the ratio or probability \( p = 1/n \) defines the supervisor’s “span of control”, or, alternatively, “the number of employees a supervisor can handle effectively” (Williamson 1967: 127). Here, the

\(^6\) The assumption of unproductive supervisors and effortless supervision (see below) is common in the principal-supervisor-agent theory of organizational hierarchies (e.g., Tirole 1986; Chang and Lai 2002; Kucuksenel and Saygi 2019).
assumption, for simplicity, is that the supervisor can interact with no more than a worker per unit of time.

At stage 2 (the production stage), workers choose between high and low effort, respectively yielding high and low output. As it is common in the principal-agent literature, we assume that workers can provide low effort at no cost, while high effort entails a strictly positive cost $e > 0$.

At stage 3 (the supervision stage), the supervisor is randomly assigned to one of the $n$ workers in the team to (effortlessly) evaluate his/her performance. The situation we have in mind is one where the relevant time period is sufficiently short (e.g., a morning) so that the supervisor can pick one and only one of the $n$ members in her team to check upon her doing. In this setting, workers choose their effort level at the beginning of stage 1, without being able to anticipate whether they will be inspected by the supervisor. After that, the supervisor randomly decides whom to oversee and the game moves to stage 2.

Upon meeting, we assume that both the supervisor and her supervisee develop some relational feelings towards each other, either positive (altruism) or negative (spite) see Sect. 3.3 for further discussion. Of course, this does not necessarily imply that the supervisor and the unsupervised supervisees have no relational concerns towards each other, but, more simply, that such feelings are irrelevant when it comes to evaluating the possibility of supervisory infractions.

Finally, after (perfectly) observing the worker’s effort, the supervisor decides whether to provide a true or false evaluation of her supervisee’s performance see Sect. 3.2. for further discussion. Then, after the principal receives the supervisor’s evaluation, he (imperfectly) assesses the validity of latter and pay wages accordingly. To avoid obfuscating the mechanisms at play, we assume that when the supervisor’s evaluation is truthful, the principal recognizes the latter as such. Conversely, and in line with Chang and Lai (2002), we assume that when the supervisor commits misconduct, there is a probability $0 < q < 1$ that she is detected by the principal. To summarize, the sequence of events is as follows:

1. The principal chooses the optimal team size and the optimal size of the bonus payments.
2. Workers choose their effort level. Production takes place for the period.
3. One of the $n$ workers is randomly evaluated by the supervisor.
4. The supervisor and the supervisee develop some relational feelings towards each other.
5. The supervisor decides whether to provide a truthful evaluation of her supervisee.
6. The principal assesses the supervisor’s evaluation and pay wages accordingly.

### 3.2 Supervisory violations

We study the possibility of two different types of supervisory infractions. First, when the worker provides high-effort, the supervisor may decide to give a negative evaluation of his performance. In this case, if the principal does not detect the supervisor’s misconduct ($q = 0$), he punishes the hard-worker with the low wage...
(type-1 error) and rewards the dishonest supervisor with the high wage (type-2 error). Conversely, if the principal detects the supervisor’s misconduct ($q = 1$), he rewards the hard-worker with the high wage and punishes the dishonest supervisor with the low wage.7

Second, when the worker provides low-effort, the supervisor may decide to give a positive evaluation of his performance. In this case, if the principal does not detect the supervisor’s misconduct ($q = 0$), he rewards both the shirking-worker and the dishonest-supervisor with the high-wage (type-2 errors). Conversely, when the principal detects the supervisor’s misconduct ($q = 1$), he punishes both the shirking-worker and the dishonest supervisor with the low wage (Table 1).

### 3.3 Other-regarding preferences

In our framework, the key reason for which supervisors may decide to issue a false evaluation of their workers’ performance (and that workers may provide low effort) is that both may find it rational to accommodate their other-regarding preferences. As usual, we model the latter through utility interdependence. In particular, we assume that the supervisor’s utility, $V$, is given by the sum of her payoff, $v$, and the representative worker’s payoff, $u$. Similarly, we assume that the utility of a representative worker, $U$, is given by the sum of his/her payoff and the payoff of the supervisor. In addition, we assume that the psychological gains that supervisors and workers derive from their other-regarding preferences must be weighted for a couple of parameters $\alpha$ and $\beta$, where $\alpha$ (resp., $\beta$) measures the supervisor’s (resp., worker’s) altruism (or spite). Hence, we have that $V = v + \alpha u$ and $U = u + \beta v$. Clearly, positive values of $\alpha$ or $\beta$ indicate altruism, while negative values of $\alpha$ or $\beta$ indicate spite. In case of altruism, we assume that the players’ other-regarding preferences are drawn from two uniform distribution between 0 and 1 with probability density function $g(\alpha)$ and $h(\beta)$, while, in case of spite, these preferences are drawn from a uniform distribution between $-1$ and 0 with probability density function $j(\alpha)$ and $k(\beta)$. Finally, we assume that all agents are perfectly informed of whether the individuals they are interacting with are altruistic or spiteful, but cannot observe the exact value of $\alpha$ and $\beta$. One may think of this assumption as describing a

| Type I violation (w. prob.$\delta$) | Type II violation (w. prob.$e$) |
|-----------------------------------|----------------------------------|
| The supervisor gives a negative evaluation of a hard-worker’s performance | The supervisor gives a positive evaluation of a shirking-worker’s performance |
| ↓ | ↓ |
| The hard-worker receives $w_l$ (Type-I error) | The shirker receives $w_h$ (Type-II error) |

$^7$ In addition, we may assume that the supervisor receives an additional penalty for her dishonesty. Despite this would have the quantitative effect of discouraging further the issuing of false reports, it would not alter our results qualitatively. Hence, for the sake of simplicity, we omit this possibility.
situation where agents have a generally correct though imprecise perception of the social environment. Hence, we assume that players calculate the equilibrium probability of a given behavior from the probability density functions described above.

### 3.4 Game’s equilibrium (definition)

The game tree in Fig. 1 serves as a helping tool to visualize the structure of interactions across the second and third stage of the game. As anticipated, the probability $1 < \gamma < 0$ that the representative worker will shirk, the probability $0 < \delta < 1$ that the supervisor will give a negative evaluation of a hard-worker’s performance (type-I violation) and the probability $0 < \varepsilon < 1$ that the supervisor will give a positive evaluation of a shirker’s performance (type-II violation) will be determined endogenously as functions of the agents’ other-regarding preferences. In addition, while workers take the probability $p$ of being evaluated by the supervisor as exogenously given, recall that this is defined by the principal’s decision concerning $n$ during the first stage of the game, since $p = 1/n$. Finally, the probability $q$ that the principal will detect a forgiving supervisor, as anticipated, is exogenous.
The game is solved by backward induction. In the third stage of the game, we shall look for a couple \(\{\delta^*, \varepsilon^*\}\) that describes, respectively, the equilibrium probability that that the supervisor will give a negative evaluation of a hard-workers’ performance and the equilibrium probability that she will give a positive evaluation of a shirking-workers’ performance. In the second stage of the game, we shall look for the equilibrium probability \(\gamma^*\) that the representative worker will provide high effort. Hence, the subgame perfect equilibrium is described by a triple \(\{\gamma^*, \delta^*, \varepsilon^*\}\). In the first stage of the game, we shall look for an optimal triple \(\{n^*, w^*, s^*\}\) that describes the principal’s profit-maximizing decision concerning the optimal team size and the optimal wage premia.

### 3.5 Simplifications and assumptions

Of the above assumptions, the following three are admittedly limiting. First, as already put forward by Mintzberg (1971), in real-world organizations supervisors play a variety of different roles, that typically include planification, organization, coordination, and control. In a more sophisticated setting, a supervisor may have to choose how to allocate her time among these different activities. This may involve relevant tradeoffs that are not accounted for by our model. For instance, she may need to optimally sacrifice some of her productive time (for instance, dedicated to organization and coordination) to ensure that her guidelines are correctly implemented by low-level workers.

Second, we have assumed that the supervisor can inspect the agency of her subordinates at no cost, since she exerts no effort when randomly evaluating the performance of 1 of the \(n\) employees in her team. Alternatively, we may assume that inspecting the workers’ performance entails some monitoring costs for the supervisor, for instance, because she has to exert some effort to check upon her supervisees, as in Prendergast’s (2007) analysis of bureaucrats. The key implication is that the principal should find ways not only to impede the appointed middle-man to misreport, but also, to induce the latter to put forward her supervisory effort. Albeit more realistic, effort-incentives have been extensively studied in principal-agent settings. Hence, we omit this possibility, as to better focus on the channels we wish to shed light upon.

Third, the principal may use the extra-profit stemming from high-effort to finance the incentive-pays of both workers and supervisor. In this case, the size of the bonuses would depend on labor productivity. The upshot is that both the workers’ and the supervisor’s decisions to misbehave would depend less on their relational feelings and more on the expected reduction of their bonuses. However, there are at least two channels that may weaken this incentivizing mechanism. First, when the team is large, both shirkers and dishonest supervisors may acknowledge that the marginal effect of their misbehavior on the size of their bonus is negligible. Second, because of team production, the usual temptation to free-ride on the others’ effort may create a classic Prisoner-Dilemma leading workers to exert low effort. Hence, even when these channels are taken into account, we believe that room is left for
relational misbehaviors to occur, despite, probably, less pervasively than suggested by the present paper.

4 Analysis

4.1 The supervision stage: type I violations

To derive the equilibrium probability $\delta^*$ that the supervisor will commit what we call a “type-I violation” and provide a negative evaluation of a hard-worker’s performance, we compute the following expected utility functions, that describe, respectively, the payoffs of an honest and dishonest supervisor interacting with a hard-worker:

$$V(d = 0) = s_h + \alpha(w_h - e)$$

$$V(d = 1) = q[s_l + \alpha(w_h - e)] + (1 - q)[s_h + \alpha(w_l - e)]$$

When the supervisor’ evaluation is truthful ($d = 0$), both she and her supervisee receive the high wage. Conversely, when she decides to commit misconduct ($d = 1$), there is an exogenous probability $q$ that the principal will detect her misbehavior. In this case, the principal punishes the dishonest supervisor with the low wage and rewards the hard-worker with the high wage. Conversely, with probability $1 - q$, the supervisor gets away with her infraction. In this case, the principal unduly rewards the dishonest supervisor with the high wage (type-II error) and punishes the hard-worker with the low wage (type-I error). To rule out the possibility of trivial cases, we impose the following limitation upon set of parameters’ value:

**Assumption 1** The principal’s monitoring ability $q$ is not so high as to make supervisory infractions irrational. Parametrically, this corresponds to $q < w/(w + s)$.

It is straightforward to see that a supervisor will commit a type-I violation if and only if $V(d = 1) \geq V(d = 0)$, or alternatively:

---

8 As anticipated in the introduction, we may allow workers suspecting their supervisor to be biased against them to send a signal to the principal at a private cost. Whether or not the principal will believe the workers’ signal, in turn, may become a matter of hard evidence (s/he who presents the more reliable report will be believed by the principal) or of relative trust. Indeed, when the principal trusts the supervisor more than the supervisee, he may ignore the worker’s signal or even take this as a further demonstration of guilt, and vice-versa. The key implication of allowing for this reasonable possibility is as follows. When (and provided that) the worker finds it rational to send the signal to the principal, the expected value of $q$ plausibly increases, implying that a supervisor who issues a negative evaluation of a hard-worker would face a greater probability of being detected. In this framework, the probability $q$ would be infraction-contingent, so that $q = q_h$ in case of type-II violations and $q = q_l$ in case of type-II violations, with $q_h > q_l$. While this would straightforwardly reduce the frequency of type-II violations in the organization, it would not alter any of our results qualitatively, but for the limit-case where the increase in the expected value of $q$ is so significant to discourage this type of supervisory wrongdoing altogether.
$$qs + \alpha(1 - q)w \leq 0$$

Equation (1) describes a “dishonesty condition” from which we derive the equilibrium probability $\delta^*$ that the supervisor will give a negative evaluation of a hard-worker’s performance, which is given in the following Proposition:

**Proposition 1** When the supervisor is altruistic and $\alpha \geq 0$, we have that $\delta^* = 0$. Conversely, when the supervisor is spiteful and $\alpha < 0$, we have that $\delta^* = \frac{[(1 - q)w - qs]/(1 - q)w}$. In addition, when $\alpha < 0$, $\delta^*$ is increasing in $w$ and decreasing in $s$ and $q$.

**Proof:** See the Appendix The key message from Proposition 1 is that the equilibrium probability of type-I violations chiefly depends on the supervisor’s other-regarding preferences. Intuitively, when the supervisor is altruistic and cares for her supervisee, she has no incentive to issue a false evaluation of her performance. Conversely, when the supervisor is spiteful, she may decide to misreport her supervisee’s performance to satisfy the psychological pleasure of watching a hard-worker being denied the bonus payment by the principal.

The first comparative statics predicts that type-I violations are more likely when the worker’s bonus $w$ is high. The intuition is as follows. Since spiteful supervisors are psychologically rewarded when their supervisees run the risk of being punished by the principal, the higher is the bonus $w$, the more severe will be the implication of not receiving the latter and thus, the higher is the subjective utility of a spiteful supervisor. Interestingly, this entails that raising the workers’ bonus will create greater incentives for resentful supervisors to commit misconduct. The second and third comparative statics are intuitive and predict that the equilibrium rate of type-I violations decreases when either the probability of being caught $q$ or the bonus payment $s$ is high.

### 4.2 The supervision stage: type II violations

Following steps identical to those in the case of type-I violations, we derive the equilibrium probability $\varepsilon^*$ that the supervisor will provide a positive evaluation of a shirker’s performance. To do so, we compute the expected utility functions of an honest and dishonest supervisors interacting with a shirker:

$$V(\varepsilon = 0) = s_h + \alpha w_l$$

$$V(\varepsilon = 1) = q(s_l + \alpha w_l) + (1 - q)(s_h + \alpha w_h)$$

When the supervisor is truthful in informing the principal that the worker provides low-effort ($\varepsilon = 0$), the principal rewards her with the high wage and punishes the shirker with the low wage. Conversely, when she misreports the loafing worker’s performance ($\varepsilon = 1$), the principal detects this violation with probability $q$. In this case, he punishes both the dishonest supervisor and the loafing-supervisee with the low wage. Conversely, with probability $1 - q$, the principal does not recognize the supervisor’s infraction and unduly rewards both her and the loafing-supervisee.
with the high wag (type-II error). As before, it is straightforward to see that a supervisor will commit a type-II violation if and only if \( V(\varepsilon = 1) \geq V(\varepsilon = 0) \), or alternatively:

\[
qs - z(1 - q)w \leq 0
\]

(2)

The “dishonesty condition” described in equation (2) is similar to condition (1), but for the fact that the second term that captures the supervisor’s psychological utility is now negatively signed. Hence, we can put forward the following Proposition:

**Proposition 2** When the supervisor is spiteful and \( z \leq 0 \), we have that \( \varepsilon^* = 0 \). Conversely, when the supervisor is altruistic and \( z > 0 \), we have that \( \varepsilon^* = \delta^* = [(1 - q)w - qs] / (1 - q)w \). In addition, when \( z > 0 \), \( \varepsilon^* \) is increasing in \( w \) and decreasing in \( s \) and \( q \).

**Proof:** See the Appendix. The message of Proposition 2 is almost “symmetrical”, so to speak, to that of Proposition 1. As before, the probability that the supervisor will commit a type-II violation depends on her-other regarding preferences. Intuitively, when the supervisor is spiteful and derives psychological pleasure from observing a shirker being denied the bonus by the principal, she has no incentive to issue a positive evaluation of her performance. Conversely, when she is altruistic, she may decide to turn a blind eye upon shirkers to avoid the psychological cost of punishing the latter.

The signs of the second and third comparative statics, in turn, convey the same messages discussed in Proposition 1: the supervisor will be less inclined to commit misconduct when either the probability of being caught \( q \) or the bonus payment \( s \) is high. That the equilibrium probability \( \varepsilon^* \) is increasing in \( w \), in turn, can be rationalized as follows. When the supervisor is altruistic, she suffers a psychological utility loss when her supervisee gets punished by the principal and this, in turn, increases with the severity of the punishment. The upshot, as before, is that providing workers with higher bonuses has the unintended effect of raising the supervisory incentives to commit misconduct.

### 4.3 The production stage

To derive the endogenous probability \( \gamma^* \) that the representative worker will provide high-effort, we compute:

\[
U(\gamma = 1) = (1 - p)w_l + p\Phi_1 - e
\]

\[
U(\gamma = 0) = (1 - p)w_l + p\Phi_0
\]

9 In an efficiency-wage setting such as that of Chang and Lai (2002), the worker’s compensation at this state would be \( w_h \). However, since the first terms in both \( U(\gamma = 1) \) and \( U(\gamma = 0) \) are identical and have no effect on the supply of labor effort, leaving an employment rent to the worker would be irrational from the principal’s viewpoint.
With probability $1 - p$, the worker is not evaluated by the supervisor, regardless of his/her effort choice. In this case, the expected wage payment is $w_l$. Conversely, when his/her performance is verified which happens with probability $p$ the worker’s expected compensation (and psychological utility) depend on the supervisor’s behavior, as can be seen from the following equations:

$$
\Phi_1 = (w_h + \beta s_h)(1 - \delta) + [q(w_h + \beta s_l) + (1 - q)(w_l + \beta s_h)]\delta
$$

$$
\Phi_0 = (w_l + \beta s_h)(1 - \epsilon) + [q(w_l + \beta s_l) + (1 - q)(w_h + \beta s_h)]\epsilon
$$

A first and intuitive remark that can be drawn by looking at the above equations, is that, in the absence of supervision ($p = 0$), workers would always find it rational to provide low effort. This is consistent with the standard wisdom from efficiency wage theory, that postulates that the wage and the supervision incentive should always be used as complements (Shapiro and Stiglitz, 1984). Given Propositions 1 and 2, it is clear that the worker’s expected utility when s/he interacts with the supervisor varies according to the other regarding preferences of the latter see Table 2. Indeed, making use of the equilibrium values $\delta^*$ and $\epsilon^*$ given in Propositions 1 and 2 and rearranging, we see that $U(\gamma = 1) \geq U(\gamma = 0)$ if:

$$
w + s + \beta s \epsilon^* - e/pq \geq 0
$$

(3)

if $\alpha > 0$, and

$$
w + s - \beta s \delta^* - e/pq \geq 0
$$

(4)

if $\alpha < 0$. Equations (3) and (4) describes the “no-shirking-condition” of the model, that varies according to whether the supervisor develops positive or negative relational feelings towards her supervisee. Given the heterogeneity in the above condition, the equilibrium probability $\gamma^*$ will also vary across the four combinations of $\alpha$ and $\beta$. It nonetheless remains that we can put forward two general results: equation (3) makes it clear that the workers’ altruism has a positive effect on labor effort only if the supervisor is altruistic as well. Complementarily, equation (4) conveys a similar message: spiteful workers will be more inclined to put forward their labor effort if their supervisor too develops negative relational feelings. The intuition is as follows.

When supervisors care for their employees, workers know that if they shirk, with probability $\epsilon^*$ the altruistic supervisor will conceal their infraction, even though this entails running the risk of being caught by the principal. To avoid the associated psychological costs, altruistic workers will be thus inclined to provide effort to avoid the supervisor to face such a situation. Reciprocally, when supervisors develop negative relational feelings towards her supervisees, spiteful workers who gain pleasure from their supervisors’ misfortune anticipate that if they provide high effort, with probability $\delta^*$ they will be given a negative evaluation because of the...
supervisors’ spite, and this, in turn, exposes the latter to the possibility of being caught by the principal.

The mixed situations where supervisors and workers develop diverging relational feelings, in turn, have a negative effect on the supply of labor effort. The behavioral channels are as follows. On the hand, resentful workers interacting with altruistic supervisors may decide to provide low-effort internalizing the probability of being given a positive evaluation by their supervisor. On the other hand, when workers care for their supervisor’s well-being but this feeling is not reciprocated, they may decide not to put forward their labor effort to avoid the situation where a resentful supervisor provide a negative evaluation of a hard-worker’s performance because of her spite towards the latter.

To investigate further the mechanisms we have just described, we put forward the following Proposition:

**Proposition 3** When workers and supervisors have reciprocal relational feelings – i.e., \( a > 0 \) and \( \beta > 0 \) or \( a < 0 \) and \( \beta < 0 \) – we have that \( \gamma^* = 1 - \left[ e - pq(w + s) \right]/pqs_0p \). Conversely, when workers and supervisors have diverging relational feelings – i.e., \( a > 0 \) and \( \beta < 0 \) or \( a < 0 \) and \( \beta > 0 \) – we have that \( \gamma^* = -\left[ e - pq(w + s) \right]/pqs_0p \), where \( \varphi = e^* \) if \( a > 0 \) and \( \varphi = \delta^* \) if \( a < 0 \). Finally, while \( \gamma^* \) is decreasing in \( e \) and increasing in \( p \) regardless of the value of \( \alpha \), when \( \alpha > 0 \), \( \gamma^* \) is increasing in both \( q \) and \( s \), while \( \gamma^*_{w} \) cannot be unambiguously signed. Conversely, when \( \alpha < 0 \), \( \gamma^* \) is decreasing in \( w \), while \( \gamma^*_q \) and \( \gamma^*_s \) cannot be unambiguously signed.

**Proof: See the Appendix.** Proposition 3 clarifies that the extent to which relational incentives undermines organizational performance depends on whether agents develop reciprocal or diverging concerns towards each other. In addition, it shows that there exist situations (\( \gamma^* \) is decreasing in \( w \) and \( s \)) where the high-wage program imposed on supervisors and employees cannot be relied upon as a worker discipline device.

The remaining comparative statics can be commented as follows. First, when the displeasure of exerting effort is high, workers have lesser incentives to fulfill their organizational duties (\( \gamma^* < 0 \)). In turn, that \( \gamma \) is increasing in the supervisor’s span of control, \( p = 1/n \), and in the principal’s monitoring ability \( q \), is consistent with one of the key predictions of efficiency wage theory, namely, that tighter monitoring induces workers to increase their labor effort. However, when \( \gamma \) is decreasing in \( q \), raising the rate at which the principal detects the supervisor’s infractions has a negative effect of labor effort. Again, this is due to the distortion of relational incentives. The mechanisms are as follows.

---

**Table 2** Worker’s expected utility (net of effort cost) when \( p = 1 \)

| \( \Phi_1 \) | \( \Phi_0 \) |
|---|---|
| \( \alpha > 0 \) | \( w_h + \beta s_h \) | \( w_l + \beta s_h - [q(w + \beta s) - w]e^* \) |
| \( \alpha < 0 \) | \( w_h + \beta s_h + [q(w - \beta s) - w]e^* \) | \( w_l + \beta s_h \) |

---

Supervise me if you can. Relational feelings, incentive pays and...
On the one hand, resentful shirkers know that with probability $e^*$ they will receive a positive evaluation from a caring supervisor and that, with probability $q$, this act of mercy will lead the supervisor to suffer a wage-loss because of her excess of altruism. This, in turn, will convey a psychological gain to spiteful workers. When such gains are large enough, $\gamma$ is decreasing in $q$. On the other hand, when workers are altruistic but interact with a spiteful supervisor, they know that if they decide to provide high-effort they will receive a negative evaluation with probability $d^*$, and that, this, in turn, will lead the supervisor to suffer a wage loss because of her excess of spite with probability $q$. Anticipating this, workers may find it rational to provide low effort to avoid the psychological cost associated to the supervisor’s loss. In this case, $\gamma$ is decreasing in $q$.

4.4 The design stage

At stage 1, the principal maximizes his expected profits by selecting an optimal triple $\{w^*, s^*, n^*\}$ internalizing the supervisor’s and worker’s decision at stage 2 and 3. To keep things simple, we assume that the values of $w_l$ and $s_l$ are exogenously determined by the outside market conditions and we further normalize $w_l = s_l = 0$. The normalization is made for the sake of tractability. By assumption, since $w_h = w_l + w$ and $s_h = s_l + s$, we have that $w_h = w$ and $s_h = s$. Recall that during each productive period, only one of the $n$ workers is evaluated by the supervisor. Hence, while the remaining $n - 1$ workers receive the baseline compensation regardless of their effort choice, the payments of the idiosyncratic couple of worker and supervisor who actually interact in production depend on their behaviors at stage 2 and 3. With these facts in mind, the principal’s problem is given by:

$$
\max_{w, s, n} \Pi = f(x) - \gamma c(\delta, w, s) - (1 - \gamma)k(\epsilon, w, s)
$$

s.t.

$$
\gamma = \gamma^*(w, s, n) \\
\delta = \delta^*(w, s) = \epsilon = \epsilon^*(w, s) \\
c(\delta, w, s) = (1 - \delta)(w + s) + \delta[qw + (1 - q)s] \\
k(\epsilon, w, s) = (1 - \epsilon)s + \epsilon(1 - q)(w + s)
$$

where $x = n\gamma$ is the firm effective workforce given by the size of the team times the average effort level and $f(x)$ is an only-labor production function that satisfies the usual conditions $f_x > 0$ and $f_{xx} < 0$.\footnote{The modelling strategy of the principal’s profit is borrowed from Chang and Lai (1999).} Hence, the first order conditions for optimal profits are given by:

$$
\Pi_n = f_x(\gamma + n\gamma_n) - \gamma_n(c - k) = 0 \quad (5)
$$
\[ \Pi_w = f_s n \gamma_w (c - k) - \gamma (c_w - k_w) = 0 \quad (6) \]
\[ \Pi_s = f_s n \gamma_s (c - k) - \gamma (c_s - k_s) = 0 \quad (7) \]

We assume that the second-order conditions are satisfied. Before proceeding, in the may be useful to advance the following Lemma, which will be handy for the analysis that follows:

**Lemma 1** In equilibrium, \( c - k > 0 \), while \( c_w - k_w > 0 \) and \( c_s - k_s < 0 \) when \( \alpha > 0 \) and \( c_w - k_w < 0 \) and \( c_s - k_s \geq 0 \) when \( \alpha < 0 \).

**Proof:** See the Appendix From the set of equations reported above, the following remarks are worth drawing. First, from Proposition 3, we already know the average work-effort is increasing in the supervisor’s span of control (\( \gamma_p^* > 0 \)).\(^{12}\) The upshot is that workers will be employed until the marginal cost of hiring an additional unit of labor equals the marginal productivity of the latter minus the marginal decrease in the supervisor’s monitoring ability, as suggested by equation (6). This is consistent with the framework developed by Williamson (1967), where the size and structure of the firm is fundamentally constrained by a series of control loss phenomena similar to those that we have just described.

Second, given the statement in Lemma 1 and the fact that the high-wage program imposed on worker and supervisors may be impotent due to the distortion of relational incentives (\( \gamma_w < 0 \) and \( \gamma_s < 0 \)), we see that there exists situations where the first order conditions in eqns (7) and (8) are violated (\( \Pi_w < 0 \) and \( \Pi_s < 0 \)). In this case, bonus payments cannot be relied upon as a worker discipline device.

### 4.5 On the inefficiency of bonus payments

A rather large literature exists supporting the view that the effectiveness of incentive pays should not be taken for granted. More precisely, most of the existing literature analyzed the issue focusing on efficiency wages. Thus, Campbell and Kamlani (1997) and Nickell and Nicolitas (1997), for instance, show that the one-way causation from wage to productivity is overstated, whereas Manning and Thomas (1997) and Autor (2003) conclude that the evidence on efficiency wages is inconclusive. A number of other studies conversely find support to some of the key predictions of efficiency wage theory, for instance, that the wage and the monitoring incentives can be used as substitutes to motivate work effort – see, for instance, Cappelli and Chauvin (1991)). Even this evidence, however, is not uncontroversial: Groshen and Krueger (1990) find that while the wage and the supervision incentives are substitutes in the nursing occupation, they are complements in other three occupations, while Leonard (1987) finds a positive relationship between the two instruments.

Theoretically, the ambiguous effect the of efficiency wage has been explained in different ways. Lazear (1989) shows that when wages depend on relative

\(^{12}\) Formally, this is given by the fact that \( \gamma_p > 0 \), which entails that \( \gamma_s < 0 \), given the assumption that \( p = 1/n \).
performances, workers will have an incentive not only to boost their productivity, but also, to jeopardize their co-workers’. Perri (1994), in turn, highlight that employees may divert some of their production time to lobby their supervisors for higher wages, whereas Chang and Lai (2002: 30) show that increasing the workers’ wage “opens up room for arbitrage to urge collusion between workers and supervisors”, as the former may bribe the latter to avoid the punishment of losing their job. From the viewpoint of industrial psychology, the incentives that leverage on the workers’ extrinsic motivation (such as the wage) have the unintended result of crowding out their intrinsic motivation and morale (Frey 1993).

With this respect, our contribution is twofold. First, while Chang and Lai (2002) show that the wage premium paid to supervisors can always be relied upon when it comes to motivating workers, we find that there exist situations where this is not true due to the distortion of relational incentives. Second, we show that the “inefficiency of the efficiency wage” discussed in Chang and Lai (2002) may extend to bonus payments as well.

In addition, our theory challenges another key prediction of efficiency wage theory, namely, that the wage and the supervision incentive can be used as substitute to induce work effort. Indeed, when both \( r_p^s \) and \( r_w^s \) are positively signed, tightening workplace or raising the wage level yield identical effect on organizational compliance. In this case, small teams with a high supervisor-worker ratio or high compensation levels may equally do when it comes to disciplining workers. Conversely, when the efficiency wage is ineffective \( (r_w^s < 0) \), the monitoring and the wage incentives should be used as complements. This is in line with the evidence presented by Leonard (1987) and the theory developed by Chang and Lai (1999), who expand on Frey’s (1993) intrinsic motivation theory and show that when monitoring crowds out work morale, the wage and the supervision incentives cannot be used as substitutes.

5 Conclusion

In this paper, we develop a simple game-theoretic framework to study the impact of other-regarding preferences in a three-tier organization where a principal hires a supervisor to monitor a team of potentially shirking workers. Both workers and supervisors are motivated by a wage premium that is paid in case of good performance. The key hypothesis is that workers and supervisors develop relational concerns (either positive or negative) after interacting with each other. The implications of this simple assumptions are multiple.

First, since altruistic supervisors resent punishing loafing workers, they may decide to indulge their supervisees’ misconduct and provide a positive and false evaluation of their performance. In doing so, however, they run the risk of being caught by the principal. Anticipating this possibility, altruistic workers will be less inclined to shirk, while spiteful agents who gain psychological pleasure from their supervisors’ misfortune will have an extra incentive to do so. Hence, we find that while the supervisors’ concerns for their workers’ wellbeing have a negative effect on labor effort, but also, that this effect is mitigated by the workers’ concerns for
their supervisors. In the same vein, we show that resentful supervisors may find it rational to provide a negative and false evaluation of a hard-worker’s performance because of their spite towards the latter. Given the possibility that the principal may discover this violation and deny the bonus to the supervisor, we also show that spiteful workers interacting with spiteful supervisors have a relational incentive to provide high effort.

These reasonable behavioral mechanisms have several implications for the use of incentive payments in organizations. Expanding on the line of research that rationalizes the “inefficiency of the efficiency wage” (Chang and Lai 2002), in fact, we show that there exist situations where relational incentives jeopardize the high-wage program imposed on workers and supervisors. In this case, bonus payments have the unintended effect of reducing the probability of high effort and thus, cannot be relied upon as a worker discipline device.

We are aware that a number of countervailing mechanisms may limit the explanatory power of our theory. For instance, both workers and supervisors may develop relational feelings towards the principal, and, depending on the sign of the latter, these may either reinforce or weaken the mechanisms outlined by the present paper. Similarly, when social norms are allowed into the picture, agents may be intrinsically motivated to fulfill their organizational duties (Frey 1993; Chang and Lai 1999, 2002). In this case, a principal may find it rational not only to invest in some kind of incentives scheme, but also, to complement or substitute the latter with the costly creation of shared norms and codes of conduct.\footnote{For a model analyzing how norms may be consciously created by agents through an investment process, see Corneo and Jeanne (1997).} In this framework, organizational culture is likely to have a pivotal role (Camerer and Vepsalainen 1988; Aktas et al. 2011; Pathiranage et al. 2020). Indeed, the results we put forward have rather to do with the nature of the relationships between supervisors and workers and, more broadly, with the culture that prevails within a firm than with the specific relationships between pairs of supervisors and agents. At least, these results should not be interpreted as suggesting that firms should tailor their work contracts to each supervisor, agent or pair of supervisor and agents. That would be obviously unmanageable. What we show, conversely, suggests that a good atmosphere, good relationships and a good corporate culture such as one promoted by positive altruistic relationships is important for the performance of a firm. In other words, our paper suggests that principals should create bonding opportunities within and across the different hierarchical layers, strive to counterfeit workplace bullying, mobbing and excessive gossiping and tighten the sense of organizational citizenship that bond together all the agents in the organization.

Moreover, a sizeable stream of research has shown that critical mass effects are key to the diffusion of deviant behaviors. Indeed, as recalled by Antoci and Sacco (1995: 89) “in an environment where corruption is the norm, its deterrence is relatively expensive and thus corruption tends to persist. Vice versa, in an environment where corruption is relatively infrequent, it can be detected and punished at relatively little cost”. Hence, the extent to which degenerate behaviors becomes established as social conventions may depend on the status-quo rate of the...
latter. Even when these channels are taken into account, however, none of our results is altered qualitatively. Hence, we leave the possibility of further behavioral refinements for future research.

Finally, we believe that the similarities between our contribution and Prendergast’s (2007) tentatively suggests some possible other applications of our setting. In his model, a principal hires a set of bureaucrats (or “middle-men”) that choose their effort level to correctly assign a benefit to a client (a “citizen”). The working assumption is that former may be biased towards the latter, either positively or negatively. The implication is that there exist situations where this produces undesirable societal outcomes, for instance, like those where tax inspectors “go soft” on tax-evaders by internalizing their preferences for not being investigated. Hence, although Prendergast does not explicitly talk about relational feelings, we believe that his study is worth mentioning to show the possible alternative applications of our framework. More generally, we believe that the three-tier hierarchy studied in this paper, along with its specific relational problems, can provide a simplified representation of a number of other social scenarios, yielding relevant insights beyond relationships in firms, such as, for instance, educational studies or political economy. Mutatis mutandis, in fact, our framework could be applied to the tree-layered interaction involving parents-teachers-students, as well as governors-officers-citizens.

Appendix: Proofs

Proof of Proposition 1

That Eq. (1) is never satisfied when \( z \geq 0 \) is straightforward. In this case, \( \delta^* = 0 \). To prove the second part of the Proposition, we define \( \hat{z} = -\frac{qs}{1-q} \) as the value of \( z \) that solves Eq. (1) as an equality. This is the level of spite that makes the supervisor with \( z = \hat{z} \) just indifferent between giving a positive or a negative evaluation of her supervisee’s performance. Given that \( z \) is uniformly distributed over \([0, q] \) when \( z \leq 0 \), the equilibrium probability that the supervisor will commit a type-I violation is given by \( \delta^* = \frac{\hat{z}}{\int j(z)dz} = 1 + \hat{z} \). Substituting \( \hat{z} \) in this expression gives the value of \( \delta^* \) reported in Proposition 1. The third part of the Proposition, in turn, follows from the fact that \( \delta^*_w = \frac{qs}{(1-q)w} > 0 \); \( \delta^*_q = -\frac{q}{(1-q)^2} < 0 \) and \( \delta^*_s = -\frac{s}{(1-q)w} < 0 \), where subscripts, as usual, denote partial derivatives.

Proof of Proposition 2

That Eq. (2) is never satisfied when \( z \leq 0 \) is straightforward. In this case, \( \varepsilon^* = 0 \). To prove the second part of the Proposition, we define \( \tilde{z} = \frac{qs}{1-q} \) as the value of \( z \) that solves Eq. (2) as an equality. This is the level of altruism that makes the supervisor with \( z = \tilde{z} \) just indifferent between giving a positive or a negative evaluation of her
supervisee’s performance. Given that \( z \) is uniformly distributed over \([0, 1]\) when \( z > 0 \), the equilibrium probability that the supervisor will commit a type-II violation is given by \( \varepsilon^* = \int_{\tilde{z}} g(z) dz = 1 - \tilde{z} \). Substituting \( \tilde{z} \) in this expression gives the value of \( \varepsilon^* \) reported in Proposition 1. Since in equilibrium it holds that \( \varepsilon^* = \delta^* \), the proof of the third part of Proposition 2 is identical to that of Proposition 1.

\[ \tilde{z} \]

\textbf{Proof of Proposition 3}

\textbf{Altruistic supervisor, altruistic worker}

Consider the case where the supervisor is altruistic and \( z > 0 \). In this case, define \( \hat{\beta} = \frac{e - pq(w + s)}{pq(z)^2} \) as the value of \( \beta \) that satisfies Eq. (3) as an equality. This is the level of altruism (or spite) that makes the worker with \( \beta = \hat{\beta} \) just indifferent between providing high or low effort. When the worker who is overseen by the supervisor develops altruistic feelings, recall that \( \beta \) is uniformly distributed over \([0, 1]\). Hence, to have an interior solution, we must have that:

\[ e - pq(w + s) > 0 \quad (8) \]

Provided that condition (8) is satisfied, the equilibrium probability \( \gamma^* \) that the worker provides high effort when \( z > 0 \) and \( \beta > 0 \) is given by \( \int_{\hat{\beta}}^1 g(\beta) d\beta = 1 - \hat{\beta} \). Substituting \( \hat{\beta} \) in this expression gives the value of \( \gamma^* \) reported in Proposition 3.

\textbf{Altruistic supervisor, spiteful worker}

When the worker who is overseen by the supervisor develops spiteful feelings, recall that \( \beta \) is uniformly distributed over \([-1, 0]\). Hence, to have an interior solution, we must have that:

\[ e - pq(w + s) < 0 \quad (9) \]

Provided that condition (9) is satisfied, the equilibrium probability \( \gamma^* \) that the worker provides high effort when \( z > 0 \) and \( \beta < 0 \) is given by \( \int_{\hat{\beta}}^0 k(\beta) d\beta = -\hat{\beta} \). Substituting \( \hat{\beta} \) in this expression gives the value of \( \gamma^* \) reported in Proposition 3.

\textbf{Spiteful supervisor, altruistic worker}

Now consider the case where the supervisor is spiteful and \( z < 0 \). In this case, define \( \tilde{\beta} = -\frac{e - pq(w + s)}{pq(z)^2} \) as the value of \( \beta \) that satisfies Eq. (4) as an equality. This is the level of altruism (or spite) that makes the worker with \( \beta = \tilde{\beta} \) just indifferent between
providing high or low effort. When the worker who is overseen by the supervisor develops altruistic feelings, recall that $\beta$ is uniformly distributed over $[0, 1]$. Hence, to have an interior solution, condition (5) must be satisfied. Provided that this is the case, the equilibrium probability $\gamma^*$ that the worker provides high effort when $\alpha < 0$ and $\beta > 0$ is given by $\int_0^\beta k(\beta)d\beta = \tilde{\beta}$. Substituting $\tilde{\beta}$ in this expression gives the value of $\gamma^*$ reported in Proposition 3.

**Spiteful supervisor, spiteful worker**

When the worker who is overseen by the supervisor develops spiteful feelings recall that $\beta$ is uniformly distributed over $[-1, 0]$. Hence, to have an interior solution, condition (4) must be satisfied. Provided that this is the case, the equilibrium probability $\gamma^*$ that the worker provides high effort when $\alpha < 0$ and $\beta < 0$ is given by $\int_{-1}^\beta k(\beta)d\beta = 1 + \tilde{\beta}$. Substituting $\tilde{\beta}$ in this expression gives the value of $\gamma^*$ reported in Proposition 3.

**Comparative statics**

The exact expressions of the comparative statics reported in the last part of Proposition 3 are given by:

$$\gamma_e = -\frac{1}{p\varphi qs}$$

$$\gamma_p = \frac{e}{p^2\varphi qs}$$

$$\gamma_q = \frac{e}{p\varphi q^2s} - \frac{1}{\varphi_q}\left[\frac{e - q(w + s)}{pq}\right]$$

$$\gamma_s = \frac{e - qw}{p\varphi qs^2} - \frac{1}{\varphi_s}\left[\frac{e - q(w + s)}{pq}\right]$$

$$\gamma_w = \frac{1}{p\varphi s} - \frac{1}{\varphi_w}\left[\frac{e - q(w + s)}{pq}\right]$$

where $\varphi = \delta^*$ if $\alpha > 0$ and $\varphi = \epsilon^*$ if $\alpha < 0$ – recall that in equilibrium, $\epsilon^* = \delta^*$. While it is clear that $\gamma_e < 0$ and $\gamma_p > 0$ are always satisfied, the signs of the other comparative statics depends on which between conditions (8) or (9) hold. To start, recall that from Provision 1 and 2 we already know that $\varphi_q < 0$, $\varphi_s < 0$ and $\varphi_w > 0$. In addition, observe that condition (4) must be satisfied to have an interior solution when workers and supervisors have reciprocal relational feelings. Hence, regardless of whether $\alpha > 0$ and $\beta > 0$ or $\alpha < 0$ and $\beta < 0$, we have that $\gamma_q > 0$, $\gamma_s > 0$ and
Finally, observe that that condition (5) must be satisfied to have an interior solution when workers and supervisors have reciprocal relational feelings. Hence, regardless of whether $\alpha > 0$ and $\beta < 0$ or $\alpha > 0$ and $\beta < 0$, we have that $\gamma_q \geq 0$, $\gamma_s \geq 0$ and $\gamma_w < 0$.

**Proof of Lemma 1**

**Altruistic supervisors**

Define $c - k = \Delta$. Hence, the exact expression of $\Delta$, $\Delta_w$ and $\Delta_s$ when $\alpha > 0$ is given by:

$$
\Delta = w - \varepsilon^*(1 - q)w - qs
$$

$$
\Delta_w = [1 - \varepsilon(1 - q)] + \varepsilon^*_w((1 - q)w - qs)
$$

$$
\Delta_s = \varepsilon q - \varepsilon^*_s[(1 - q)w - qs]
$$

According to Assumption 1, $(1 - q)w - qs > 0$. Hence, we have that $\Delta > 0$ if $\varepsilon^* < w/[(1 - q)w - qs]$, which is always satisfied according to Assumption 1. In addition, from Proposition 2, we already know that $\varepsilon^*_w > 0$ and $\varepsilon^*_s < 0$. Given this, it is straightforward to check that $\Delta_w > 0$ and $\Delta_s > 0$ are always satisfied when $\alpha > 0$.

**Spiteful supervisors**

The exact expression of $\Delta$, $\Delta_w$ and $\Delta_s$ when $\alpha < 0$ is given by:

$$
\Delta = w + \delta^*[(1 - q)w + qs]
$$

$$
\Delta_w = -\delta(1 - q) - \delta^*_w[(1 - q)w + qs]
$$

$$
\Delta_s = -q\delta - \delta^*_s[(1 - q)w + qs]
$$

According to Assumption 1, $(1 - q)w - qs > 0$. Hence, $\Delta > 0$ is always satisfied. From Proposition 1, we already know that $\delta^*_w > 0$ and $\delta^*_s < 0$. Given this, it is straightforward to check that $\Delta_w < 0$ and $\Delta_s \geq 0$ when $\alpha < 0$.

---

**Author’s contribution** All authors have provided equal contribution to the development of the paper, from its initial conception to its final drafting.

**Funding** Open access funding provided by Università degli Studi di Torino within the CRUI-CARE Agreement. This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.
Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Aktas E, Cicek I, Kiyak M (2011) The effect of organizational culture on organizational efficiency: the moderating role of organizational environment and CEO values. Procedia Soc Behav Sci 24:1560–1573
Antoci A, Sacco PL (1995) A public contracting evolutionary game with corruption. J Econ 61(2):89–122
Artz B, Goodall AH, Oswald AJ (2020) How common are bad bosses? Ind Relat 59(1):3–39
Autor DH (2003) Efficiency wage, Neoclassical and non-neoclassical evidence. mimeo, MIT.
Bac M, Kucuksenel S (2006) Two types of collusion in a model of hierarchical agency. J Inst Theoretical Econ 162(2):262–276
Baker G, Gibbons R, Murphy KJ (1994) Subjective performance measures in optimal incentive contracts. Quart J Econ 109:1125–1156
Bandiera O, Barankay I, Rasu I (2010) Social Incentives in the Workplace. Rev Econ Stud 77(2):417–558
Barnby T, Eberth B, Ma A (2012) Incentives, learning, task difficulty, and the Peter Principle: Interpreting individual output changes in an organisational hierarchy. Labour Econ 19(1):76–81
Berman EM, West JP, Richter MN (2002) Workplace relations: friendship patterns and consequences (According to Managers). Public Adm Rev 62:217–230
Brown M, Falk A, Fehr E (2004) Relational contracts and the nature of market interactions. Econometrica 72:747–780
Brown M, Falk A, Fehr E (2012) Competition and relational contracts: the role of unemployment as a disciplinary device. J Eur Econ Assoc 10:887–907
Bull C (1987) The existence of self-enforcing implicit contracts. Quart J Econ 102:147–159
Camerer C, Vepsalainen A (1988) The economic efficiency of corporate culture. Strateg Manag J 9:115–126
Campbell CM III, Kamlani KS (1997) The reasons for wage rigidity: evidence from a survey of firms. Quart J Econ 112:759–789
Cappelli P, Chauvin K (1991) An interplant test of the efficiency wages hypothesis. Quart J Econ 106:769–787
Cartwright E, Xue L, Brown C (2021) Are people willing to Tell Pareto White Lies? a review and new experimental evidence. Games 12:1–23
Casadesus-Masanell R (2004) Trust in agency. J Econ Manag Strategy 13:375–404
Chang JJ, Lai CC (1999) Carrots or sticks? A social custom viewpoint on worker effort. Eur J Polit Econ 15:297–310
Chang JJ, Lai CC (2002) Is the efficiency wage efficient? the social norm and organizational corruption. Scandinavian J Econ 104(1):27–47
Cordero Salas P, Roe B (2012) The role of cooperation and reciprocity in structuring carbon sequestration contracts in developing countries. Am J Agr Econ 94:411–418
Corneo G, Jeann O (1997) Snobs, bandwagons, and the origin of social customs in consumer behavior. J Econ Behav Organ 32:333–348
Cortina LM, Magley VJ, Williams JH, Langhout RD (2001) Incivility in the workplace: incidence and impact. J Occup Health Psychol 6(1):64–80
Supervise me if you can. Relational feelings, incentive pays and...

Crapanzano R, Mitchell M (2005) Social Exchange theory: an interdisciplinary Review. Journal of Management 31(6), 874–900.

Dur R. (2009) Gift Exchange in the Workplace: Money or Attention? J Eur Econ Assoc 7:550–560

Dur R, Glazer A (2008) Optimal contracts when a worker envies his boss. J Law Econ Organ 24:120–137

Dur R, Tichem J (2015) Altruism and Relational Incentives in the Workplace. J Econ Manag Strategy 24(3):485–500

Falk A, Fehr E, Fischbacher U (2005) driving forces behind informal sanctions. Econometrica 73(6):2017–2030

Frey BS (1993) Does monitoring increase work effort? the rivalry with trust and loyalty. Econ Inq 31(4):663–670

Gätcher S, Herrmann B (2006) The limits of self-governance in the presence of spite: experimentalevidence from Urban and Rural Russia. Iza Discussion Paper n 2236

Gibbons R (1998) Incentives in organizations. J Econ Perspectives 12:115–132

Gibbs M (2013) Design and Implementation of Pay for Performance in Thomas. In: WF, (ed) CR and Shughart II. Oxford Handbook in Managerial Economics. Oxford University Press, Oxford, UK, pp 397–423

Gibbs M, Merchant KA, Van der Stede WA, Vargus ME (2004) Determinants and effects of subjectivity in incentives. Account Rev 79:409–436

Griffeth RW, Riordan CM (1995) The opportunity for friendship in the workplace: an undernested construct. J Bus Psychol 10(2):141–154

Groshen EL, Krueger AB (1990) The structure of supervision and pay in hospitals. Ind Labour Relation Rev 43:S134–S146

Harper J, Adkins A (2015) Employees want more from their managers. Gallup Workplace. www.gallup.com/workplace/236570/employees-lot-managers.aspx.

Herrmann B, Christian Thöni C, Gächter S (2008) Antisocial punishment across societies. Science 319(7):1362–1367

Kragl J, Schmid J (2009) The impact of envy on relational incentive contracts. J Econo Behav 72(2):766–779

Kucuksenel S, Saygili K (2019) Other-regarding preferences in organizational hierarchies. J Econ 126:201–2019

Laffont JJ, Martimort D (1997) the firm as a multicontract organization. J Econ Manag Strategy 6:201–234

Lazear EP (1989) Pay equality and industrial politics. J Polit Econ 97:561–580

Lazear EP (2004) The pete principle: a theory of decline. J Polit Econ 112(1):S141–S163

Lee S, Persson P (2011) Authority versus loyalty: social incentives and governance. NYU Working Paper, No. FIN-10–001.

Leonard J (1987) Carrots and sticks: pay, supervision and turnover. J Law Econ 5:S136–S152

Levin J (2003) Relational incentive contracts. Am Econ Rev 97:835–857

MacLeod WB, Malcomson JM (1989) Implicit contracts, incentive compatibility, and involuntary unemployment. Econometrica 57:447–480

MacLeod WB, Parent D (2000) Job characteristics and the form of compensation. Res Labor Econ 18:177–242

Manning A, Thomas J (1997) A simple test of the shirking model. Centre of Economic Performance Discussion Paper no. 374.

Marshall A (1980) Principles of economics, Palgrave classics. Palgrave Macmillan, London

Millward NM, Stevens D (1992) Workplace industrial relations in transition: The ED/ESRC/PSI/ACAS Surveys. Dartmouth, Aldershot

Mintzberg H (1971) Managerial work: analysis from observation. Manage Sci 18(2):97–110

Moerbeek HHS, Need A (2003) Enemies at work: can they hinder your career? Social Networks 25:67–82

Morrison RL (2008) Negative relationships in the workplace: Associations with organisational commitment, cohesion, job satisfaction and intention to turnover. J Manag Organ 14:330–344

Nickell S, Nicolitas D (1997) Wages, restrictive practices and productivity. Labour Econ 4:201–221

Non A (2012) Gift-exchange, incentives, and heterogeneous workers. Games Econom Behav 75:319–336

Norris AL, Marcus DK, Mercer SH, Ziegler-Hill V (2014) The psychology of spite and the measurement of spitefulness. Psychol Assess 26(2):563–574

Pathiranage Y, Lakhmini J, Abeysekera R (2020) A literature review on organizational culture towards corporate performance. J Manag Account Econ 7:522–544

Perri TJ (1994) Influence activity and executive compensation. J Econ Behav Organ 24:169–181
Peter LJ, Hull, R (1969) The peter principle, William Morrow & Co Inc. (Pan Books edition 1970).
Prendergast C (1999) The provision of incentives in firms. J Econ Literature 37:7–63
Prendergast C (2007) The Motivation and Bias of Bureaucrats. American Economic Review 97(1):180–196
Prendergast C, Topel RH (1996) Favoritism in organizations. J Polit Econ 104:958–978
Rhoades L, Eisenberger R (2002) Perceived organizational support: a review of the literature. J Appl Psychol 87:698–714
Shapiro C, Stiglitz JE (1984) Equilibrium unemployment as a worker discipline device. Am Econ Rev 74(3):433–444
Shchetinin, O (2010) Altruism and career concerns. Gothenborg University. Working Papers in Economics 427.
Sliwka D (2007) Trust as a signal of a social norm and the hidden costs of incentive schemes. Am Econo Rev 97:999–1012
Southard A (2017) Spitefulness. In: Zeigler-Hill V, Shackelford TK (eds) Encyclopedia of Personality and Individual Differences. Springer, Cham. https://doi.org/10.1007/978-3-319-28099-8_1270-1
Tirole J (1986) Hierarchies and bureaucracies: on the role of collusion in organizations. J Law Econ Organ 2(2):181–214
Wagner R, Harter JK (2006). 12: The elements of great managing. Gallup Press.
Williamson O (1967) Hierarchical control and optimum firm size. J Polit Econ 75(2):123–138

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.