Optimal Contracts for Lenient Supervisors[1]

Thomas Giebe[2]
Humboldt University at Berlin

Oliver Gürtler[3]
University of Cologne

September 12, 2011

[1]We would like to thank two anonymous referees as well as Patrick Bolton, Florian Englmaier, Guido Friebel, Veronika Grimm, Karl Ludwig Keiber, Nicolas Klein, Johannes Koenen, Matthias Kräkel, Sebastian Kranz, Tymofiy Mylovanov, Judith Przemeck, Andreas Roider, Ferdinand von Siemens, Dirk Sliwka, Roland Strausz, Achim Wambach and Elmar Wolfstetter for helpful comments. Financial support by the Deutsche Forschungsgemeinschaft (DFG), grants SFB/TR 15 ("Governance and the Efficiency of Economic Systems"), is gratefully acknowledged.

[2]Thomas Giebe, Dept. of Economics, Economic Theory I, Humboldt University at Berlin, Spandauer Str. 1, D-10099 Berlin, Germany, e-mail: thomas.giebe@wiwi.hu-berlin.de, phone: +49-30-20935773, fax: +49-30-20935619.

[3](Corresponding author) Oliver Gürtler, University of Cologne, Dept. of Economics, Albertus-Magnus-Platz, D-50923 Cologne, Germany, email: oliver.guertler@uni-koeln.de, phone: +49-221-4701450, fax: +49-221-4701816.

Abstract
We analyze optimal contracts in a hierarchy consisting of a principal, a supervisor and an agent. The supervisor is either neutral or altruistic towards the agent, but his preferences are private information. In a model with two supervisor types, we find that the optimal contract may be very simple, paying the supervisor a flat wage independent of his type and his evaluation of the agent's effort. Such a contract induces the neutral type of supervisor to report the agent's performance truthfully, while the altruistic type reports favorably independent of performance. Accordingly, overstated performance (leniency bias) may be the outcome of an optimal contract under informational asymmetries.

Keywords: Subjective performance evaluation, leniency, supervisor, private information
JEL classification: D82, D86, J33, M52

1 Introduction

Many firms use subjective measures to assess and reward their employees' performance.[1] This gives discretion to those who are assigned the task of performance evaluation. Several studies indicate that supervisors seem to exploit the power given to them. In particular, leniency bias — a practice of systematically overstating employee performance — is a regular phenomenon in subjective performance evaluation.[2] Fortune 200 company Arrow Electronics, which has introduced a subjective performance evaluation scheme in 1995, serves as an example to illustrate this point.[3] Supervisors at Arrow Electronics had to evaluate their subordinates in seven performance areas on a scale of 1 to 5 (with 5 as the highest rating). When employees were evaluated for the first time, *not a single employee* received a rating of 1 or 2 in *any* performance area. The average rating was 4.5.

One may argue that the observation of high ratings alone is not evidence of leniency, but rather of an effective incentive system which induces employees to work hard and to perform well. Three observations contrast this argument. First, firms seem to be dissatisfied with the outcome of performance appraisals and seek ways to lower the ratings. In the case of Arrow Electronics it seems to have been obvious to the participants that ratings were lenient. Indeed, former CEO Steve Kaufman decided to have all employees rated again, while giving the supervisors instructions on what the general distribution of ratings should be.[4] Second, field studies show that there are systematic differences in evaluations when employees are evaluated by different supervisors, with some supervisors constantly giving better ratings than others.[5] In combination with the observation of generally good ratings, this suggests that at least some of the supervisors decide to overstate performance. Third, it is sometimes observed that supervisors give the highest ratings to those employees that they work closely with. This indicates that supervisors inflate performance ratings if they have a personal relationship

[1] Smith et al. (1996) present results from a survey of mid-western US companies. About 92 percent of the respondents state that performance appraisal reviews and feedback sessions are required in their companies. MacLeod and Parent (1999) report that about 20 to 30 percent of all US workers receive some form of subjectively determined reward every year.

[2] See e.g. Medoff and Abraham (1980) and the excellent survey by Prendergast (1999).

[3] For details regarding the performance evaluation scheme at Arrow Electronics see Hall and Madigan (2000).

[4] As reported by Hall and Madigan (2000), Kaufman said: "The board of directors gave me a 2 on *two* of my seven rankings, and if I get two 2s, sure as hell someone, somewhere in the company should get at least one."

[5] See, for example, Kane et al. (1995).

with their subordinates.[6]
While there is consensus that leniency in performance evaluation is generally relevant, there are different views on the consequences of leniency for firms. Typically, leniency is argued to be detrimental to firms. This is because leniency reduces the effect of an employee's performance on his rating and, therefore, incentives to put forth effort to affect performance. There is, however, a more positive view of leniency too. Negative ratings are sometimes argued to lead to demotivation of employees or to fairness concerns. Obviously, these problems are avoided if supervisors are lenient and give only positive ratings to their subordinates.[7]
In the current paper, we follow the first view and focus on a situation where leniency is detrimental to firms, by weakening the link between pay and performance. Our aim is to analyze whether a supervisor's compensation can be structured such as to mitigate the problems caused by leniency. In particular, we want to answer the following interrelated questions. Can the compensation of a supervisor be designed such that leniency is eliminated, i.e., such that the supervisor reports performance truthfully? If so, is it always in the firm's interest to eliminate leniency?
In order to answer these questions, we consider a model with a three–tiered hierarchy consisting of a principal, a supervisor, and an agent. The production of output requires the agent's effort. In order to motivate the agent, the principal hires a supervisor whose task it is to monitor and report the agent's performance. The supervisor is assumed to be either altruistic or neutral towards the agent and, hence, may be lenient.
We distinguish different settings where the supervisor's preference for the agent (his type) might be private or public information and where the parties might be wealth–constrained.[8] We derive optimal contracts for these settings and investigate whether leniency prevails if contracts are structured optimally.
We find that, if the supervisor's type is public information, the optimal contract eliminates leniency by punishing the supervisor for good ratings of the agent. The first-best solution can be implemented regardless of wealth constraints. Under private information, it is feasible to eliminate leniency by

[6]Hall and Madigan (2000) call this effect the "Wilson factor", named after a supervisor at Arrow Electronics who consistently displayed the kind of behavior just described.

[7]See e.g. a recent study by Bol (2011) that analyzes compensation data from a Dutch financial service provider and finds that leniency is positively associated with agents' future performance.

[8]Most of the literature on alternative preferences assumes common knowledge of the parties' preferences and so does not consider the problem we address. See, however, von Siemens (2011a,b) on the screening of inequity–averse agents.

offering an incentive–compatible menu of contracts inducing correct evaluation of the agent. Due to excessive rents, however, it can be optimal to have very simple contracts, paying the supervisor a flat wage independent of his type and how he evaluates the agent.[9] This induces a neutral supervisor to always report the agent's performance truthfully, while an altruistic supervisor reports favorably independent of performance. This implies that leniency is consistent with optimal contractual arrangements under informational asymmetries. Under unlimited liability, the first–best solution can be implemented even though leniency occurs.
The paper is organized as follows: In the next section we relate the paper to the literature. In Section 3 we present the model and the first–best solution. Section 4 looks at the public information case, while Section 5 deals with private information. Section 6 gives a discussion and Section 7 concludes. All proofs are in the Appendix.

2 Related Literature

The paper is closely related to the literature on biases in subjective performance evaluation, e.g., Baker et al. (1988), Murphy (1992), Harris (1994), Prendergast and Topel (1996), Prendergast (2002), MacLeod (2003), or Grund and Przemeck (2012). Particularly relevant for the present study are the papers by MacLeod (2003) and Grund and Przemeck (2012). MacLeod (2003) considers a principal-agent model and assumes that both parties receive subjective estimates of the agent's performance. If there is no correlation between the estimates of principal and agent, it is found that the optimal contract pays the agent a bonus unless the principal observes the lowest possible level of performance. MacLeod (2003) argues that this finding provides a possible explanation for the observation of lenient ratings. Similar to MacLeod (2003), we analyze optimal contracts. However, the main differences are that we consider a principal-supervisor-agent hierarchy and that leniency arises from the supervisor's preferences. We find that both features are consistent with the empirical evidence discussed in the Introduction. In the model of Grund and Przemeck (2012), there is a supervisor who has to rate two agents who receive pay depending on their own performance rating. Agents do not like to receive unequal payments (they are inequality-averse) and the supervisor wants the agents to obtain high utility. This implies that he has to rate them well, and not too differently. A rationale for leniency

[9]In our model, the supervisor's only task is to monitor the agent. The term "flat wage" indicates that the supervisor's wage does not depend on how he rates the agent. Of course, a supervisor might have other tasks with variable pay.

bias and centrality bias is thus given.[10] Grund and Przemeck (2012) and the present paper have in common that leniency is due to the supervisor's preferences. In contrast to our paper, Grund and Przemeck (2012) take contracts as exogenously given and do not analyze the principal's decision problem. Moreover, they assume common knowledge of preferences, while the private information setting is at the core of our analysis. Summarizing, our paper differs from MacLeod (2003) and Grund and Przemeck (2012) in that we combine altruistic preferences with a contract-theoretical analysis, asking if leniency can be, resp. should be, mitigated through contract design. Thus, the present paper extends our knowledge of leniency in several ways. It finds that leniency can always be eliminated by means of incentive contracts if the supervisor is lenient because he cares for the agent's well-being (regardless of whether his preferences are private information). The paper also finds that it can be optimal not to eliminate leniency.

The paper is also related to the literature on optimal contracts in three-tiered hierarchies consisting of a principal, a supervisor and an agent, see, e.g., Tirole (1986), Strausz (1997) or Vafai (2004). These papers differ from ours in two respects. First, they typically assume hard information, i.e., a supervisor may conceal but not misrepresent information. In our model, information is soft.[11] Second, in those models the supervisor has publicly known standard (selfish) preferences. Our model, in contrast, relaxes the standard selfish preferences assumption. This makes the principal's problem more complicated since she has to take into account the different types' incentive constraints when designing optimal contracts.

A paper that is similar to ours in that it focuses on non-standard (but publicly known) preferences in a principal-supervisor-agent relationship is the paper by Lee and Persson (2010). They assume that supervisor and agent are altruistic towards each other. The agent chooses some action which affects his performance for the principal. The supervisor wants the agent to perform well because he gets some credit for this. In contrast to our paper, the supervisor does not evaluate the agent, but may monitor and possibly overturn the agent's action. The authors find that the equilibrium crucially depends on how altruistic the players are. If altruism is high, the supervisor will not monitor the agent. Still, the agent tries to perform well in order to make sure that the supervisor is credited for good performance. If altruism is low, these implicit incentives do not work and the supervisor decides to actively monitor the agent.

[10]Centrality bias refers to the practice of not sufficiently differentiating between agents' performance when performance is evaluated subjectively.

[11]See Laffont and Martimort (1997) and Faure-Grimaud et al. (2003) for models with soft information.

Finally, the paper is related to literature combining adverse selection and moral hazard, such as Laffont and Tirole (1986), McAfee and McMillan (1987) or Lewis and Sappington (1997). In those papers, the player who chooses an unobservable action has superior information. In contrast, our supervisor has private information (his type) and a hidden action (to report effort truthfully or not) while the agent's action is hidden from the principal but not from the supervisor.

Summarizing, our contribution can be clearly distinguished from the existing literature. We consider a principal-supervisor-agent hierarchy with soft performance information. We take the supervisor's tendency for leniency as given and analyze its consequences for the design of optimal contracts. We analyze the model under public and private information about preferences.

3 The model

A principal, P, a supervisor, S, and an agent, A, play the following basic game. At stage 1, P offers a menu of contracts simultaneously to A and S. A contract obliges A to exert effort and S's task is to monitor and report A's effort. The contract specifies an effort level requested from A as well as wages for A and S that (may) depend on reported performance. At stage 2, either both A and S simultaneously sign the contract(s) or the game ends. At stage 3, A chooses effort and at stage 4, S observes and reports A's effort to P. Then payments are made according to the contract and the game ends.

The cost of A's effort, $e \geq 0$, is $c(e)$. It is twice differentiable and satisfies $c(0) = 0$, $c'(0) = 0$, $c'(e) > 0$ for $e > 0$, $c'(e) = \infty$ for $e \to \infty$, and $c''(e) > 0$. Effort produces an output, $f(e)$, that accrues to P, where $f(0) = 0$, $f'(e) > 0$ and $f''(e) \leq 0$.

Effort and output are unobservable for P and not verifiable by third parties. Moreover, there is no other objective measure of e. Hence, P cannot write an explicit performance contract to motivate A. Thus, it is S's task to monitor A and to observe and report the agent's effort choice.[12]

The contract obliges S to send a report $m \in \{y, n\}$ to P where $m = y$ indicates that A has chosen (at least) the requested effort while $m = n$ means underperformance. We assume this report to be in written form.[13] This implies that third parties are able to learn the content of the supervisor's report by looking at his written statement so that the content of reports

[12] Of course, S may have other tasks as well.

[13] In practice, subjective performance evaluations are typically put down in writing. Supervisors at Arrow Electronics, for instance, had to fill out an evaluation sheet for each employee they had to rate.

is verifiable. We further assume that S observes A's effort perfectly and costlessly.[14]

All parties are risk-neutral and have zero reservation value. A central assumption is that S may care for A's well-being and thus has an incentive to distort the effort report. We model S's preferences as[15]

$$U_S = w_S + \hat{\lambda} U_A, \tag{1}$$

where w_S denotes the income of S and U_A is A's utility. The parameter $\hat{\lambda} \in [0, 1)$ measures how strongly S cares for A's well-being. We assume that S reports honestly if he is indifferent.[16]

A's utility is given by

$$U_A = w_A - c(e), \tag{2}$$

with w_A as A's income.

Except for Section 4, we assume that S's preference, or type, is private information, i.e., a supervisor of type i, denoted by $\mathrm{S_i}$, knows that his preference parameter is $\hat{\lambda} = \lambda_i$, while for P and A, $\hat{\lambda}$ is a random variable.

Throughout the paper we assume that there are two S–types. Supervisor $\mathrm{S_1}$ has standard (selfish) preferences, $\lambda_1 = 0$. We call him *neutral.* The other type, $\mathrm{S_2}$, is *altruistic*, with $\lambda_2 \in (0, 1)$. $\mathrm{S_1}$ (resp. $\mathrm{S_2}$) occurs with probability q (resp. $1 - q$).

We confine analysis to pure strategies and (menus of) incentive–compatible contracts where S self–selects into the contract designed for his type and A makes the demanded effort. A contract designed for $\mathrm{S_i}$ is of the form $\{a_i, b_i, e_i, w_{yi}, w_{ni}\}$. It specifies the requested (minimum) effort level, e_i, $\mathrm{S_i}$'s wages $\{w_{yi}, w_{ni}\}$ conditional on report $m \in \{y, n\}$. Moreover, A receives a base wage a_i, and, if S reports $m = y$, a bonus b_i.

P offers a *menu* of contracts and, at stage 2, A and S simultaneously sign this *menu.* Then P may choose to conceal from A which particular *contract* from the menu is selected by S. In that case, A cannot infer S's type but he knows each type's incentives. For that case, we assume that it is ex-post verifiable, which particular contract has been selected by S. Note that, in principle, A's wage might depend on S's type also when the true type is concealed from A but then, obviously, the demanded effort must be equal for both types.

[14]As an example, consider a supervisor and an agent who share an office, possibly working on the same project. Here, S might observe A's effort without additional effort or cost.

[15]See Rotemberg and Saloner (1993), Prendergast and Topel (1996), Prendergast (2002), Sliwka (2007) or Rotemberg (2009, 2011) for similar specifications.

[16]Note that this assumption is not needed for the existence of our equilibria but it ensures uniqueness.

We call S *lenient* if he reports favorably, $m = y$, independent of A's performance.
As a benchmark, consider the first-best solution assuming that effort is contractible and no contractual frictions arise. Here, P does not need S.[17] She simply imposes the effort level that solves

$$\max_{e,w_A} f(e) - w_A \qquad s.t. \;\; w_A - c(e) \geq 0. \tag{3}$$

Hence, the first-best effort, e_{FB}, is implicitly given by

$$f'(e_{FB}) = c'(e_{FB}). \tag{4}$$

We denote $f_{FB} := f(e_{FB})$ and $c_{FB} := c(e_{FB})$.

4 Complete Information

As another benchmark case, suppose S's type $\hat{\lambda}$ is common knowledge.[18] The optimal contracts solve the following program:

$$\begin{aligned}
&\max_{a,b,w_y,w_n,e} && f(e) - a - b - w_y \\
s.t.\;\; &(IC_A) && a + b - c(e) \geq a \\
&(IC_{Sn}) && w_n + \hat{\lambda}(a - c(\tilde{e})) \geq w_y + \hat{\lambda}(a + b - c(\tilde{e})),\ \forall\, \tilde{e} < e \\
&(IC_{Sy}) && w_y + \hat{\lambda}(a + b - c(\bar{e})) \geq w_n + \hat{\lambda}(a - c(\bar{e})),\ \forall\, \bar{e} \geq e \\
&(IR_A) && a + b - c(e) \geq 0 \\
&(IR_S) && w_y + \hat{\lambda}(a + b - c(e)) \geq 0
\end{aligned} \tag{5}$$

Supposing that it is optimal to induce positive effort, optimality requires that S reports truthfully.[19] The objective of the above program maximizes P's profit, i.e., output minus wage payments, given that S reports truthfully and A prefers positive effort. This implies that S earns the wage w_y while A earns a bonus, receiving $a+b$. The equilibrium is supported by the following constraints. The constraint (IC_A) induces the requested effort. If the agent expects the supervisor to report performance truthfully, two effort levels may be optimal for him, either the requested one or an effort of zero. The

[17] Note that P cannot do better by leaving a rent r to A and hiring a supervisor at a negative wage $w_S = -\lambda r$. This is due to $\lambda < 1$.

[18] We drop the subscript i.

[19] Intuitively, if S lies with positive probability, A's incentives are weakened and implementing a given effort becomes more expensive or impossible.

constraint ensures that the payoff from exerting the requested effort level and getting the bonus exceeds the payoff from zero effort. (IC_{Sy}) and (IC_{Sn}) make S report A's effort truthfully. By reporting $m = y$ instead of $m = n$, the supervisor changes his wage from w_n to w_y, while the agent's payoff is increased by the bonus payment, which gives S an additional utility of $\hat{\lambda}b$. The two constraints guarantee that S prefers to choose $m = y$ if the agent has provided (at least) the requested effort level and $m = n$ otherwise. In fact, (IC_{Sy}) and (IC_{Sn}) together imply that S must be indifferent between reports. Finally (IR_A) and (IR_S) ensure participation, i.e., a nonnegative equilibrium utility.

Proposition 1. *The optimal contract under complete information is* $a^* = 0$*,* $b^* = c_{FB}$*,* $e^* = e_{FB}$*,* $w_y^* = 0$*,* $w_n^* = \hat{\lambda}c_{FB}$*. It implements the first-best solution and eliminates leniency.*

The optimal contract extracts all rents from S and A. A is just compensated for effort cost, while S gets zero wage (and zero payoff, due to A's zero utility). If $\hat{\lambda} > 0$, the optimal contract has $w_n^* > w_y^* = 0$, i.e., it rewards an altruistic S for giving a truthful negative performance report. This is necessary since a positive report implies that A gets the bonus b^* in which S participates with utility $\hat{\lambda}b^*$. Thus, a truthful negative report requires that S is compensated for his "loss" due to the unpaid bonus to A. However, in equilibrium, A exerts effort and S reports truthfully, implying that w_n^* is not payoff-relevant. Truthful reporting makes effort basically verifiable and, since rents to A and S can be avoided, this allows P to implement the first–best solution. Wages are nonnegative which makes the contract feasible even under wealth constraints.

5 Incomplete Information

Suppose the distribution of S–types is common knowledge while the realization λ_i is $\mathrm{S_i}$'s private information. Following standard contract theory, we distinguish whether there are restrictions on feasible wages or not.[20] We analyze unlimited liability in 5.1 and limited liability in 5.2. Similar to standard results, this distinction has important consequences for the design of optimal contracts in our paper.

Limited liability refers to a situation where wages must be nonnegative. Under unlimited liability, instead, wages are not restricted. A lower wage bound might have different reasons. S and A could be wealth-constrained and,

[20] See, e.g., Laffont and Martimort (2001).

thus, simply unable to make a payment to P. Alternatively, there could be minimum-wage regulations that rule out negative wages. Under both interpretations it seems that there are jobs, for which the assumption of unlimited liability is more appropriate. Managerial jobs are less likely to be affected by wealth-constraints and minimum-wage regulations. Here, it may be reasonable to assume unrestricted wages. In contrast, lower-level employees are typically subject to limited liability.

5.1 Unlimited liability

Proposition 2. *If the players are unlimitedly liable, the optimal contract is unique, type–independent, and satisfies* $a^* = \frac{q-1}{q} c_{FB} < 0$, $b^* = \frac{c_{FB}}{q}$, $e^* = e_{FB}$, $w_y^* = w_n^* = 0$. *A does not learn S's type. The contract implements the first-best solution and the altruistic supervisor is lenient.*

As is well-known from contract theory, a contract that leaves rents to other players than the principal generally leads to an inefficient solution. If positive rents cannot be avoided, the principal faces a trade-off between efficiency and rent extraction and deviates from the first-best solution in order to reduce the ensuing rents.
By this argument, it is easy to understand that a contract is optimal for the principal if it induces the desired effort level from the agent and extracts all rents from agent and supervisor.
The contract described in Proposition 2 achieves both of these objectives. Essentially, it provides incentives only through the neutral supervisor. Since this type occurs with probability less than one, this requires stronger incentives via a larger bonus for A. Then, A's rent is extracted via a negative base wage.
In particular, the contract pays the same flat wage to both S–types. A flat wage induces S_1 to report truthfully, since his own wage makes him indifferent between reports and he does not care about A's wages. S_2 is lenient, i.e., reports favorably independent of A's effort. This is because his wage is not affected by his report, but the bonus payment to A increases his utility. A does not know which S–type he is facing. S_2's leniency does not provide effort incentives, since if S_2 occurs, A gets the bonus anyway. Thus, positive effort can only be induced via the truthfully reporting S_1. Since S_1 occurs with probability q, the *expected* bonus, qb, must cover the effort cost. Thus, efficient effort is achieved by setting $qb^* = c_{FB}$. But then, the *actual* bonus exceeds effort cost, $b^* > c_{FB}$. Thus, A's base wage is negative in order to extract these rents from him. Finally, since A receives his reservation value of zero, both types of supervisor receive the same payoff. Their flat wage is

chosen such that supervisor rents are extracted as well. Limited liability of S is unproblematic since S's flat wage is nonnegative.
An implication of Proposition 2 is that leniency is not necessarily detrimental to performance. In principle, the presence of a lenient supervisor lowers incentives by weakening the link between pay and performance. As we have shown, providing stronger incentives through the neutral supervisor combined with a negative base wage for A completely neutralizes the negative effect of leniency. The negative base wage, however, makes the contract infeasible if A is wealth-constrained.

5.2 Limited liability

Limited liability is defined as

$$a_i \geq 0,\ a_i + b_i \geq 0, w_{yi} \geq 0, w_{ni} \geq 0, \quad i \in \{1, 2\}. \tag{6}$$

Recall that we only consider incentive–compatible (menus of) contracts, i.e., supervisors self–select according to their types. While we assume that A knows the menu of contracts, P may choose to conceal from A which particular contract was chosen by S. Then A only knows each type's incentives.
It is easy to see that (IC_{Sy}) and (IC_{Sn}) (see (5)) hold simultaneously for S_i if and only if $w_{yi} + \lambda_i b_i = w_{ni}$. Thus, truthful reporting of effort requires that S_i is indifferent between reports $m \in \{y, n\}$. If S_i were not indifferent, he would prefer one of the reports in which case his report would be independent of effort. But then A's optimal effort is zero. Positive equilibrium effort therefore requires that at least one S-type reports truthfully.
Accordingly, we partition the potential (menus of) contracts into the following cases, depending on which S-type reports effort truthfully and whether the type is revealed to A.[21]

truthful report by	A does not learn S's type	A learns S's type
S_1 only	(A)	(D)
S_2 only	(B)	(E)
S_1 and S_2	(C)	(F)

In the following, we look at each of those cases in a separate lemma. For the moment, we only solve for the optimal contracts as a function of given requested effort level, ignoring the optimal level of effort. However, the optimal requested efforts are strictly positive with the exception of cases (D)

[21]More precisely, in the cases (D) to (F), the supervisor's contract choice is revealed to A. In equilibrium, this implies that A learns the supervisor's type.

and (E). There, A learns S's type and, by design, one type does not report effort truthfully. If A faces this type then he will not exert positive effort. For that S-type, the only feasible effort level is zero.
After each lemma, we compare the results with those under complete information in order to see how the optimal contracts are constrained by the combination of asymmetric information and limited liability (again, ignoring the optimal level of effort). Recall that the limited liability constraint is not binding under complete information, see Proposition 1.
In order to simplify notation in Lemmas 1 to 6 and their proof, we suppress $\lambda_1 = 0$ altogether and denote $\lambda := \lambda_2$, $f := f(e)$, $c := c(e)$, $f_i := f(e_i)$, $c_i := c(e_i)$, $i \in \{1, 2\}$.

Lemma 1. *Consider case (A). For given positive requested effort, the set of optimal contracts is characterized by*

$$\begin{aligned} S_1 : &\quad w_{y1} = w_{n1} = 0, a_1 = 0, b_1 = \frac{c}{q}, \\ S_2 : &\quad w_{y2} = w_{n2} = 0, a_2 \geq 0, a_2 + b_2 \geq 0, b_2 \neq 0, a_2 + \max\{0, b_2\} = \frac{c}{q}. \end{aligned} \tag{7}$$

The contract for S_1 is similar to the optimal complete-information contract with one exception: b_1 is larger than b^*. In order to see why, recall that only S_1 reports effort truthfully, while S_2's report is effort independent. A knows that effort does not pay if S_2 occurs. Therefore, only b_1 (not b_2) can incentivize A. The *expected* bonus b_1 must cover the effort cost, $qb_1 = c \iff b_1 = c/q$.
Since S_2 reports independently of effort, S_2's contract does not have to incentivize A. It only has to make sure that neither S_1 nor S_2 have an incentive to choose the other type's contract. First, S_2 gets zero wages (as well) in order to prevent S_1 from choosing S_2's contract. Second, S_2 must get a rent of $\lambda c/q$ in order to make his contract exactly as attractive as S_1's contract. This rent is paid indirectly via A's wages which, due to $\lambda_1 = 0$, is not attractive for S_1. Since S_2's contract can ignore A's incentives, there is some freedom in choosing a_2 and b_2, as long as limited liability is observed.

Lemma 2. *Consider case (B). For given positive requested effort, the set of* *optimal contracts is characterized by*

$$
\begin{aligned}
S_1: \quad & \left(a_1 \geq 0, b_1 = -a_1, w_{y1} = \frac{\lambda c}{1-q}, w_{n1} \in [0, w_{y1}), w_{n1} + \lambda a_1 \leq w_{y1}, \right) \\
& or \left(a_1 = 0, b_1 \geq 0, w_{y1} \in [0, w_{n1}), w_{y1} + \lambda b_1 \leq w_{n1}, w_{n1} = \frac{\lambda c}{1-q}\right), \\
S_2: \quad & w_{y2} = 0, w_{n2} = \frac{\lambda c}{1-q}, a_2 = 0, b_2 = \frac{c}{1-q}.
\end{aligned} \tag{8}
$$

We compare the truth-telling supervisor's contract (here S_2) with the optimal complete-information contract from Proposition 1 (again, ignoring the *level* of effort). Similar to case (A), the only difference is the size of the bonus (b_2 as compared to b^*) which has to be adjusted in the same way. Note that w_{n2} is structurally the same as under complete information: $w_{n2} = \lambda b_2$ as compared to $w_n^* = \hat{\lambda} b^*$.

Similar to case (A), there is some freedom in choosing the contract for the other supervisor (S_1) who reports independently of effort. Again, this contract must provide the same rent as S_2's contract, it can ignore A's incentives, and it must not be attractive for S_2. In addition, limited liability has to be observed. The difference to case (A) is that, here, the rents for S_1 must be provided *directly* through wage payments since S_1 does not participate in A's wages, i.e., either w_{n1} or w_{y1} must be equal to $w_{n2} = \lambda c/(1-q)$ (while the other wage must be lower, respectively, as dictated by case (B)).

Lemma 3. *Consider case (C). For given positive requested effort, the set of optimal contracts is characterized by*

$$
\begin{aligned}
& w_{y1} = w_{n1} = w_{n2} = \lambda b_2, w_{y2} = 0, a_1 = 0, a_2 = b_1, \\
& (b_1, b_2) = \begin{cases} \left(0, \frac{c}{1-q}\right) & \text{if } \lambda < \left(\frac{1-q}{q}\right)^2 \\ \left(\frac{c}{q}, 0\right) & \text{if } \lambda > \left(\frac{1-q}{q}\right)^2 \\ \left(t, \frac{c-qt}{1-q}\right), t \in \left[0, \frac{c}{q}\right], & \text{if } \lambda = \left(\frac{1-q}{q}\right)^2 \end{cases},
\end{aligned} \tag{9}
$$

S_2's wages are similar to those of the complete information case. As a consequence of incomplete information, S_1 will only choose the right contract if one offers him the same income, $w_{y1} = w_{n1} = \lambda b_2$, while his wages would be zero under complete information.

Since both types report effort truthfully, both contracts can be used to incentivize A. As it turns out, however, depending on the parameters λ and q,

one of the types is generally more efficient in providing incentives and one would set the other type's bonus equal to zero, as we explain in the following. Under incomplete information, each supervisor type must be prevented from choosing the other type's contract. This is complicated because offering type i a rent (directly through wages or indirectly through A's bonus) makes his contract generally more attractive for type j. If, as a response, one makes j's contract more attractive as well, then it might also become more attractive for type i, requiring a larger rent for i, and so on. This incentive structure is illustrated by the fact that under the optimal contracts, a_2 depends on b_1 and w_{y1} depends on b_2 (and all of those are payoff-relevant in equilibrium). Now, if S_1 is unlikely to occur (low q), then setting incentives using b_1 is not effective, since 1) A only cares about the *expected* bonus qb_1, which means we need a large bonus since q is small, and, 2) a larger b_1 requires a larger a_2 which has to be paid with high probability (large $1-q$). Similarly, if S_2 is unlikely (small $1-q$) then incentivizing A through b_2 is equally ineffective due to a low *expected* bonus, $(1-q)b_2$ and the fact that $w_{y1} = b_2$ has to be paid with high probability.
Therefore, generally, only one bonus will optimally be positive, while the bonus is then equal to those paid in cases (A) and (B), with a similar explanation, respectively.[22]

Lemma 4. *Consider case (D). Since types are revealed to A and A knows that* S_2 *reports independently of effort, we have* $e_2 = 0$*. For* $e_1 > 0$*, and for arbitrarily small* $\epsilon > 0$*, the set of optimal contracts is characterized by*

$$\begin{aligned} & w_{y1} = w_{n1} = w_{y2} = \hat{w}, w_{n2} = 0, a_1 = a_2 = 0, b_1 = c_1, \\ & (\hat{w}, b_2) = \begin{cases} \left(\tilde{\epsilon}, \frac{(\epsilon-\tilde{\epsilon})}{\lambda}\right), \ \tilde{\epsilon} \in [0,\epsilon], & \text{if } \lambda = 1-q \\ \left(0, \frac{\epsilon}{\lambda}\right) & \text{if } \lambda > 1-q \\ (\epsilon, 0) & \text{if } \lambda < 1-q \end{cases}, \\ & \text{or,} \\ & w_{y1} = w_{n1} = w_{n2} = \hat{w}, w_{y2} = 0, a_1 = 0, b_1 = c_1, \\ & (\hat{w}, a_2, b_2) = \begin{cases} \left(\tilde{\epsilon}, \frac{(\epsilon-\tilde{\epsilon})}{\lambda}, -\frac{(\epsilon-\tilde{\epsilon})}{\lambda}\right), \ \tilde{\epsilon} \in [0,\epsilon], & \text{if } \lambda = 1-q \\ \left(0, \frac{\epsilon}{\lambda}, -\frac{\epsilon}{\lambda}\right) & \text{if } \lambda > 1-q \\ (\epsilon, 0, 0) & \text{if } \lambda < 1-q \end{cases}. \end{aligned} \tag{10}$$

[22]In the case $\lambda = \left(\frac{1-q}{q}\right)^2$ in (9) both types are equally effective in providing incentives. Then it does not matter how a given amount of bonus is allocated among the two contracts, as long as the expected bonus is equal to effort cost, $qb_1 + (1-q)b_2 = c$. Thus, (only) in this case we can set the bonuses equal to those under complete information.

The contracts in Lemma 4 might look complicated but, in fact, they are easily explained. The positive parameter ϵ (as well as $\tilde{\epsilon} \leq \epsilon$) is arbitrarily small. It is necessary in order to prevent S_2 from reporting truthfully, as required in case (D). Essentially, ϵ makes S_2 favor one of his two reports.
For $\epsilon \to 0$, we get $\hat{w} = a_2 = b_2 = 0$ which is an intuitive result, as we explain in the following.
Supposing $\epsilon \to 0$, the contract for S_1 is similar to the complete-information result. That contract, with $a_1 + b_1 - c_1 = 0$ does not provide any rents to S_1 and A.
For S_2, however, the contract is different from the complete-information case. Recall that, under (D), S_2 reports independently of A's effort and S_2's type is revealed to A, so we can only implement $e_2 = 0$. Therefore, it is optimal to avoid any rents for S_2 if possible. Since S_1's contract is not attractive for S_2, we can indeed pay zero rents to S_2.

Lemma 5. *Consider case (E). Since types are revealed to A and A knows that* S_1 *reports independently of effort, we have* $e_1 = 0$*. For* $e_2 > 0$*, the set of optimal contracts is characterized by*

$$\begin{aligned} S_1: \quad & \left(w_{y1} = \lambda c_2, w_{n1} \in [0, \lambda c_2), a_1 \geq 0, b_1 = -a_1, w_{n1} + \lambda a_1 \leq \lambda c_2\right), \\ & \text{or,} \\ & \left(w_{y1} \in [0, \lambda c_2), w_{n1} = \lambda c_2, a_1 = 0, b_1 \geq 0, w_{y1} + \lambda b_1 \leq \lambda c_2\right), \\ S_2: \quad & w_{y2} = 0, w_{n2} = \lambda c_2, a_2 = b_2 = c_2. \end{aligned} \tag{11}$$

Here, only S_2 reports effort truthfully and his contract is similar to the complete-information case with the exception of $a_2 = c_2$ (as compared to $a^* = 0$).
This difference can be explained as follows. Due to incomplete information, S_1 must be prevented from choosing S_2's contract. Truthtelling by S_2, however, requires $w_{n2} = \lambda b_2 > 0$ which is attractive for S_1. Thus, S_1 must get the same rent through his contract (equal to $\lambda b_2 = \lambda c_2$). But this rent is not present in the complete-information contracts. Therefore, one cannot achieve zero rents for S_2 either since S_2 can always choose S_1's contract and get that rent. The optimal way of providing that rent for S_2 is through $a_2 > 0$ (raising b_2 or w_{y2} instead would both require larger wages for S_1 again and would, thus, not solve the problem).
Lemma 5 shows that there is some freedom in choosing S_1's wages. This is because $e_1 = 0$ and, thus, A can be ignored in that case. What matters is that S_1 gets that rent, $\max\{w_{y1}, w_{y2}\} = \lambda c_2$, observing limited liability and choosing a_1 and b_1 in a way that makes S_1's contract unattractive for S_2.

Lemma 6. *Consider case (F). Supposing positive efforts, the set of optimal contracts is characterized by*

$$\begin{aligned} S_1: \quad & w_{y1} = w_{n1} = \lambda c_2, a_1 = 0, b_1 = c_1, \\ S_2: \quad & w_{y2} = 0, w_{n2} = \lambda c_2, a_2 = b_2 = c_2. \end{aligned} \tag{12}$$

Case (F) is closer to the complete-information case than the previous cases in the sense that A is informed about the type he is facing and each type reports truthfully in equilibrium. In contrast to the complete-information case, however, we have to ensure that each type self-selects into the right contract. This constraint explains the deviations from the complete-information result: First, truth-telling by S_2 requires that $w_{n2} = \lambda_2 b_2 > 0$ (similar to the complete-information result). This, however, is attractive for S_1 who can achieve a positive utility by choosing S_2's contract. In order to prevent that, one has to give S_1 a rent through his own contract, by setting $w_{y1} = w_{n1} = \lambda_2 b_2$, too (ensuring truth-telling as well). Second, giving S_1 a rent through positive wages implies that S_1's contract becomes more attractive for S_2, as we explain in the following. Suppose $a_2 = 0$. Then, S_2's contract would be the same as the complete-information contract, implying zero rents for S_2. However, by choosing S_1's contract, S_2 would obtain a positive wage (and zero indirect utility since A's utility is zero under S_1's contract). Therefore, we have to give S_2 a rent. What is the optimal way of doing that? There are three options: raising b_2, w_{y2} or a_2. As it turns out, the first two again require higher wages for S_1 while raising a_2 does not change S_1's incentives.[23] Therefore, it is optimal to set $a_2 = b_2$ (resulting in a rent of $\lambda_2 b_2$ as required).

Note that so far, in Lemmas 1 to 6, we characterized the optimal contracts for each case for given requested effort levels only. From now on, we compare those contracts, taking into account that, for each case, the principal maximizes expected profits by choosing optimal effort levels. These optimal effort levels are, of course, generally different between cases (A) to (F). Therefore, all following statements about optimality refer to maximum profits under optimal effort choice. Denoting P's profits from the optimal contracts by $\pi^{(A)}$, $\pi^{(B)}$, etc., we find[24]

Lemma 7.

1. *The optimal contract in case (F) strictly dominates the contracts in cases (D) and (E).*

[23] Raising w_{y2} is directly attractive for S_1 while raising b_2 requires a larger w_{n2} in order to achieve truth-telling by S_2. That, again, is attractive for S_2.

[24] Recall that we denote $\lambda := \lambda_2$.

2. *The optimal contract in case (C) is always exactly as profitable as the most profitable contract among cases (A) and (B). In particular,*

$$\pi^{(C)} = \begin{cases} \pi^{(A)} & \text{if } \lambda \geq \left(\frac{1-q}{q}\right)^2 \\ \pi^{(B)} & \text{if } \lambda < \left(\frac{1-q}{q}\right)^2 \end{cases} = \max\left\{\pi^{(A)}, \pi^{(B)}\right\}. \tag{13}$$

3. *Depending on the model parameters, each of the cases (A), (B), (C) and (F) can contain the globally optimal contract(s). In particular,*

   - *(A) (resp. (C)) can be globally optimal if the neutral supervisor occurs with sufficiently large probability.*
   - *(B) (resp. (C)) contains the globally optimal contract if the altruistic supervisor occurs with sufficiently large probability.*
   - *(F) contains the globally optimal contract when the maximum profits of (A) and (B) are sufficiently close to each other.*[25] *Moreover, it can be globally optimal if the neutral supervisor occurs with sufficiently large probability.*

In the following, we look at the statements of Lemma 7 in more detail.
Why does (F) outperform both (D) and (E)? In fact, $\pi^{(F)}$ is even arbitrarily close to $\pi^{(D)}+\pi^{(E)}$.[26] The point is that under (D) and (E), only one supervisor type reports truthfully and A learns the type. Thus, whenever the other type occurs, effort is zero. Nevertheless, P has to pay the other type a rent in order to induce him to choose the right contract. But P does not get anything in return from that type. Under (F), however, both types report truthfully which means that effort is always positive and, since A learns the type, the effort can be tailored to each type separately. Essentially, (F) can be seen a combination of (D) and (E).
Under (A) and (B) only one type reports truthfully while A does not learn the type. Thus, A will always exert positive effort, in order to earn the bonus from the truth telling type. Since the other type reports independent of effort, A is not incentivized through that other type, regardless of whether that type reports $m = y$ or $m = n$. However, since the truth telling type occurs only with some probability (q, resp. $1 - q$), the bonus must be sufficiently large since A only cares for the *expected* bonus payment. Now, as shown in the proof of Lemma 3, and contrary to the result for contracts (F) above, under

[25] In order to be precise, this is for parameters λ and q sufficiently close to $\lambda = (1-q)^2/q^2$. At this point, $\pi^{(A)} = \pi^{(B)} = \pi^{(C)}$.

[26] The difference is due to an arbitrarily small $\epsilon > 0$, see the proof of Lemma 7.

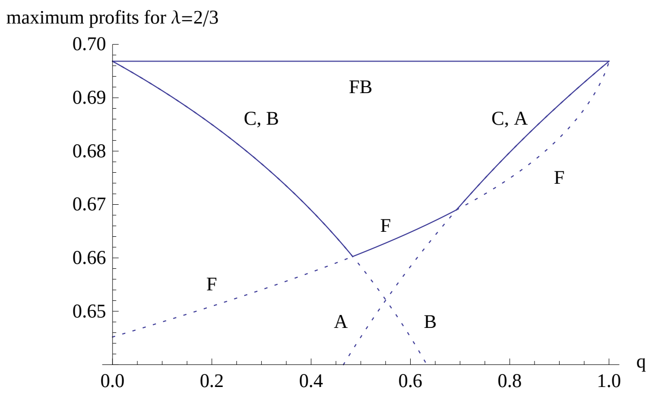

Figure 1: Optimal contracts for $f(e) = e$, $c(e) = e^{10}$ and $\lambda = 2/3$.

(C) the principal makes use of only one supervisor type in the sense that P pays a positive bonus only for one of the two types, although both types report truthfully. This is because, generally, one of the two types is less efficient as a means of inducing effort (see the discussion after Lemma 3). Thus, (C) is essentially only equivalent to the better one of (A) and (B): P makes use of the more profitable one of the two truth telling supervisors.

Finally, what is the intuition for the globally optimal contracts? For contracts (A) and (B) (and, thus, (C), see the discussion above) P sets incentives only through one supervisor type. Her profit is "distorted" away from first-best because of various rents, as discussed previously. The more likely the "right" type becomes, the less rents are to be paid in expectation. In the limit ($q \to 1$ resp. $q \to 0$), these rents disappear, making (A), resp. (B) (and, thus, (C)), first-best. Importantly, for $q \to 1$, (A) and (C) (resp. (B) and (C) for $q \to 0$) approach the complete-information setting, since then the type is "almost" commonly known. We have shown, however, that (F) outperforms (A), (B) (and, thus, (C)) if none of the types is sufficiently likely. Intuitively, (A), (B) and (C) are geared to one of the types, respectively, making these contracts better if that type is likely, while (A) and (B) become arbitrarily bad in the opposite case. (F), in contrast, sets incentives through both types. Therefore, (F) is comparably better in an intermediate range where the other contracts perform worst.

Figure 1 illustrates parts 2 and 3 of Lemma 7 for $f(e) = e$, $c(e) = e^{10}$ and $\lambda = 2/3$. It also shows the first–best profit (FB).[27]

In the following, we derive the main result. By Lemma 7, there is always a globally optimal contract either in (C) or in (F), where both supervisor types report effort truthfully. Moreover, whenever we are sufficiently close

[27]The figure look very similar for, e.g., $f(e) = \sqrt{e}$ and $c(e) = e^2/2$.

to standard (selfish) preferences, in the sense of a sufficiently large q, then contract case (A) can become globally optimal. In that case, we find an especially simple optimal contract:
By Lemma 1, the following contract is then optimal: $w_{y1} = w_{y2} = w_{n1} = w_{n2} = 0$ and $a_1 = a_2 = 0$, $b_1 = b_2 = c/q$. With these wages, S_2 is lenient and always reports favorably. Obviously, this is a type-independent contract (i.e. the menu of contracts can be replaced by a single contract) since neither P nor A need to learn S's type. This contract is close to the optimal contract under unlimited liability (which is also type-independent, see Proposition 2). The difference is that here we cannot have a negative base wage for A. Thus, P cannot extract A's rents, which lowers the optimal effort and therefore the bonus. We summarize our main result:

Proposition 3.

1. *There is always an optimal contract where both supervisor types are induced to report truthfully, i.e., leniency is eliminated.*
2. *If the neutral supervisor type is sufficiently likely, it can be optimal not to eliminate leniency. The corresponding contract can be simple: type-independent, paying the supervisor a flat wage regardless of the agent's effort.*

6 Discussion

Leniency In the present paper we modeled the "negative" view of leniency, i.e., leniency only weakens the link between the agent's performance and reward. Nevertheless, our main result is that optimal contracts do not necessarily eliminate leniency. In particular, under incomplete information and unlimited liability, leniency prevails under the unique optimal contract (Proposition 2). There, the agent is uncertain about the supervisor's type and incentives are given through the neutral supervisor only. The agent knows that only the neutral type reports truthfully, while A would get a bonus even without effort whenever the supervisor happens to be altruistic. Thus, sufficient effort incentives require that the expected bonus due to the neutral type alone must cover the effort cost, which implies that the bonus exceeds actual effort cost. But this is not detrimental to P's profit since A's rents can be extracted through a negative base wage.
Under incomplete information and limited liability, leniency can always be eliminated by offering a menu of incentive-compatible contracts. Essentially,

the supervisor types can always be separated because the altruistic type participates in the agent's utility and, thus, can be paid rents indirectly via the agent's wages while the neutral supervisor prefers a direct wage payment.[28] Since those contracts might imply excessive rents, however, a different, and much simpler kind of contracts can be optimal if the supervisor is neutral with sufficiently large probability. There, similar to the unlimited liability case, the lenient reporting of the altruistic type would essentially be accepted since its elimination is too expensive. Then incentives are provided only through the (very likely) neutral supervisor type (Proposition 3).

If A knows S's type Suppose A is limitedly liable and knows S's type when he chooses effort. Then contracts (A), (B), and (C) are not feasible because they require concealing the type. Then the unique optimal contract is (F). Therefore, it is never in P's interest that A knows S's type although it does not matter whenever (F) is globally optimal, anyway.

Practical Implications We have shown that, in principle, the cases (A), (B), (C), and (F) can contain optimal contracts. While there is only one feasible contract in (F), there are many optimal wage combinations in the other cases.
From a practical perspective, some contracts are simpler or seem to be more realistic than others. In particular, we can use the following two criteria to evaluate contracts. First, consider contracts with type–independent wages (for S and A). There, P can offer a single contract rather than a menu of contracts, uncertainty for A is reduced since A's wages are independent of S's private information, and, neither P nor A need to infer S's type. Second, consider flat wages for S, that do not depend on how A is evaluated. These wages are the simplest and seem to be the most prevalent in reality.[29]
The contract from part 2 of Proposition 3, contained in (A), fits both criteria and it is the unique optimal contract in the model with standard preferences where S is neutral for sure. We have seen that it can remain optimal if the probability of this neutral type is sufficiently large. As indicated before, however, the contract requires A not to learn S's type. In practice, A may

[28] Note that separation of types, i.e., eliminating leniency, is always feasible with contracts of type (F), also with arbitrarily many supervisor types. However, satisfying the many incentive constraints leads to excessive rents such that the optimal required efforts are (near) zero for many of the supervisor types. Thus, these contracts perform much worse with more types unless some types have high probability.

[29] As mentioned before, "flat" here means that the supervisor's wage is independent of the *content* of his report. It does not rule out that the supervisor's pay has other, performance-related, components that refer to other tasks.

learn this type in two ways, either from direct communication with S or from observation of S's behavior. Communication between S and A may not be credible because the neutral type of supervisor is always indifferent between communicating to be of the one or the other type. A possible way to prevent A from learning the type from observation of S's behavior is to introduce job rotation to keep the interaction between agent and supervisor short.[30]

The first criterion is also satisfied by (B) if we set $a = 0$, $b = c/(1-q)$, $w_y = 0$, and $w_n = \lambda c/(1-q)$. This contract has $w_n > w_y$ and the neutral supervisor always reports negatively. In practice, we might observe this if supervisors are rewarded on the basis of cost savings. Contracts (C) and (F) satisfy neither criterion. Yet, these contracts eliminate leniency and induce supervisors to report effort truthfully. In this sense, given that either (C) or (F) is optimal, Arrow Electronics' various efforts to tackle the leniency problem might be justified.

Collusion A full analysis of collusion is beyond the scope of this paper. However, even without a detailed collusion model, we can evaluate the potential gains from collusion under the optimal contracts. The most obvious way of collusion is that the agent exerts zero effort, saving the effort cost, and pays the supervisor a bribe in order to make him report favorably such that the agent gets the bonus anyway.

At first glance, collusion should be more profitable in the presence of an altruistic supervisor as compared to the model with standard (neutral) preferences because any utility gain of the agent implies an additional indirect gain for the supervisor.

However, the potential gains from collusion are actually lower than under the assumption of standard preferences, for several reasons.[31] First, by the altruistic preferences, for any given bribe paid by the agent, the agent's loss is larger than the supervisor's gain.[32] Second, the optimal contracts imply various rents to S or A which lead to efforts below the first–best level and thus reduce the potential gains from collusion. Third, the optimal contracts already take into account the supervisor's tendency for favorable reports. They induce the altruistic supervisor to tell the truth by punishing good reports through lower wages. Thus, the altruistic supervisor's only gain

[30]See e.g. Tirole (1986) and Prendergast and Topel (1993) for different motives of job rotation in a principal-supervisor-agent relationship.

[31]A detailed analysis of the collusion problem under all optimal contracts is available on request.

[32]A bribe payment of t to the supervisor reduces the agent's utility by t. The bribe has two effects on the supervisor. His utility increases directly by t but it also decreases by λt through the agent's utility loss. The supervisor's net gain is $(1-\lambda)t < t$.

from collusion may be the bribe, as in the setting with neutral supervisor. Fourth, under the simple contract from part 2 of Proposition 3, the altruistic supervisor is lenient anyway and thus the agent only wants to bribe the neutral supervisor. But since he does not know which type he is facing, the bribe is wasted with positive probability.
All in all, the potential rents that can be shared between the colluding parties are lower than those under standard preferences. These gains have to be weighed up against the cost of collusion, like, e.g., the risk of detection and punishment.

7 Conclusion

We looked at the optimal design of contracts in a principal–supervisor–agent relationship where the supervisor may be lenient because he cares for the agent's well–being. This corresponds to the "negative" view on leniency, i.e., that leniency weakens the link between pay and performance.
Nevertheless, the main insight is that, under incomplete information and limited liability, it can be optimal to have a very simple contract that does not eliminate leniency although this is feasible. In particular, this contract pays the supervisor a flat fee, independent of his performance reports, and does not reveal his type. Under this contract the supervisor who cares for the agent is lenient. This kind of contract is always optimal under unlimited liability.
We also show that there always exists an optimal contract that eliminates leniency. This might be a justification for firms' efforts to reduce leniency.

8 Appendix

Proof of Proposition 1. It is easily verified that the proposed contract satisfies (5). The principal's profit simplifies to $\pi = f_{FB} - c_{FB}$. □

Proof of Proposition 2. The proof is done in two steps. First, we show that the proposed contract yields the first-best solution and is therefore optimal. Second, we show that it is the only contract with this property. **(1)** If $w_y^* = w_n^* = 0$, S_1 reports A's effort truthfully, while the other type reports $m = y$. Thus, A chooses effort e, if

$$a + b - c(e) \geq a + (1-q)\,b \quad \Leftrightarrow \quad b \geq \frac{c(e)}{q}. \tag{14}$$

Thus, b^* induces $e = e_{FB}$. S and A participate since their payoffs are zero. P gets the first–best profit: $\pi = f_{FB} - b^* - a^* = f_{FB} - c_{FB}$. **(2)** P's profit can be written as $\pi = f(e) - w_S - w_A$, where w_S and w_A are the wage payments to S and A. From S's participation constraints it follows that $w_S \geq -\lambda U_A$. Hence, P's profit is not above $f(e) + \lambda U_A - w_A$ or $f(e) + \lambda U_A - (U_A + c(e))$. As $\lambda < 1$, first-best requires $U_A = 0$, i.e., A must not receive a rent. If the contract is such that no supervisor type reports A's effort truthfully, then positive effort cannot be induced, i.e. A chooses $e = 0$. Moreover, a contract can only be optimal if S accepts it and no type of supervisor receives a positive rent. The latter requirement implies $\max\{w_y, w_n\} \leq 0$. Otherwise, S_1 receives a strictly positive rent. As $U_A = 0$, the only contract satisfying $\max\{w_y, w_n\} \leq 0$, inducing at least one type of S to report y truthfully and ensuring participation of S has $w_y = w_n = 0$. □

In order to save on space and notation, we prove Lemmas 1-6 in a combined proof, avoiding the repetition of similar arguments and conditions. Along the way, we also derive P's profits for each case.

Proof of Lemmas 1 to 6. In all cases below we will see that the expected bonus payment covers the expected effort cost. This, together with (6) takes care of A's and S's participation constraints. In cases (D) and (E), A learns S's type. Whenever he faces a type who does not have incentives to report truthfully, then the only implementable effort is zero. In all other cases, i.e., when A knows (or expects with positive probability) that S reports effort truthfully, then strictly positive effort is feasible and optimal. Throughout the proof we assume strictly positive requested effort (and thus strictly positive cost and output) unless indicated otherwise. At the end of the proof, we show why this is optimal. Throughout the proof, we repeatedly refer to "relevant" wages and "optimal" wages. The first term refers to those wage components that are payoff–relevant on the equilibrium path. The second term refers to wages that minimize the expected value of total relevant wages in equilibrium (to S and A) for *given* effort. The proof proceeds by minimizing the expected total relevant wages for each case observing all relevant constraints. We explicitly give those constraints but the easy task of verifying that the proposed wages satisfy the constraints is left to the reader. We also characterize the set(s) of non–relevant wages that support the equilibrium in each case. First, we state the set of conditions that have to be satisfied by the optimal contract(s). Limited liability is defined by (6). Note that negative bonuses, $b_i < 0$, are generally feasible. S_i's incentive constraints for truthful reporting of A's effort, given that P demands that effort be at least

$e = \hat{e}$, are

$$w_{ni} + \lambda_i \left(a_i - c_i \left(\tilde{e}\right)\right) \geq w_{yi} + \lambda_i(a_i + b_i - c_i \left(\tilde{e}\right)) \quad \forall\, \tilde{e} < \hat{e} \tag{15}$$
$$w_{yi} + \lambda_i(a_i + b_i - c_i \left(\bar{e}\right)) \geq w_{ni} + \lambda_i(a_i - c_i \left(\bar{e}\right)) \quad \forall\, \bar{e} \geq \hat{e}, \tag{16}$$

which is equivalent to $w_{yi} + \lambda_i b_i = w_{ni}$. Thus, truthful reporting implies

$$\mathrm{S}_1 : \quad w_{y1} = w_{n1} \tag{17}$$
$$\mathrm{S}_2 : \quad w_{y2} + \lambda b_2 = w_{n2}. \tag{18}$$

Moreover, wages $w_{yi} + \lambda_i b_i > w_{ni}$ (resp. $< w_{ni}$) induce $\mathrm{S_i}$ to always report $m = y$ (resp. $m = n$) regardless of effort. In such a case, A will not exert effort if he learns S's type (as under contract cases (D), (E)). Under any contract, $\mathrm{S_i}$ (where $\mathrm{S_j}$ is the other type) selects the right contract only if his corresponding payoff, π_{Si}, is at least as large as the payoff from imitating the other type,

$$\pi_{Si} \geq \max \left\{w_{yj} + \lambda_i(a_j + b_j - c_j), w_{nj} + \lambda_i(a_j - c_j)\right\}. \tag{19}$$

For S_1 (where $\lambda_1 = 0$) we have $\pi_{S1} \in \{w_{y1}, w_{n1}\}$ and thus, depending on the relation of w_{y1} and w_{n1}, (19) simplifies to

$$w_{y1} = w_{n1} \;\Rightarrow\; (19): \; w_{y1} = w_{n1} \geq \max \left\{w_{y2}, w_{n2}\right\}, \tag{20}$$
$$w_{y1} > w_{n1} \;\Rightarrow\; (19): \; w_{y1} \geq \max \left\{w_{y2}, w_{n2}\right\}, \tag{21}$$
$$w_{y1} < w_{n1} \;\Rightarrow\; (19): \; w_{n1} \geq \max \left\{w_{y2}, w_{n2}\right\}. \tag{22}$$

Cases (A), (B), (C) Here, P conceals S's type from A. Therefore, obviously, the requested (and equilibrium) efforts are the same for both types, $e_1 = e_2 =: e > 0$ (and c, f without subscript). **Case (A)** Here, only S_1 reports effort truthfully, i.e., (17) holds while (18) does not. We divide (A) into subcases (A1) where S_2 always reports $m = y$, and (A2) where the report is always $m = n$. Since only S_1's report depends on effort, A's incentive constraints for (A1) and (A2) are the same:

$$\begin{aligned} &(A1): q(a_1 + b_1) + (1-q)(a_2 + b_2) - c \geq qa_1 + (1-q)(a_2 + b_2) \iff b_1 \geq \frac{c}{q} \\ &(A2): q(a_1 + b_1) + (1-q)a_2 - c \geq qa_1 + (1-q)a_2 \iff b_1 \geq \frac{c}{q}. \end{aligned} \tag{23}$$

In addition to the above and (6), we have to satisfy constraints (17), (19), and (20), while (18) is violated. Collecting and simplifying, we get

$$
\begin{aligned}
& b_1 \geq \frac{c}{q} && (24)\\
& w_{y1} = w_{n1} && (25)\\
& w_{y1} \geq \max\{w_{y2}, w_{n2}\} && (26)\\
(A1):\ & w_{y2} + \lambda b_2 > w_{n2} && (27)\\
(A1):\ & w_{y2} + \lambda(a_2 + b_2) \geq \max\{w_{y1} + \lambda(a_1 + b_1), w_{n1} + \lambda a_1\} && (28)\\
(A2):\ & w_{y2} + \lambda b_2 < w_{n2} && (29)\\
(A2):\ & w_{n2} + \lambda a_2 \geq \max\{w_{y1} + \lambda(a_1 + b_1), w_{n1} + \lambda a_1\}. && (30)
\end{aligned}
$$

(A1) By (24), $b_1 > 0$. Together with (25), the RHS of (28) evaluates to $w_{y1} + \lambda(a_1 + b_1)$. By (26), $w_{y1} - w_{y2} \geq 0$. Thus, (28) implies

$$
\begin{aligned}
& w_{y1} - w_{y2} \leq \lambda(a_2 + b_2 - (a_1 + b_1)) && (31)\\
\Rightarrow\ & a_2 + b_2 \geq a_1 + b_1. && (32)
\end{aligned}
$$

The relevant wages are a_i, b_i, w_{yi}, while w_{ni} are not relevant, $i \in \{1, 2\}$. We minimize total relevant wages. First, (24) holds with equality, $b_1 = \frac{c}{q}$. Second, the cheapest way to satisfy (32) is with $a_1 = 0$ (limited liability) and $a_2 + b_2 = \frac{c}{q}$. Third, the cost-minimizing payoff-relevant supervisor wages are $w_{y1} = w_{y2} = 0$. This requires $w_{n1} = w_{n2} = 0$, by (25) and (26). In equilibrium, only the sum $a_2 + b_2$ is relevant. Thus, there is a degree of freedom in choosing a_2 and b_2 as long as $b_2 > 0$ (in order to satisfy (27)). It follows that P's maximum profit is

$$\pi^{(A1)} = f - q(a_1 + b_1 + w_{y1}) - (1 - q)(a_2 + b_2 + w_{y2}) = f - \frac{c}{q}. \qquad (33)$$

(A2) By (24), $b_1 > 0$. Together with (25), the RHS of (30) evaluates to $w_{y1} + \lambda(a_1 + b_1)$. By (26), $w_{y1} - w_{n2} \geq 0$. Thus, (30) implies

$$
\begin{aligned}
& w_{y1} - w_{n2} \leq \lambda(a_2 - (a_1 + b_1)) && (34)\\
\Rightarrow\ & a_2 \geq a_1 + b_1. && (35)
\end{aligned}
$$

The relevant wages are a_1, a_2, b_1, w_{y1}, and w_{n2}, while b_2, w_{n1}, and w_{y2} are not. First, (24) holds with equality, $b_1 = \frac{c}{q}$. Second, the cheapest way to satisfy (35) is with $a_1 = 0$ and $a_2 = \frac{c}{q}$. Third, the cheapest relevant S–wages are $w_{y1} = w_{n2} = 0$. By (25) and (26), this requires $w_{n1} = w_{y2} = 0$. By (29), we must set $b_2 < 0$ and, by (6), $a_2 + b_2 \geq 0$. Since b_2 is not relevant, we get $b_2 \in [-a_2, 0) = [-c/q, 0)$. Then P's profit is the same as under (A1):

$$\pi^{(A2)} = f - q(a_1 + b_1 + w_{y1}) - (1 - q)(a_2 + w_{n2}) = f - \frac{c}{q}. \qquad (36)$$

Summarizing (A1) and (A2), we can express a_2 and b_2 in a more concise way,

$$a_2 \geq 0, a_2 + b_2 \geq 0, b_2 \neq 0, a_2 + \max\{0, b_2\} = \frac{c}{q}, \tag{37}$$

where the first two conditions ensure limited liability, the third prevents truth telling by S_2 while the last one summarizes the fact that A's total rents must be equal to c/q. Since (33)=(36), we denote

$$\pi^{(A)} = f - \frac{c}{q}. \tag{38}$$

Case (B) Here, only S_2 reports effort truthfully, i.e., (18) holds, while we have either subcase (B1), where (21) holds, or subcase (B2) where (22) is satisfied. Positive effort requires $b_2 > 0$ since S_1's report is independent of effort. Similar to (23) we get $b_2 \geq c/(1-q)$. Collecting and simplifying all relevant constraints, we need to satisfy

$$b_2 \geq \frac{c}{(1-q)} \tag{39}$$

$$w_{y2} + \lambda b_2 = w_{n2} \tag{40}$$

$$w_{y2} + \lambda(a_2 + b_2) \geq \max\{w_{y1} + \lambda(a_1 + b_1), w_{n1} + \lambda a_1\} \tag{41}$$

$$(B1) : w_{y1} > w_{n1} \tag{42}$$

$$(B1) : w_{y1} \geq \max\{w_{y2}, w_{n2}\} \tag{43}$$

$$(B2) : w_{y1} < w_{n1} \tag{44}$$

$$(B2) : w_{n1} \geq \max\{w_{y2}, w_{n2}\} \tag{45}$$

(B1) The relevant wages are a_i, b_i, w_{yi}, while w_{ni} are not relevant, $i \in \{1, 2\}$. Choose b_2 and w_{y2} as low as possible, i.e., $b_2 = c/(1-q)$, by (39), and $w_{y2} = 0$. By (40) this leads to $w_{n2} = \lambda c/(1-q)$. By (43), it follows that $w_{y1} = \lambda c/(1-q)$. By (42) and (6), $w_{n1} \in [0, \lambda c/(1-q))$. For the remaining relevant wages, it is feasible to have the minimum values $a_2 = 0$ and $a_1 + b_1 = 0$. Note that only the sum $a_1 + b_1$ but not its composition is payoff–relevant. Inserting the above into (41), we get

$$\frac{\lambda c}{1-q} \geq \max\left\{\frac{\lambda c}{1-q}, w_{n1} + \lambda a_1\right\}. \tag{46}$$

Since w_{n1} is not relevant and a_1 only matters indirectly, as part of the sum $a_1 + b_1$, there is some degree of freedom as long as the value of the RHS is not affected. Precisely, we need $w_{n1} + \lambda a_1 \leq \lambda c/(1-q)$, where $a_1 \geq 0$, $b_1 = -a_1$, and $w_{n1} \in [0, \lambda c/(1-q))$, by (42). P's maximum profit is

$$\begin{aligned} \pi^{(B1)} &= f - q(a_1 + b_1 + w_{y1}) - (1-q)(a_2 + b_2 + w_{y2}) \\ &= f - c\left(1 + \frac{q}{1-q}\lambda\right). \end{aligned} \tag{47}$$

(B2) Here, the relevant wages are a_1, a_2, b_2, w_{n1}, and w_{y2}, while b_1, w_{y1} and w_{n2} are not relevant. By (39), $b_2 = c/(1-q)$. We set $w_{y2} = 0$ and, by (40), $w_{n2} = \lambda c/(1-q)$. By (45), we cannot pay less than $w_{n1} = \lambda c/(1-q)$. For w_{y1}, which is not relevant, we get, by (44), $w_{y1} \in [0, \lambda c/(1-q))$. For the remaining relevant wages, we claim that $a_1 = a_2 = 0$ is feasible. Inserting the above into (41), we get

$$\frac{\lambda c}{1-q} \geq \max\left\{w_{y1} + \lambda b_1, \frac{\lambda c}{1-q}\right\}, \tag{48}$$

which is satisfied, provided that w_{y1} and b_1 (which are not relevant) are set such that they don't increase the RHS of (48), i.e., $w_{y1} + \lambda b_1 \leq \lambda c/(1-q)$. Together with $b_1 \geq 0$ (by limited liability and $a_1 = 0$) and $w_{y1} \in [0, \lambda c/(1-q))$ (see above), we have determined all wages. P's profit becomes

$$\begin{aligned} \pi^{(B2)} &= f - q(a_1 + w_{n1}) - (1-q)(a_2 + b_2 + w_{y2}) \\ &= f - c\left(1 + \frac{q}{1-q}\lambda\right). \end{aligned} \tag{49}$$

Since (47)=(49), we denote

$$\pi^{(B)} = f - c\left(1 + \frac{q}{1-q}\lambda\right). \tag{50}$$

Case (C) Here, S_1 and S_2 report effort truthfully, i.e., (17) and (18) hold. A doesn't learn the supervisor's type and thus exerts positive effort if

$$q(a_1 + b_1) + (1-q)(a_2 + b_2) - c \geq qa_1 + (1-q)a_2 \tag{51}$$

$$\Rightarrow\ qb_1 + (1-q)b_2 \geq c, \tag{52}$$

By limited liability, (6), a negative bonus, $b_i < 0$, is feasible as long as $a_i + b_i \geq 0$. For the moment, suppose that $b_i \geq 0$. We derive the optimal contract(s) for that case and afterwards prove that $b_i < 0$ is never optimal. Inserting (18) into the RHS of inequality (20), we get

$$w_{y1} \geq w_{y2} + \lambda b_2. \tag{53}$$

Under (C), (19) simplifies to

$$w_{y2} + \lambda(a_2 + b_2) \geq \max\{w_{y1} + \lambda(a_1 + b_1), w_{n1} + \lambda a_1\}. \tag{54}$$

By (17) and $b_i \geq 0$, the RHS of (54) is equal to $w_{y1} + \lambda(a_1 + b_1)$. After replacing w_{y1} by the RHS of (53) and simplifying, we get

$$a_2 \geq a_1 + b_1. \tag{55}$$

First, a_1 and a_2 are relevant. From (55) we get $a_1 = 0$ (limited liability) and $a_2 = b_1$. Second, (52) is binding. Third, w_{y1} and w_{y2} are relevant. From (53) and $b_i \geq 0$ we get $w_{y2} = 0$ and $w_{y1} = \lambda b_2$. By (17) and (18), it follows that $w_{n1} = w_{n2} = \lambda b_2$. Thus, P's expected profit is, for $b_i \geq 0$,

$$\begin{aligned}\pi^{(C)} &= f - q(a_1 + b_1 + w_{y1}) - (1-q)(a_2 + b_2 + w_{y2}) \\ &= f - q\left(b_1 + \lambda b_2\right) - (1-q)\left(b_1 + b_2\right) \\ &= f - b_1 - b_2(1 - q + q\lambda). \end{aligned} \tag{56}$$

P's maximization problem becomes

$$\max_{b_1, b_2, e} f(e) - b_1 - b_2(1 - q + q\lambda) \tag{57}$$

$$s.t.\ qb_1 + (1-q)\, b_2 = c\,(e)\,. \tag{58}$$

By (58), P can achieve a certain fixed increase in effort by increasing either b_1 by k units or b_2 by $\frac{qk}{1-q}$ units. By (57), the corresponding cost of this additional effort are either k or $(1-q+q\lambda)\,\frac{qk}{1-q}$, respectively. Hence, if $k > (1-q+q\lambda)\,\frac{qk}{1-q}$, or, equivalently, if $\lambda < \left(\frac{1-q}{q}\right)^2$, P sets $b_1 = 0$ and $b_2 = c/(1-q)$ and expected profit becomes

$$\pi^{(C)} = f - c\left(1 + \frac{q}{1-q}\lambda\right). \tag{59}$$

Similarly, if $\lambda > \left(\frac{1-q}{q}\right)^2$, $b_2 = 0$ and $b_1 = c/q$ are optimal and

$$\pi^{(C)} = f - \frac{c}{q}. \tag{60}$$

Finally, if $\lambda = \left(\frac{1-q}{q}\right)^2$, both, b_1 and b_2 are equally effective (and costly). In that case, (59) and (60) are equal and there is a degree of freedom in setting b_1 and b_2, i.e., they may both be strictly positive (as long as (58) holds). Now we rule out the optimality of negative bonuses, $b_i < 0$. Consider the set of optimal wages derived above where $b_1 = 0$. Now suppose b_1 is reduced by any amount $t > 0$ to $b_1 = -t$. Then, by (52), b_2 increases, $b_2 \geq tq/(1-q) + c/(1-q)$. By (6), the negative b_1 must be compensated by $a_1 \geq t$. By (18), w_{n2} increases and, thus, by (20), w_{y1} and w_{n1} increase as well. As before, we need $w_{n2}, w_{y1}, w_{n1} \geq \lambda b_2$ but b_2 is larger now. Consider (54). Since $w_{y1} = w_{n1}$ and $a_1 \geq t$ now, the RHS of (54) has increased (as compared to the case $b_i \geq 0$). Thus, the LHS must also increase. This would be achieved by setting $a_2 \geq t$, since raising w_{y2} or b_2 would again increase w_{n2}, w_{y1} and w_{n1}. Applying the cost minimization argument, we get an expected profit of

$$\pi^{(C),b_1<0} = f - c\left(1 + \frac{q}{1-q}\lambda\right) - t\left(1 + \frac{q^2}{1-q}\lambda\right), \tag{61}$$

which is less than (59). By the previous analysis, P can get the maximum of (59) and (60) with $b_i \geq 0$. Thus, $b_1 < 0$ is never desirable. A similar argument can be made for $b_2 < 0$. Consider the set of optimal wages derived above where $b_2 = 0$. Now suppose b_2 is reduced by some amount $t > 0$ to $b_2 = -t$. We get $b_1 = c/q + t(1-q)/q$. Then $a_1 = 0$ and, by (18), $w_{n2} = 0$ and $w_{y2} = \lambda t$. It follows that $w_{y1} = w_{n1} = \lambda t$. Inserting into (54), we get

$$\lambda t + \lambda(a_2 - t) \geq \lambda t + \lambda\left(\frac{c}{q} + t\frac{1-q}{q}\right) \iff \lambda a_2 \geq \lambda\frac{c+t}{q}. \tag{62}$$

Note that, for any given t, in order to satisfy (62), raising w_{y2} on the LHS of (54) does not help: it would in turn increase w_{y1} and thus the RHS. Thus we need $a_2 = (c+t)/q$. We get a profit of

$$\pi^{(C),b_2<0} = f - \frac{c}{q} - t\left(\lambda + \frac{1-q}{q}\right), \tag{63}$$

which is less than (64), and thus never desirable (see the argument above). Summarizing the above, we get

$$\pi^{(C)} = \begin{cases} f - c\left(1 + \frac{q\lambda}{1-q}\right) & \text{if } \lambda < \left(\frac{1-q}{q}\right)^2 \\ f - \frac{c}{q} & \text{if } \lambda \geq \left(\frac{1-q}{q}\right)^2. \end{cases} \tag{64}$$

Cases (D), (E), (F) Here, P reveals the type report to A before A chooses effort. Then A chooses type-dependent efforts, e_1 and e_2. **Case (D)** Here, only S_1 tells the truth, i.e., (17) holds while (18) does not. We divide the case into (D1) where S_2 always reports $m = y$, and (D2) where he reports $m = n$. Since S_2 reports independent of effort, A chooses $e_2 = 0$ (and thus $c_2 = f_2 = 0$). Therefore, A's incentive constraint is $a_1 + b_1 - c_1 \geq a_1$, or, $b_1 \geq c_1$. Collecting (and simplifying) all incentive constraints (see (17), (19), and (20)), we need to satisfy (6) and

$$b_1 \geq c_1 \tag{65}$$

$$w_{y1} = w_{n1} \tag{66}$$

$$w_{y1} \geq \max\{w_{y2}, w_{n2}\} \tag{67}$$

$$(D1): w_{y2} + \lambda b_2 > w_{n2} \tag{68}$$

$$(D1): w_{y2} + \lambda(a_2 + b_2) \geq \max\{w_{y1} + \lambda(a_1 + b_1 - c_1), w_{n1} + \lambda(a_1 - c_1)\} \tag{69}$$

$$(D2): w_{y2} + \lambda b_2 < w_{n2} \tag{70}$$

$$(D2): w_{n2} + \lambda a_2 \geq \max\{w_{y1} + \lambda(a_1 + b_1 - c_1), w_{n1} + \lambda(a_1 - c_1)\}. \tag{71}$$

(D1) By (65) and (66), the RHS of (69) simplifies to $w_{y1} + \lambda(a_1 + b_1 - c_1)$. Since $w_{y1} \geq w_{y2}$, by (67), and $\lambda > 0$, inequality (69) gives us

$$w_{y1} - w_{y2} \leq \lambda(a_2 + b_2 - (a_1 + b_1 - c_1)) \tag{72}$$

$$\Rightarrow \; a_2 + b_2 \geq a_1 + b_1 - c_1. \tag{73}$$

The relevant wages are a_i, b_i, and w_{yi} ($i \in \{1, 2\}$), while w_{ni} are not relevant. We cannot pay less than $b_1 = c_1$ and $a_1 = 0$. Then (73) would be satisfied for any $a_2 + b_2 \geq 0$ (which must hold by (6)). Now consider (68). Suppose we want to achieve a difference of some $\epsilon > 0$ between the LHS and the RHS of (68) and do so in a cost–minimizing way. On the LHS of (68), both w_{y2} and b_2 are relevant. Thus, we set the RHS (which is not relevant) as low as possible, $w_{n2} = 0$. Although w_{y2} and b_2 need to be as small as possible, in order to satisfy (68) we must set one of them (or both) strictly positive. Raising the LHS of (68) to $\epsilon > 0$ requires $b_2 = \epsilon/\lambda$ and costs $(1-q)\epsilon/\lambda$ in expectation while $w_{y2} = \epsilon$ leads to $w_{y1} = w_{n1} = \epsilon$ (by (67) and (66)) and costs $q\epsilon + (1-q)\epsilon = \epsilon$. Thus, if $\lambda > 1-q$, a positive b_2 is cheaper and we set $b_2 = \epsilon/\lambda$, $w_{y2} = w_{y1} = w_{n1} = 0$ while if $\lambda < 1-q$, we set $b_2 = 0$ and $w_{y2} = w_{y1} = w_{n1} = \epsilon$. Finally, suppose $\lambda = 1-q$. Then it is optimal to have $w_{y2} = w_{y1} = w_{n1} = \tilde{\epsilon} \leq \epsilon$ and $b_2 = (\epsilon - \tilde{\epsilon})/\lambda$. Finally, $a_2 = 0$. Note that $b_2 < 0$ does not make sense since, by (6), this must be offset with $a_2 > 0$ while, by (68) and (67), it requires even higher w_{y2} and w_{y1}. The principal's expected profit becomes

$$\begin{aligned}\pi^{(D1)} &= q(f_1 - a_1 - b_1 - w_{y1}) + (1-q)(-a_2 - b_2 - w_{y2}) \\ &= \begin{cases} q(f_1 - c_1) - \epsilon & \text{if } \lambda \leq 1-q \\ q(f_1 - c_1) - \frac{1-q}{\lambda}\epsilon & \text{if } \lambda > 1-q \end{cases}\end{aligned} \tag{74}$$

(D2) By (65) and (66), the RHS of (71) simplifies to $w_{y1} + \lambda(a_1 + b_1 - c_1)$. Since $w_{y1} \geq w_{n2}$, by (67), and $\lambda > 0$, inequality (71) gives us

$$w_{y1} - w_{n2} \leq \lambda(a_2 - (a_1 + b_1 - c_1)) \tag{75}$$

$$\Rightarrow \; a_2 \geq a_1 + b_1 - c_1. \tag{76}$$

The relevant wages are a_1, b_1, w_{y1}, a_2, and w_{n2}, while w_{n1}, b_2, and w_{y2} are not relevant. We cannot pay less than $a_1 = 0$ and $b_1 = c_1$. Consider (70) and for the moment suppose that we want to achieve a difference of $\epsilon > 0$ between the RHS and LHS of (70). The wages on the LHS of (70) are not relevant, but the RHS is, thus, $w_{y2} = 0$. The only way to make the LHS negative is with $b_2 < 0$, i.e., $b_2 = -\epsilon/\lambda$. However, by (6), this requires at least $a_2 = \epsilon/\lambda$ which is relevant and costs $(1-q)\epsilon/\lambda$ in expectation. The only alternative is

$w_{n2} = \epsilon$. This, however, is relevant and, by (67), it requires $w_{y1} = \epsilon$ which is also relevant, and, by (66), $w_{n1} = \epsilon$ which is not relevant. The total cost of this is $q\epsilon + (1-q)\epsilon = \epsilon$. Thus, if $\lambda < (1-q)$, we set $w_{n2} = w_{y1} = w_{n1} = \epsilon$ and $a_2 = b_2 = 0$ while if $\lambda > (1-q)$ we set $w_{n2} = w_{y1} = w_{n1} = 0$ and $b_2 = -\epsilon/\lambda$ and $a_2 = \epsilon/\lambda$. Finally, if $\lambda = (1-q)$, we get $w_{n2} = w_{y1} = w_{n1} = \tilde{\epsilon} \leq \epsilon$, $b_2 = -(\epsilon - \tilde{\epsilon})/\lambda$ and $a_2 = (\epsilon - \tilde{\epsilon})/\lambda$. P's expected profit becomes

$$\begin{aligned} \pi^{(D2)} &= q(f_1 - a_1 - b_1 - w_{y1}) + (1-q)(-a_2 - w_{n2}) \\ &= \begin{cases} q(f_1 - c_1) - \epsilon & \text{if } \lambda \leq 1-q \\ q(f_1 - c_1) - \frac{1-q}{\lambda}\epsilon & \text{if } \lambda > 1-q, \end{cases} \end{aligned} \tag{77}$$

which is equal to $\pi^{(D1)}$. Thus, we denote

$$\pi^{(D)} = \begin{cases} q(f_1 - c_1) - \epsilon & \text{if } \lambda \leq 1-q \\ q(f_1 - c_1) - \frac{1-q}{\lambda}\epsilon & \text{if } \lambda > 1-q. \end{cases} \tag{78}$$

Case (E) Only S_2 reports effort truthfully, i.e., (18) holds while (17) does not. We divide (E) into (E1) where S_1 always reports $m = y$ and (E2) where $m = n$. Since S_1 reports independent of A's effort, A chooses $e_1 = 0$ (and, thus, $c_1 = f_1 = 0$) while A's incentive constraint for e_2 is $a_2 + b_2 - c_2 \geq a_2$, or, $b_2 \geq c_2$. Collecting (and simplifying) all relevant constraints (see (6), (18), (19), (21) for (E1), and (22) for (E2)), we need to satisfy

$$b_2 \geq c_2 \tag{79}$$
$$w_{y2} + \lambda b_2 = w_{n2} \tag{80}$$
$$w_{y2} + \lambda(a_2 + b_2 - c_2) \geq \max\{w_{y1} + \lambda(a_1 + b_1), w_{n1} + \lambda a_1\} \tag{81}$$
$$(E1): w_{y1} > w_{n1} \tag{82}$$
$$(E1): w_{y1} \geq \max\{w_{y2}, w_{n2}\} \tag{83}$$
$$(E2): w_{y1} < w_{n1} \tag{84}$$
$$(E2): w_{n1} \geq \max\{w_{y2}, w_{n2}\}. \tag{85}$$

(E1) The relevant wages are a_i, b_i, w_{yi}, for $i \in \{1, 2\}$, while w_{ni} are not relevant. Set b_2 to its minimum value, $b_2 = c_2$, see (79). We cannot pay less than $w_{y2} = 0$. It follows, by (80), that $w_{n2} = \lambda c_2$. By (83), we get $w_{y1} = \lambda c_2$. Inserting these values in (81), we get

$$\lambda a_2 \geq \max\{\lambda c_2 + \lambda(a_1 + b_1), w_{n1} + \lambda a_1\}. \tag{86}$$

The LHS of (81) is S_2's equilibrium payoff, so every component is relevant. By the above and (6), the RHS of (86) cannot be less than λc_2. It is feasible to have both sides equal to λc_2: Consider the LHS of (81). Raising w_{y2} or b_2 is not an option: both increase w_{n2}, by (80), and therefore w_{y1}, by (83), which would increase the RHS of (86) to a value above λc_2. Thus, we set $a_2 = c_2$ which makes the LHS of (86) equal to λc_2. Then, in order to keep the RHS equal to λc_2, we need $a_1 + b_1 = 0$ and $w_{n1} + \lambda a_1 \leq \lambda c_2$. There is some degree of freedom in setting the remaining wages, as long as we observe these two conditions, as well as (6) and (82). Thus, we get $a_1 \geq 0$, $b_1 = -a_1$, and $w_{n1} \in [0, \lambda c_2)$. The principal's profit is therefore

$$\begin{aligned}\pi^{(E1)} &= q(-a_1 - b_1 - w_{y1}) + (1-q)(f_2 - a_2 - b_2 - w_{y2}) \\ &= (1-q)(f_2 - c_2) - c_2(1 - q + q\lambda).\end{aligned} \tag{87}$$

(E2) The relevant wages are a_1, w_{n1}, a_2, b_2, and w_{y2}, while b_1, w_{y1}, and w_{n2} are not relevant. Set b_2 to its minimum value, $b_2 = c_2$, see (79). We cannot pay less than $w_{y2} = 0$. It follows, by (80), that $w_{n2} = \lambda c_2$. By (85), we get $w_{n1} = \lambda c_2$. Inserting these values in (81), we get

$$\lambda a_2 \geq \max\left\{w_{y1} + \lambda(a_1 + b_1), \lambda c_2 + \lambda a_1\right\}. \tag{88}$$

The LHS of (81) is S_2's equilibrium payoff, so every component is relevant. Look at the second term on the RHS of (88). Since $a_1 \geq 0$, the RHS cannot be less than λc_2. Therefore, on the LHS, we cannot have $a_2 < c_2$. It is, however, feasible (and, thus, cost-minimizing) to have $a_2 = c_2$. This obviously requires $a_1 = 0$ and $w_{y1} + \lambda b_1 \leq \lambda c_2$, while $w_{y1} \in [0, \lambda c_2)$ and $b_1 \geq 0$ in order to satisfy (84) and (6), respectively. Recall that w_{y1} and b_1 are not relevant. The principal's profit is equal to $\pi^{(E1)}$:

$$\begin{aligned}\pi^{(E2)} &= q(-a_1 - w_{n1}) + (1-q)(f_2 - a_2 - b_2 - w_{y2}) \\ &= (1-q)(f_2 - c_2) - c_2(1 - q + q\lambda).\end{aligned} \tag{89}$$

Thus, we denote

$$\pi^{(E)} = (1-q)(f_2 - c_2) - c_2(1 - q + q\lambda). \tag{90}$$

Case (F) Here, both S–types report truthfully, see (17) and (18). Since A learns S's type, his incentive constraints are $a_i + b_i - c_i \geq a_i$, or, $b_i \geq c_i$ for $i \in \{1, 2\}$. A chooses positive but (potentially) different efforts, e_1, e_2. Collecting (and simplifying) all relevant constraints (see (6), (17), (18), (19), and (20)), we need to satisfy

$$b_1 \geq c_1, b_2 \geq c_2 \tag{91}$$

$$w_{y1} = w_{n1} \tag{92}$$

$$w_{y1} \geq \max\{w_{y2}, w_{n2}\} \tag{93}$$

$$w_{y2} + \lambda b_2 = w_{n2} \tag{94}$$

$$w_{y2} + \lambda(a_2 + b_2 - c_2) \geq \max\left\{w_{y1} + \lambda(a_1 + b_1 - c_1), w_{n1} + \lambda(a_1 - c_1)\right\} \tag{95}$$

The relevant wages are a_i, b_i, w_{yi} for $i \in \{1, 2\}$, while w_{ni} are not relevant. By (91), the minimum bonuses are $b_1 = c_1$ and $b_2 = c_2$. This and (92) imply that the RHS of (95) is equal to $w_{y1} + \lambda(a_1 + b_1 - c_1)$. Set $w_{y1} = w_{y2} + \lambda b_2$ since combining (93) and (94) implies that it cannot be lower than that. Then (95) delivers

$$a_2 - c_2 \geq a_1 + b_1 - c_1. \tag{96}$$

By $b_1 = c_1$ and (6), the RHS of (96) is nonnegative, which implies that we cannot have less than $a_2 = c_2$. We set the remaining wages equal to their minimum feasible values: $w_{y2} = 0$, then, by (94), $w_{n2} = \lambda c_2$. As shown above, $w_{y1} = w_{y2} + \lambda b_2 = \lambda c_2$, and, by (92), $w_{n1} = \lambda c_2$. Finally, $a_1 = 0$. P's profit becomes

$$\begin{aligned}\pi^{(F)} &= q(f_1 - a_1 - b_1 - w_{y1}) + (1-q)(f_2 - a_2 - b_2 - w_{y2}) \\ &= q(f_1 - c_1) + (1-q)(f_2 - c_2) - c_2(1 - q + q\lambda). \end{aligned} \tag{97}$$

Collecting results, we find that in all cases P's expected profits, $\pi^{(A)}$–$\pi^{(F)}$, are weighted sums of output and effort cost (with constant weights). Thus, by our model assumptions we can confirm that the optimal requested efforts are positive. An exception is $\pi^{(D)}$: Here, a small amount is subtracted that is independent of c and f. Although the optimal (smallest) amount ϵ does not exist, it can always be chosen such that positive effort strictly dominates zero effort. □

Proof of Lemma 7. In the proof of Lemmas 1 to 6 we derived the maximum profits in cases (A) to (F) (denoted by $\pi^{(A)}$ to $\pi^{(F)}$, respectively) for given efforts. **1.** A comparison of (97) with (90) and (78) shows that, for the same effort level, $\pi^{(F)} > \pi^{(D)} + \pi^{(E)}$: The first term of $\pi^{(F)}$ is larger than $\pi^{(D)}$ and the two remaining terms in $\pi^{(F)}$ are equal to $\pi^{(E)}$. Since $q \in (0, 1)$, the assertion already holds if one inserts the optimal efforts of (D) and (E) in $\pi^{(F)}$. Thus, contracts as in (D) and (E) are strictly dominated by (F). **2.** The first equality in (13) obviously holds, by inspection of (38), (50) and (64). It is straightforward to show that, for any positive effort e, $\pi^{(A)} \gtreqless \pi^{(B)}$ iff $\lambda \gtreqless (1-q)^2/q^2$, which gives us the second equality in (13). **3.** By inspection of (38), (50), (64) and (97), it is obvious that by raising q, (A) (and thus (C)) monotonically approaches the first-best solution (i.e. the profit function $f - c$) while (B) becomes arbitrarily bad. (F) approaches the first-best solution as well since P finds it optimal to set $e_2 = 0$ (implying $c_2 = f_2 = 0$) if q approaches 1. If, instead, one lowers q, then it is, again, obvious that (B) (and thus (C)) monotonically approaches the first-best solution, while now (A) becomes arbitrarily bad. In contrast to the preceding argument, (F) is outperformed by (B) and (C) as soon as q is sufficiently small (since $\pi^{(F)}$

approaches $f_2 - 2c_2$ if q tends to zero). In order to see that (F) is optimal if the maximum profits of (A) and (B) are sufficiently close to each other, consider the following argument. Suppose $\lambda = (1-q)^2/q^2$. This is the case where (A), (B) and (C) perform worst, i.e., $\pi^{(A)} = \pi^{(B)} = \pi^{(C)} = f - c/q$. Denote the optimal effort for these contracts by e^{ABC} and in $\pi^{(F)}$, set $e_1 = e_2 = e^{ABC}$. Then $\pi^{(F)}$, see (97), simplifies to $f - c/q$ as well. However, neither e_1 nor e_2 have been chosen optimally for (F). In fact, it is obvious from (97) that they are generally different from each other. Thus, the optimal contract for (F) outperforms (A), (B), and (C). The following example illustrates this result. It also shows that (A) can be globally optimally if q is sufficiently large. Suppose $f(e) = e$ and $c(e) = e^{10}$. Figure 1 plots the maximum profits for (A), (B), (C) and (F). It is drawn for $\lambda = 2/3$ and it also shows the first–best profit (FB). The figure looks very similar for, e.g., $f(e) = \sqrt{e}$ and $c(e) = e^2/2$. □

References

Baker, G., Jensen, M. C. and Murphy, K. J.: 1988, Compensation and incentives; practice vs. theory, *Journal of Finance* **3**, 593–616.

Bol, J. C.: 2011, The determinants and performance effects of managers' performance evaluation biases, *The Accounting Review (forthcoming)* .

Faure-Grimaud, A., Laffont, J. J. and Martimort, D.: 2003, Collusion, delegation and supervision with soft information, *Review of Economic Studies* **70**, 253–279.

Grund, C. and Przemeck, J.: 2012, Subjective performance evaluation and inequality aversion, *Applied Economics* **44 (17)**, 2149–2155.

Hall, B. J. and Madigan, C.: 2000, Compensation and performance evaluation at Arrow electronics, *Harvard Business School Case 800-290* .

Harris, M. M.: 1994, Rater motivation in the performance appraisal context: A theoretical framework, *Journal of Management* **20**, 737–756.

Kane, J., Bernardin, H., Villanova, P. and Peyrefitte, J.: 1995, Stability of rater leniency: Three studies, *Academy of Management Journal* **38**, 1036–1051.

Laffont, J.-J. and Martimort, D.: 1997, Collusion under asymmetric information, *Econometrica* **65**, 875–911.

Laffont, J.-J. and Martimort, D.: 2001, *The Theory of Incentives: The Principal–Agent Model*, Princeton University Press.

Laffont, J. and Tirole, J.: 1986, Using cost observation to regulate firms, *Journal of Political Economy* **94**, 614–641.

Lee, S. and Persson, P.: 2010, Authority versus loyalty: Social incentives and modes of governance, *NYU Working Paper No. FIN-10-001* .

Lewis, T. R. and Sappington, S. E. M.: 1997, Information management in inventive problems, *Journal of Political Economy* **105**, 796–821.

MacLeod, W. B.: 2003, Optimal contracting with subjective evaluation, *American Economic Review* **93**, 216–240.

MacLeod, W. B. and Parent, D.: 1999, Jobs characteristics and the form of compensation, *Research in Labor Economics* **18**, 177–242.

McAfee, P. and McMillan, J.: 1987, Competition for agency contracts, *Rand Journal of Economics* **18**, 296–307.

Medoff, J. L. and Abraham, K. G.: 1980, Experience, performance, and earnings, *The Quarterly Journal of Economics* **95**(4), 703–36.

Murphy, K. J.: 1992, Performance measurement and appraisal: Motivating managers to identify and reward performance, *in* W. Bruns (ed.), *Performance Measurement, Evaluation, and Inventives*, Harvard Business School, pp. 37–62.

Prendergast, C.: 1999, The provision of incentives in firms, *Journal of Economic Literature* **37**(1), 7–63.

Prendergast, C.: 2002, Uncertainty and incentives, *Journal of Labor Economics* **20**, S115–S137.

Prendergast, C. and Topel, R.: 1993, Discretion and bias in performance evaluation, *European Economic Review* **37**, 355–365.

Prendergast, C. and Topel, R.: 1996, Favoritism in organizations, *Journal of Political Economy* **104**, 958–978.

Rotemberg, J. J.: 2009, Attitude–dependent altruism, tournout and voting, *Public Choice* **140**, 223–244.

Rotemberg, J. J.: 2011, Fair pricing, *Journal of the European Economic Association (forthcoming)* .

Rotemberg, J. J. and Saloner, G.: 1993, Leadership style and inventives, *Management Science* **39**, 1299–1318.

Sliwka, D.: 2007, Trust as a signal of a social norm and the hidden costs of incentive schemes, *American Economic Review* **97**, 999–1012.

Smith, B. N., Hornsby, J. S. and Shirmeyer, R.: 1996, Current trends in performance appraisal: An examination of managerial practice, *Advanced Management Journal* **61 (3)**.

Strausz, R.: 1997, Delegation of monitoring in a principal-agent relationship, *Review of Economic Studies* **64**, 337–357.

Tirole, J.: 1986, Hierarchies and bureaucracies, *Journal of Law, Economics, and Organization* **2**, 181–214.

Vafai, K.: 2004, Delegation and opportunism, *Journal of Institutional and Theoretical Economics* **160**, 498–521.

von Siemens, F.: 2011a, Heterogeneous social preferences, screening, and employment contracts, *Oxford Economic Papers* **63 (3)**, 499–522.

von Siemens, F.: 2011b, Social preferences, sorting, and competition, *Scandinavian Journal of Economics (forthcoming)* .