**Preprints of the
Max Planck Institute for
Research on Collective Goods
Bonn 2012/14**

Communicating Subjective Evaluations

Matthias Lang

MAX PLANCK SOCIETY

Preprints of the
Max Planck Institute
for Research on Collective Goods Bonn 2014/14

Communicating Subjective Evaluations

Matthias Lang

September 2009, this version April 2017

Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn
http://www.coll.mpg.de

Communicating Subjective Evaluations*

Matthias Lang**

April 2017

Abstract

Consider managers evaluating their employees' performance. Should managers justify their subjective evaluations? To scrutinize this question, I model justification: Suppose a manager's evaluation is private information. To justify her evaluation, she can gather additional information that is uninformative about her employee's effort. I show that the manager justifies her evaluation if and only if the evaluation indicates bad performance. The justification assures the employee that the manager has not distorted the evaluation downwards. For good performances, however, the manager pays a constant high wage without justification. Empirical literature demonstrates that subjective evaluations are lenient and discriminate poorly between good performances. I show that both effects occur in optimal contracts without any biased behavior.

JEL classifications: D82, D86, J41, M52

Keywords: Communication, Justification, Subjective Evaluation, Information Acquisition, Centrality, Leniency, Disclosure

1 Introduction

This paper analyzes communication in a principal-agent model in which the principal's performance measure is unobservable to the agent and nonverifiable by third parties. As verifiable, i.e., objective, performance measures are often unavailable, such subjective measures are widely used in practice.[1] Their subjectivity allows the principal to choose whether and how to disclose and to justify her evaluation of the agent's work. Hence, a hold-up problem arises: The principal wants to incentivize the agent to exert work effort, but these incentives depend on an appropriate evaluation in the end. Therefore, HR departments and personnel policies place great emphasis on feedback and communication of evaluations.[2] Nevertheless, subjective evaluations are lenient

*I thank Helmut Bester, Yves Breitmoser, Christoph Engel, Leonardo Felli, Ludivine Garside, Thomas Giebe, Alia Gizatulina, Olga Gorelkina, Dominik Grafenhofer, Paul Heidhues, Martin Hellwig, Sergei Izmalkov, Sebastian Kodritsch, Johannes Koenen, Daniel Krähmer, Marco Ottaviani, Régis Renault, Ilya Segal, Caspar Siegert, Roland Strausz, and Stefan Terstiege for very helpful discussions, and the audiences at the World Congress of the Game Theory Society, Econometric Society NASM and EM, CESifo Applied Micro, EEA annual meeting and seminars in Milan (Bocconi), Mannheim, Berlin, Bonn, Maastricht and Munich for comments.

**Free University Berlin, Dept. of Economics, Boltzmannstr. 20, Berlin, Germany, lang@uni-bonn.de

[1]The extensive use of subjective performance measures is confirmed by Gibbs et al. (2009), Dessler (2008, p. 339), Porter et al. (2008, p. 148), Levin (2003), MacLeod and Parent (1999), and Murphy (1993). The reason is that agents can manipulate objective performance measures or multitask problems. Consequently, Gibbons (1998, p. 120) concludes that "objective performance measures typically cannot be used to create ideal incentives."

[2]See, for example, Dessler (2008, Chapter 9) or Porter et al. (2008, Chapter 8).

and wage dispersion for the best evaluations is low.[3] These empirical observations are referred to as leniency bias and centrality bias, respectively – supposedly arising from supervisors' mistakes. In response, a whole industry has sprung up to provide training for supervisors. Alternatively, "some companies go so far as to rate employees on a bell curve," requiring supervisors to match a given distribution with their evaluations, as forcefully advocated by, e.g., Jack Welch, a former and very renowned CEO of General Electric (New York Times, 2013).[4] I show that both responses may be misplaced, because leniency and centrality are optimal and do not require any bias. Both effects result from optimal contracting with standard preferences.

In addition to explaining leniency and centrality, the methodological contribution of this paper is to model justification. By justification, I refer to a message that transmits information previously unknown by the recipient and that is partially verifiable by the recipient. In the model, an agent (he) works for a principal (she) who privately receives information about the agent's performance, like reports from colleagues, observations of the agent at work or of the agent's output. By random encounters or joint observations, the agent learns a very small fraction of the principal's information. These shared signals, however, are uninformative about the evaluation of the agent's work by the principal. The principal has two options. Either she reports directly her evaluation or she justifies her evaluation by acquiring additional information. Justification of subjective evaluations is a common HR practice: "92% require a review and feedback session as part of the appraisal process." (Dessler, 2008, p. 366) The principal's message is not necessarily truthful and providing justification is costly. The agent replies with an unverifiable message about the shared signals. As the messages are the only third-party enforceable information, the contract just depends on these messages. I study the resulting communication pattern: in equilibrium the principal justifies only bad evaluations. In this case, wages increase in the evaluation. For good evaluations, the principal in equilibrium saves the hassle of explaining them and simply pays a high wage. This yields pooling and wage compression at the top: leniency, i.e., expected evaluations exceed expected performance, and centrality, i.e., variation in performance exceeds variation in wages at the top, arise endogenously from optimal contracting.

The intuition for this communication pattern is: First, it is never optimal to justify all evaluations, because justification is costly. Second, if the agent is evaluated positively, he suspects no deviation by the principal, because the principal has to pay higher wages for better evaluations. If the agent is evaluated negatively, the agent considers two possibilities: his performance was bad or the principal distorted her evaluation downwards to pay lower wages. To counter such

[3]For example, Suvorov and van de Ven (2009, p. 666) state that "compression of performance ratings is well documented." According to Bretz et al. (1992), 60–70% of employees get an evaluation from the best or second-best category. In addition, Murphy (1993, p. 56) reports that the top 1% of employees at a pharmaceutical company receive a pay raise just 3% higher than the median employee. The finding goes back to Taylor and Wherry (1951, p. 39). They find "a marked distortion [of evaluations] . . . with considerably poorer discrimination at the top." More recently, Puhani and Yang (2017), Kampkötter and Sliwka (2017), Golman and Bhatia (2012), and Spence and Keeping (2011) also document and confirm leniency and centrality of subjective evaluations.

[4]*New York Times*, 2013 Nov. 24, Invasion of the annual reviews, Business News p.8. See also *Wall Street Journal*, 1999 Jun. 21, Raises and Praise or Out the Door: How GE's Chief Rates and Spurs His Employees, p. B1.

suspicions, the principal justifies bad evaluations. Notice that compared to common moral-hazard settings additional incentives are necessary: ex-ante the principal wants to justify bad evaluations ex-post. Nevertheless, ex-post she wants to save on justification costs. The principal has no commitment power other than the contract. Hence, she has to design contractual terms that make it ex-post incentive compatible for her to justify the evaluation. Finally, there is a clear intuition for centrality: the agent cannot verify the evaluation without justification. Hence, without justification the principal reports an evaluation yielding lower payments. Therefore, no wage dispersion is feasible and there is pooling if the principal provides no justification.[5]

To show the strength and power of justification, I limit the agent's information as far as possible. Nevertheless, the agent can to some extent verify the principal's message although the evaluation and the agent's information are stochastically independent and, hence, uncorrelated. In contrast to previous literature, I do not assume an exogenous verification technology, type-dependent message spaces, or that messages are verifiable by a third party. All messages are unverifiable, but contractible. Hence, a third party cannot tell whether a message is truthful. The mechanism uses the fact that the principal and the agent share some observations of the environment and the processes that lead to the evaluation. These shared observations are uninformative about the evaluation and have mass zero with respect to the principal's information resulting in the evaluation. It is easy to see that justification also works if the agent has more information. The principal recalls all observations to justify the evaluation. By adding this additional information to her evaluation, she makes herself vulnerable to scrutiny. If she were to distort the evaluation, she has to lie about some observations. No matter how she distorts the evaluation, there is a strictly positive probability that the agent becomes aware of any distortion that negatively affects his wages. The reason is that the principal does not know which observations the agent has learned. As an example, consider a business analyst at a consulting firm who received praise from a client. Suppose the partners in the firm pretend that the analyst's work was bad and justify their assessment saying that the client complained about the analyst. Although the business analyst is unaware of the partners' real evaluation of her work, she knows that the partners are lying. My model of justification is much more general and not limited to employment relations. It applies more generally to mechanism design, moral hazard, and hold-up settings whenever contracting parties interact and share some information. The mechanism allows the better-informed party to provide justification.

Finally, return to the empirical observations of leniency and centrality. Both effects vanish in evaluations for development or feedback instead of wage-setting, as shown, e.g., by Dessler (2008, p. 356), Milkovich et al. (2008, p. 351), and Jawahar and Stone (1997). This observation is in line with my predictions. The principal requires incentives to report her evaluation for wage-setting truthfully. These incentives cause pooling of the best evaluations. If the evaluation is

[5] Murphy (1993, p. 49) summarizes the reasoning as follows: Principals have "nonpecuniary costs [here, information-acquisition costs] associated with performance appraisal, which leads them to prefer to assign uniform ratings rather than to carefully distinguish employees by their performance."

for development or feedback, these incentives are unnecessary, as the preferences of the principal and the agent with respect to the allocation of training are likely to be better aligned than those with respect to wages. Managers at Merck, for example, experienced that "the salary link made discussions on performance improvement difficult." (Murphy, 1993, p. 58) Psychological costs of supervisors to give bad evaluations to their subordinates yield no straightforward explanation of this pattern, since those costs should apply to evaluations for all purposes similarly.

The remainder of this paper is organized as follows. Section 2 discusses the related literature. Section 3 conveys the basic intuition in a reduced-form model. Section 4 sets up the model and characterizes the optimal communication pattern. Section 5 examines the robustness of the model. Section 6 contains the concluding remarks. All proofs are relegated to the appendix.

2 Related Literature

As in Al-Najjar et al. (2006) and Anderlini and Felli (1994), I explicitly model certain features of a language. In their papers, restrictions of the contracting language make it ex-ante impossible to describe some events that are observable to all contracting parties ex-post. These restrictions make incomplete contracts optimal. In my paper agents can write any contract ex-ante. Yet, the state of the world is private information and needs to be communicated ex-post. This communication can be supplemented by justification that makes the principal's message partly verifiable. Although I use a similar representation of the states of the world as an infinite binary sequence, their approach is conceptually and technically different from what I do here. I illustrate my model of providing justification using a setting with subjective performance measures.

There is a long literature on subjective performance measures. Usually, it is assumed that evaluations are observable and relationships are long-term. This yields implicit contracts, like for example in Goldluecke and Kranz (2013), Compte (1998), Baker et al. (1994), and MacLeod and Malcomson (1989). Then reputation effects created by the continuation value for both contracting parties allow subjective performance measures to gain credibility and to be used for the agent's incentives. Li and Matouschek (2013) and Levin (2003) drop the assumption that subjective performance measures are perfectly observable by both contracting parties. In this case, optimal contracts often have a termination form, i.e., contracts end after observing bad performance. See also Malcomson (2012) and MacLeod (2007) for recent surveys. In contrast to these repeated interactions, subjective evaluations are also used in static settings.

MacLeod (2003) was the first to implement subjective performance measures in a static setting. He assumes that the agent has a signal that is correlated with the principal's evaluation. Each party reports their information by simultaneously sending a public message. As the information structure is exogenously given, the principal cannot decide, depending on the performance measure, whether to justify her evaluation. Yet, his results correspond to two special cases of my model. If the agent's and the principal's signal are correlated, MacLeod (2003) achieves the common second-best solution similarly to obligatory or costless justification in my model in Lemma 2. If signals are uncorrelated, the optimal contract in MacLeod (2003) resembles the case of prohibitively expensive justification in my model. Economically, the main difference between my paper and MacLeod (2003) is that I endogenize communication. This allows me to discuss the resulting communication pattern. Nevertheless, the case of imperfect correlation with a binding upper limit on third-party payments by MacLeod (2003) shares some features with my optimal contracts, but the reasoning and the proofs are different. First, I do not assume an upper limit on payments. Second, the agent receives no private signals telling him that he received no information. Instead, it is the principal's incentive – resulting from the contract and justification costs – to withhold and distort her evaluation that yields the compression at the top result. MacLeod and Tan (2016) extend the model of MacLeod (2003) by considering malfeasance and more general information structures between agent and principal, like better-informed agents. In addition, they change the timing and study sequential messages with the agent or the principal sending their message first. In this dichotomy, I scrutinize authority contracts with the principal reporting first, although with justification as a different communication technology.

In the current paper, I follow a static approach. Some justification can be found in Fuchs (2007) who considers a finitely repeated principal-agent model. He shows that it is optimal for the principal to announce her subjective evaluation only once at the end of the interaction. In this case, the agent does not learn whether a good performance has already occurred. Hence, it is sufficient to penalize only the worst outcome, while paying a constant wage following all other terminal histories. Brown and Heywood (2005) and Addison and Belfield (2008) provide additional justification for a static approach. They show empirically that performance evaluations are more likely to be used for employees with shorter expected tenure.

This paper also relates to the literature on endogenous contracts, like Kvaløy and Olsen (2009). Yet, I do not assume any cost for writing specific contractual arrangements. The contract can be any functions of the messages, but justification is costly. Another paper discussing endogenous verifiability is Dewatripont and Tirole (2005). They allow both sides to exert effort to increase the probability of a verifiable message. Yet, they do not consider moral hazard and subjective evaluations. As justification allows partially verifying the performance measure, there is a parallel to the literature on costly state verification, like Hart and Moore (1998), Gale and Hellwig (1985), and Townsend (1979). These models allow an investor to verify the firm's performance at a cost. They show the optimality of debt contracts, which are similar to optimal contracts in my paper, as there is no verification for high payments. In this literature, however, contracts provide no incentives for the agent and verification is contractible, so that payments can depend on whether verification occurred as in Townsend (1979) and contracts specify when to verify as in Gale and Hellwig (1985). In my model, justification is not contractible. Hence, contracts cannot enforce justification directly and payments cannot depend on whether justification was provided. The reason is that justification need not be truthful and cannot be verified directly by one of the contracting parties, while verification is truthful and verifiable. This is also the reason why mixed strategies with respect to justification are not optimal in my setting in contrast to, e.g., Townsend (1979).

To make deterministic verification strategies optimal, Krasa and Villamil (2000) dynamically extend costly state verification. An investor can verify the firm's performance at a cost. Yet, the contract can be renegotiated, after the firm learned the state of the world. Simple debt contracts are optimal in this setting, dominating stochastic contracts. In my model, there is commitment to a contract. Hence, renegotiations are impossible. Yet, it has to be sequentially optimal for the better-informed side to provide justification. Hence, equilibrium wages without justification have to be higher than justified equilibrium wages. In Krasa and Villamil (2000), the less-informed investor verifies the firm's payments if these payments are below a threshold. This is another distinction between the literature on costly state verification and my paper. In my model, due to the nature of justification, the better-informed side chooses whether to provide justification. In contrast, the less-informed side usually chooses whether to verify. For example, Doornik (2010) applies costly state verification to moral hazard. She considers a setting where output is contractible, but private information of the principal. The principal offers the agent a payment. If the less-informed agent rejects the principal's offer, output is verified at a cost to both sides and the agent's wage is determined according to the realized output. In an optimal contract, the agent stochastically triggers verification for low wage offers. The agent is willing to verify, because contracts specify higher wages following verification. This is feasible, because contracts map messages, the fact whether verification occurred, and, if verified, the state of the world into wages. Such a contract is not feasible in my setting, as only messages are contractible. Doornik (2010) has pooling at the bottom and at the top. Again, in my model, it is the better-informed party that makes the verification decision and optimal justification is deterministic.

Furthermore, wage compression or centrality of evaluations could also be caused by fairness or trust. According to Bernardin and Orban (1990, p. 197) "trust in appraisal accounted for a significant proportion of variance in performance ratings." In my model, justification establish this trust. In Giebe and Gürtler (2012), Al-Najjar and Casadesus-Masanell (2001), and Rotemberg and Saloner (1993), this trust is created by the extent to which the principal's preferences incorporate the agent's well-being in contrast to standard preferences in my model.

Several papers consider different rationales for subjective evaluations, namely, as a signal about the agent's productivity. If the agent does not know her productivity, she can infer her productivity from the evaluation by the principal. Fuchs (2014) shows that if the principal pays a discretionary bonus for positive evaluations, the bonus payment makes her evaluation credible and allows for a separating equilibrium. Zábojník (2014) uses subjective evaluations to fine-tune the agent's effort choice in a multitasking setting. Subjective evaluations can supplement imperfect objective measures and allow the agent to learn her productivity. Suvorov and van de Ven (2009) analyze subjective evaluations as a signal about the agent's productivity if the agent is intrinsically motivated.

3 Basic Intuition and Reduced-Form Model

Here I study the basic intuition in a reduced-form model where justification directly makes the performance measure partially verifiable. Section 4 considers the full model that drops this assumption and reduces justification to an exchange of messages that need not be truthful and which cannot be verified directly by one of the parties.

3.1 Setting

Consider a risk-averse agent working for a risk-neutral principal. The principal proposes a contract to the agent. The contract specifies payments depending on messages as described later. After signing such a contract, the agent exerts effort $e \in [0, 1)$, which is unobservable by the principal. Then the principal privately learns the evaluation $p \in \mathcal{P}$ with a finite set $\mathcal{P} \subset (0, 1]$ that has at least three elements. The evaluation p is drawn from a distribution $F(p|e) = eF^H(p) + (1-e)F^L(p)$ depending on the agent's effort e. The distributions $F^H(p)$ and $F^L(p)$ have associated probability measures $f^H(p), f^L(p) > 0$ for all $p \in \mathcal{P}$. In addition, the ratio $f^H(p)/f^L(p)$ strictly increases in p. Therefore, the distribution $F(p|e)$ satisfies the monotone likelihood ratio property ensuring that a higher p indicates higher work effort. It proves helpful to use $p_{min} = \min \mathcal{P}$ and $p_{max} = \max \mathcal{P}$.

The principal communicates the evaluation p by sending a message m_P. For this purpose, she has two options: she either provides justification or she does not. If she does not provide justification, her message space is $\mathcal{P}$. If she provides justification, she pays costs κ to make her message partially verifiable and, hence, her message space is $[p, 1] \cap \mathcal{P}$. I drop this assumption in the full model in Section 4. Let $\beta \in \{0, 1\}$ denote the principal's justification decision. For $\beta = 1$, she justifies her evaluation of the agent's work. Independently of the principal's choice, but after observing this choice, the agent replies with an unverifiable message $m_A \in \{0, 1\}$. Both parties can lie and send any message from the corresponding message spaces. Finally, the contract is performed according to the messages m_P and m_A. The contract specifies the payments made by the principal $W(m_P, m_A)$ and the agent's wage $C(m_P, m_A)$ depending on these two messages. As MacLeod (2003, Proposition 2), Fuchs (2007, Proposition 1) and MacLeod and Tan (2016, Section 2.3) demonstrate, some surplus has to be destroyed in this kind of model to implement positive effort by the agent. An alternative is to use stochastic contracts as in Lang (2017). I follow the first approach and allow for $W(m_P, m_A) \geq C(m_P, m_A)$. The principal has no commitment power other than the contract.

There is a $B > 0$, such that the principal's benefit is $Bp - w - \beta\kappa$ if she pays a wage w. The agent's preferences are represented by $u(w) - d(e)$ if he chooses effort e and receives a wage w. I assume $\lim_{w \to 0} u(w) = -\infty$ with derivatives $u' > \epsilon > 0$ and $u'' < 0$. The disutility d of exerting effort is increasing and strictly convex with the limit $\lim_{e \to 1} d(e) = \infty$. Both functions are twice continuously differentiable. The agent receives a reservation utility $\bar{u}$ if he rejects the contract.

3.2 Analysis

Lemma 6 in the appendix shows that restricting attention to contracts with truthful revelation is without loss of generality. Therefore, I restrict attention to contracts with truthful revelation in this section. I discuss indirect mechanisms at the end of Section 4.[6] Grossman and Hart (1983) show that the model can be solved in two steps. First, for every level of effort e, an optimal contract and its expected costs $\Pi(e)$ for the principal are computed. The second step determines optimal effort levels e by solving

$$\max_{e\in[0,1)} B\sum pf(p|e)-\Pi(e).$$

Returning to the first step, Program A below determines an optimal contract that implements effort e by choosing payments $W(m_P,m_A)$ by the principal and $C(m_P,m_A)$ for the agent. These payments implicitly determine which evaluations to justify, $\beta(p)$. The objective is to minimize expected costs subject to several conditions. The participation constraint PC makes the agent accept the proposed contract. The agent's incentive compatibility IC guarantees that the agent chooses the desired level of effort. In addition, truth-telling has to be incentive compatible for the principal (TT_P) and the agent (TT_A). Truth-telling implies $m_P=p$, $\beta=\beta(p)$ and $m_A=\beta$. Finally, the principal's payment has to be higher than the wage for the agent.

$$\Pi(e)=\min\sum\nolimits_{p\in\mathcal{P}}(W(p,\beta(p))+\kappa\beta(p))f(p|e) \tag{A}$$

$$\text{subject to} \sum\nolimits_{p\in\mathcal{P}} u(C(p,\beta(p)))f(p|e)-d(e)\geq\bar{u} \tag{PC}$$

$$e\in\arg\max\sum\nolimits_{p\in\mathcal{P}} u(C(p,\beta(p)))f(p|e)-d(e) \tag{IC}$$

$$\kappa\beta(p)+W(p,\beta(p))\leq\kappa\tilde{\beta}+W(m_P,\tilde{\beta}) \tag{TT_P}$$

$$\forall(\tilde{\beta},m_P)\in\big\{\{0\}\times\mathcal{P},\{1\}\times([p,1]\cap\mathcal{P})\big\},\forall p\in\mathcal{P}$$

$$C(p,\beta)\geq C(p,m_A) \qquad \forall p\in\mathcal{P},\forall\beta,m_A\in\{0,1\} \tag{TT_A}$$

$$W(m_P,m_A)\geq C(m_P,m_A) \qquad \forall m_P\in\mathcal{P},\forall m_A\in\{0,1\}. \tag{1}$$

If the principal's information p were observable and contractible, we can neglect the principal's and the agent's truth-telling constraints, TT_P and TT_A.

Lemma 1. *If evaluations are contractible, optimal wages $w_e^*(p)$ only depend on the evaluation p. Wages $w_e^*(p)$ increase in p for positive effort, $e>0$. There is no justification.*

If the principal's information is subjective and justification is a choice of the principal, the additional incentive constraints for truth-telling do matter. Yet these incentive constraints do not change the equilibrium wage if there are no justification costs.

Lemma 2. *If there are no justification costs and $\kappa=0$, there is an optimal contract in which all evaluations are justified. Equilibrium wages are the same as in Lemma 1.*

[6] Notice that I prove Lemma 1 and 2 as well as Proposition 1 and 2 directly without using such revelation-principle arguments.

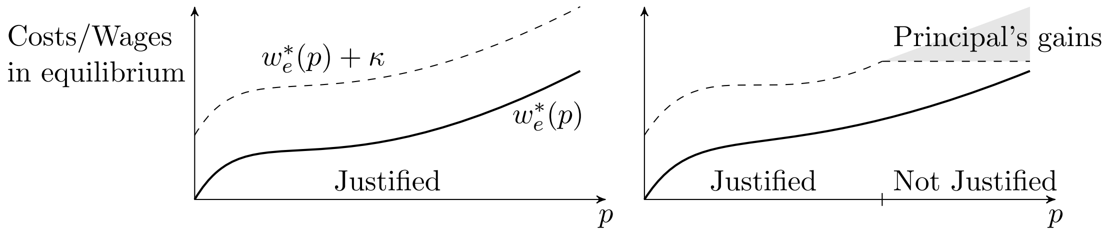

Figure 1: Idea of the Proof of Proposition 1

The left-hand side illustrates the optimal contract if all evaluations are justified. The right-hand side shows a contract in which all, but the best evaluations are justified. Thick lines denote the agent's wage in equilibrium, while dashed lines depict the principal's total costs including justification costs.

The lemma constructs the optimal contract that is unique up to out-of-equilibrium payments. Justification makes the principal's evaluation partially verifiable, so that the wage in equilibrium resembles the wage in the contractible setting. Whenever justification is costly and $\kappa > 0$, however, it is no longer optimal to justify all evaluations. To gain some intuition, suppose to the contrary that the principal justifies all evaluations. Then optimal contracts imply wages $w_e^*(p)$ in equilibrium as in Lemma 2. Yet, the principal can modify this contract to save on justification costs. The reason is that the agent does not suspect a distorted evaluation by the principal for the highest wages. Therefore, it is not optimal to justify all evaluations.

Proposition 1. *If justification is costly and $\kappa > 0$, justifying all evaluations is not optimal: In an optimal contract, there is a $p \in \mathcal{P}$ with $\beta(p) = 0$.*

The proof shows that the principal's total costs decrease if she refrains from justifying the best evaluations. By paying a high wage that is not justified, she can reduce justification costs. These efficiency gains go partly to the principal as indicated by the gray ares in Figure 1. As detailed in the proof, the new contract satisfies all constraints in Program A. Therefore, it is not optimal to justify all evaluations. It remains to determine which evaluations to justify. To find the optimal justification strategy, however, we need to know more about optimal contracts.

Proposition 2. *If an optimal contract exists, there is an optimal contract $\mathcal{W}^{**}$ and a set $\mathcal{C} \subseteq \mathcal{P}$, such that only evaluations in $\mathcal{C}$ are justified: $\beta(p) = 1$ if and only if $p \in \mathcal{C}$. In addition,*

$$C^{**}(m_P, m_A) = c(m_P)$$

$$W^{**}(m_P, m_A) = \begin{cases} w^{**}(m_P) & \text{if } m_P \in \mathcal{C} \text{ and } m_A = 1 \\ w^{**} & \text{else.} \end{cases}$$

Payments increase in the evaluation, because $w^{**}(m_P)$ is non-decreasing in $m_P \in \mathcal{C}$. The principal justifies low payments, as $w^{**}(m_P) + \kappa \le w^{**}$ for all $m_P \in \mathcal{C}$.

If the principal does not provide justification, the agent cannot verify the evaluation by the principal. Therefore, the principal's payments have to be constant in the absence of justification. Hence, there is pooling of wages with a wage of w^{**}. For the same reason, the pooling wage has to be higher than wages with justification. Otherwise, the principal would deviate by not justifying the evaluation and paying the pooling wage. The contract cannot detect such a deviation. The high pooling wage guarantees that this deviation is unprofitable.

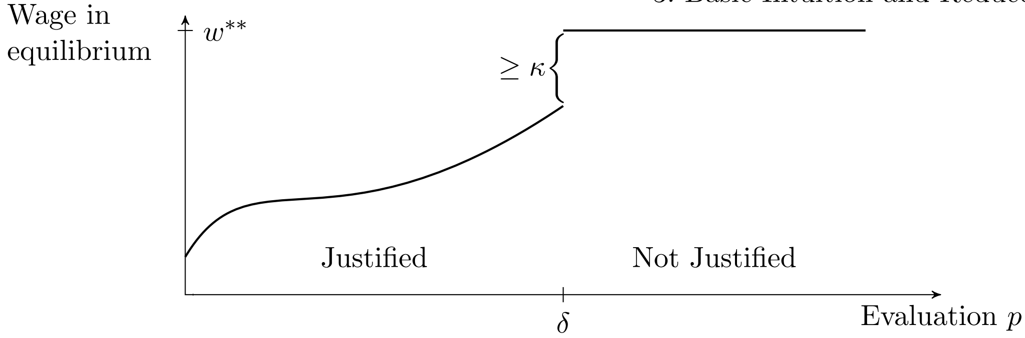

Figure 2: Proposition 3 and the Equilibrium Wage

The optimal justification strategy follows a threshold rule. Bad evaluations are justified, good evaluations are reported without justification. For ease of exposition, I draw wages as continuous, although they are discrete.

Now turn to my main result: Optimal justification strategies follow a threshold rule. The principal justifies only bad evaluations. In these cases, the agent suspects a distortion by the principal who can alleviate these fears by providing justification.

Proposition 3. *The principal optimally justifies evaluations p up to a threshold $\delta \in [0, 1)$, while she does not justify evaluations above δ.*

The proposition proves the communication pattern described in the introduction and summarized in Figure 2. The principal justifies only bad evaluations and low wages, while she remains silent on good performance. Optimal contracts reward evaluations at the top similarly. Thus, the contract eliminates wage differences that the principal would have to justify.

Before turning to formal arguments, consider the following intuition. If the optimal communication pattern does not follow a threshold rule as in Proposition 3, there were evaluations p_L and p_H such that $p_L < p_H$ and the principal justifies p_H, but does not provide justification for p_L. Such a communication pattern implies that equilibrium payments decrease in the evaluation p according to Proposition 2. The monotone likelihood ratio property ensures that decreasing equilibrium payments are not optimal. Hence, a threshold rule is optimal. More formally, such a communication pattern implies that equilibrium payments for evaluation p_H plus κ equal equilibrium payments for evaluation p_L. Adjust the contract so that the principal does not justify p_H. At the same time decrease the agent's wage for p_L and increase the wage for p_H such that the agent's expected utility remains constant. Hence, the agent's participation constraint PC is still satisfied. I show that the agent's incentive compatibility IC is slack in this modified contract due to the monotone likelihood ratio property. Therefore, the initial communication pattern cannot be optimal and the principal justifies only bad evaluations.

If justification costs κ are prohibitively high, it is not optimal to use justification and, hence, $\delta = 0$. To characterize the threshold for justification costs, I first determine the optimal contract without justification in Lemma 8 in the appendix. This contract is unique up to out-of-equilibrium payments. I just have to characterize when this contract is optimal and when it is optimal to use justification.

Proposition 4. *The principal optimally justifies some evaluations if she wants to implement positive effort $e > 0$ and justification costs κ are at most $\bar{\kappa} > 0$ with*

$$\bar{\kappa} = u^{-1}\left(\bar{u} + d(e) + \frac{f(p_{min}|e)d'(e)}{f^L(p_{min}) - f^H(p_{min})}\right) - u^{-1}\left(\bar{u} + d(e) - \frac{(1-f(p_{min}|e))d'(e)}{f^L(p_{min}) - f^H(p_{min})}\right).$$

The principal does not justify any evaluations if justification costs κ are higher than $\bar{\kappa}$.

Justifying evaluations makes the principal's promise of incentives to the agent credible and allows the principal to assure the agent that her evaluation is not distorted. Therefore, the principal wants to be transparent about her evaluations, in particular, bad evaluations, even if justifying evaluations is costly and takes place after the agent's effort choice. It remains to determine wages w^{**} and $w^{**}(p)$ from Proposition 2 to characterize optimal contracts.

Proposition 5. *Suppose $\kappa < \bar{\kappa}$ and $e > 0$. An optimal contract is determined by δ, w^{**} and $w^{**}(p)$:*

$$\min_{w(p),w,\delta} \sum_{p\leq\delta}(w(p)+\kappa)f(p|e) + (1-F(\delta|e))w \tag{C}$$

$$\text{subject to} \sum_{p\leq\delta} u\big(w(p)\big)f(p|e) + (1-F(\delta|e))u(w) - d(e) \geq \bar{u} \tag{PC}$$

$$\sum_{p\leq\delta} u\big(w(p)\big)(f^H(p) - f^L(p)) + u(w)\sum_{p>\delta}(f^H(p) - f^L(p)) = d'(e) \tag{IC}$$

The intuition for Proposition 5 is the following. Third-party payments can be avoided on the equilibrium path. In addition, the principal's truth-telling constraint is automatically satisfied for three reasons: Proposition 3 implies that justification follows a threshold rule and the contract in Proposition 5 specifies a constant wage for good evaluations. Moreover, the monotone likelihood ratio property implies that $w(p)$ is increasing in the evaluation p and that the pooling wage is above the wage for justified evaluations.

If justification costs vanish, more and more evaluations are justified. Yet, recall from Proposition 1 that it is never optimal to justify all evaluations if there are justification costs.

Proposition 6. *For $e > 0$ and sufficiently small justification costs, the principal justifies all, but the best evaluation, $\mathcal{C} = \mathcal{P} \setminus \{p_{max}\}$. The agent's wages equal $w_e^*(m_P)$.*

Similar arguments can be used to show that the number of justified evaluations $|\mathcal{C}|$ weakly decreases in the justification costs. Now turn to the full model without exogenous verifiability.

4 Evaluating the Agent's Work

4.1 Subjective Evaluations at Work

As motivation for my information structure, I begin with a brief case study. Consider performance evaluations at Arrow Electronics, a Fortune 500 company, as documented in Hall and Madigan (2000). Employees are evaluated in seven categories, capturing, for example, costumer satisfaction, their business judgment, or skills as a team worker. In each category, they receive a rating on a scale from one to five. The average rating across categories yields the result of the evaluation that is used for compensation purposes.

- At $t = 0$, principal proposes a contract to the agent.
- At $t = 1$, agent can accept the contract offer and determines effort e.
- At $t = 2$, principal learns evaluation p.
- At $t = 3$, principal either directly sends a report $m_P \in \mathcal{I}$ or learns I and provides justification $m_P \in \mathcal{I}$ at costs κ.
- At $t = 4$, agent sends a message $m_A \in [0, 1]$.
- At $t = 5$, principal makes payments according to $W(m_P, m_A)$ and $C(m_P, m_A)$.

Figure 3: Timing of the Model

Suppose the principal can acquire a signal in each category. For example, she listens to costumers praising the employee. She talks to the agent's colleagues to learn about his skills as a team worker. She observes the agent at work or the agent's output. This closely captures a practical evaluation process, as "an appraiser would use evidence from direct observation of the employee, or by reports from others, to make judgment about the appraisee's performance." (Porter et al., 2008, p. 149) These signals are subjective and private information of the principal. The agent, however, sometimes gets direct feedback from customers or is told by colleagues about their reports to the principal. Hence, he might observe one of the principal's signals.

Arrow Electronics requires managers to communicate evaluations. Suppose, the principal can choose either to tell the agent only the result of the evaluation or to justify the evaluation. Justification tells the agent all signals. Providing justification is costly, as it requires the principal to spend additional time on the evaluation.[7] The next section formalizes these notions.

4.2 Setting

The setting is the same as in Section 3 with three exceptions: the principal's and the agent's information as well as the message spaces of both players. Figure 3 summarizes the timing. As before, the principal learns the evaluation p. Then she decides whether to justify her evaluation. If the principal provides justification, she pays costs κ to privately receive signals $I(t) \in \{0, 1\}$ in different categories $t \in T = [0, 1]$. In each category t, the principal's signal declares success, $I(t) = 1$, or failure, $I(t) = 0$. Denote the set of successful categories by $I = \{t \in T | I(t) = 1\}$. Independently of her choice, she sends a message

$$m_P \in \mathcal{I} = \{X \subseteq [0, 1] | X \text{ is } \lambda\text{-measurable and } \lambda(X) \in \mathcal{P}\}$$

with the measure λ specified below. Also independently of the principal's choice, the agent observes one category S drawn uniformly from the successful categories I.[8] The principal does not

[7] Assume that the agent quits his job at Arrow Electronics afterwards. Indeed, turnover rates at Arrow Electronics could reach 20%-25%.

[8] In Section 5 the agent also learns some failed categories. In Rahman (2012), a principal instructs an agent to shirk sometimes creating shared observations between principal and monitor. These shared observations allow the principal to verify the monitor's report if the probability of shirking is positive. In my model of justification, it is sufficient that there are some shared observations, but they can be uninformative and have mass zero.

know which category the agent observes. Notice that the evaluation p and the agent's information S are stochastically independent. The agent cannot learn anything about the evaluation p from his information S. The agent replies to the principal's message with an unverifiable message $m_A \in T$. Both parties can lie and send any message from the corresponding message spaces.

As in Section 3, a probability $p \in \mathcal{P}$ is drawn from $F(p|e) = eF^H(p) + (1-e)F^L(p)$ depending on the agent's effort e. In each category $t \in T$, the principal's signal declares success, $I(t) = 1$, with probability p and failure, $I(t) = 0$, with probability $1 - p$. The signals $I(t)$ are essentially pairwise independent as defined by Sun (2006, Definition 2.7).

Lemma 3. *There is a probability space that satisfies these assumptions and guarantees a law of large numbers.*

Sun and Zhang (2009, Theorem 1) introduce an extension λ of the Lebesgue measure on T. Define the subjective evaluation as in the case study as the average $\mu = \int_T I(t)\mathrm{d}\lambda$ of the principal's signals $I(t)$ in the categories T. According to Sun (2006, Theorem 2.8), this average μ is well defined and equals p almost surely. Therefore, the evaluation μ is a sufficient statistics for the agent's effort.

4.3 Analysis

As in Section 3, I begin with a benchmark. If signals I were observable and contractible, only aggregate evaluations μ matter, but not the detailed assessments in the different categories.

Lemma 4. *If signals I are observable and contractible, optimal wages $w_e^*(\mu)$ depend only on the evaluation μ, i.e., the average of the principal's information I. Wages $w_e^*(\mu)$ are the same as in Lemma 1. There is no justification.*

As Holmström (1979) shows, with contractible information, the wage depends only on the sufficient statistics μ instead of the entire information I. In addition, better evaluations imply higher wages. If the principal's information is subjective and justification is the principal's choice, messages and justification choices do matter. If there are no justification costs, these additional incentives change optimal contracts, but do not change equilibrium wages.

Lemma 5. *If there are no justification costs and $\kappa = 0$, there is an optimal contract in which all evaluations are justified. Equilibrium wages are the same as in Lemmas 1, 2, and 4.*

The contract differs from the contracts in Lemmas 2 and 4. In particular, payments depend on whether messages agree, i.e., $m_A \in m_P$. On the equilibrium path, messages agree. Therefore, the equilibrium wage is $w_e^*(\mu)$ – the same as before. Whenever justification is costly and $\kappa > 0$, however, it is no longer optimal to justify all evaluations. The reasoning is the same as in Proposition 1.

Proposition 7. *If justification is costly and $\kappa > 0$, justifying all evaluations is not optimal: In an optimal contract, there is a $p \in \mathcal{P}$ with $\beta(p) = 0$.*

Next, determine optimal contracts using the intuition we gained in Section 3.

Proposition 8. *Suppose $\kappa < \bar{\kappa}$ and $e > 0$. The following contract is optimal:*

$$C^*(m_P, m_A) = \begin{cases} w^{**} & \textit{if } \lambda(m_P) > \delta \\ w^{**}(\lambda(m_P)) & \textit{if } \lambda(m_P) \leq \delta \end{cases}$$

$$W^*(m_P, m_A) = \begin{cases} w^{**} & \textit{if } \lambda(m_P) > \delta \\ w^{**}(\lambda(m_P)) & \textit{if } \lambda(m_P) \leq \delta \textit{ and } m_A \in m_P \\ j(w^{**} + \kappa) & \textit{else.} \end{cases}$$

Proposition 5 determines the values of δ, w^{**} and $w^{**}(p)$ for $p \leq \delta$.

The agent's wage is either constant or depends only on the measure of the principal's message m_P. Notice that, following truthful communication, the measure of the principal's message equals the measure of successful categories and, hence, the evaluation, i.e., $\lambda(m_P) = \lambda(I) = \mu$. If the principal provides justification, the agent can partially detect deviations by the principal. We say messages m_P and m_A agree if $m_A \in m_P$. Otherwise, they disagree. If the principal sends a truthful message $m_P = I$, messages agree with probability one. If messages agree, the principal's payment equals the agent's wage. If the principal reports more successful categories than the evaluation, $m_P \supset I$, messages agree with probability one, but this deviation increases the principal's payment. If the principal deviates on a set of categories with measure zero, messages still agree with probability one and the wage remains unchanged. If the principal deviates to a message m_P which yields a lower wage than a truthful report of her evaluation, messages disagree with positive probability. If messages disagree, the principal has to pay the highest payment in this contract. The proof shows that there is no profitable deviation for the principal. Hence, truth-telling is optimal.

The agent's wage equals the principal's payments in the absence of justification. Varying the agent's wage in the absence of justification means reducing the agent's wages. This reduction eases the agent's incentive compatibility IC, but makes it more difficult to satisfy the agent's participation constraint PC. The principal's costs remain unchanged. Yet, instead of varying the agent's wages, it is better for the principal to pay a constant wage in the absence of justification and to justify the evaluations with the worst likelihood ratio. Therefore, the agent's wage equals the principal's payments in the absence of justification. For $\kappa > \bar{\kappa}$, there is no justification and Lemma 8 in the appendix shows that this equality no longer holds.

As before, I concentrate on justification strategies, that optimally follow a threshold rule.

Corollary 1. *The principal optimally justifies evaluations μ up to a threshold $\delta \in [0, 1)$, while she does not justify evaluations above δ.*

The principal justifies only bad evaluations. In these cases, the agent suspects a distortion by the principal. For good evaluations, the agent does not suspect a deviation by the principal. Thus, the principal remains silent on good performance. I described this communication pattern in the introduction and summarized it in Figure 2. The comparative statics of the justification threshold δ in the justification costs κ and Proposition 6 remain valid in this model. Finally, return to leniency and centrality bias from the introduction. Begin with centrality.

Corollary 2. *For sufficiently large κ, wages set by subjective evaluations of a fully rational and unbiased principal exhibit centrality and compression at the top.*

Hence, optimal contracting results in centrality of wages. No biases or non-standard preferences are required to explain centrality in wages. To talk about leniency and centrality in evaluations in a meaningful way, it makes sense to think about indirect implementation. In addition, this indirect implementation simplifies optimal contracts. For this purpose, consider the following contract:

$$C(m_P, m_A) = \begin{cases} w^{**}\left(\lambda(m_P)\right) & \text{if } \lambda(m_P) \leq \delta \\ w^{**} & \text{else} \end{cases}$$

$$W(m_P, m_A) = \begin{cases} w^{**} & \text{if } m_P = [0, p_{max}] \\ w^{**}\left(\lambda(m_P)\right) & \text{if } \lambda(m_P) \leq \delta \text{ and } m_A = 1 \\ j(w^{**} + \kappa) & \text{else.} \end{cases}$$

It is easy to check that the contract implements the same incentives and the same utilities for principal and agent as optimal contracts in Proposition 8. Therefore, the agent chooses the same effort as before. The principal optimally reports $m_P = [0, p_{max}]$ if the evaluation p is above the threshold δ. Otherwise, she justifies the truth and sends a message $m_P = I$. Hence, she uses the messages $\mathcal{I}' = \{[0, p_{max}] \cup \{X \in \mathcal{I} | \lambda(X) \leq \delta\}$ in equilibrium. The agent optimally replies with the message $m_A = 1$ to all messages of the principal with $S \in m_P$. If $S \notin m_P$, the agent replies with $m_A = 0$. Such indirect mechanisms have a nice interpretation: The principal proposes a wage and, for $p \leq \delta$, provides justification. The agent is then asked whether he accepts or rejects the principal's offer. If the agent accepts the offered wage, the principal pays the wage to the agent. If the agent rejects the offered wage, there is a costly dispute resolution with the costs paid by the principal. Essentially, the agent can object to the principal's evaluation. This conflict resolution seems realistic, as Bretz et al. (1992, p. 332) state that "most organizations report having an informal dispute resolution system (e.g., open door policies) that employees may use to contest the appraisal outcome. About one-quarter report having formalized processes." These more realistic contracts imply leniency and centrality in evaluations.

Corollary 3. *For sufficiently large κ, evaluations by a fully rational and unbiased principal exhibit leniency, $\mathbb{E}(\lambda(m_P)) > \mathbb{E}(p)$, and centrality, $Var(p|p > \delta) > Var(\lambda(m_P)|p > \delta)$.*

This result confirms empirical observations of leniency and centrality: there is less distinction in subjective evaluations than in the underlying performance measure, in particular at the top. Taylor and Wherry (1951, p. 39) were the first to find lenient evaluations and "a marked distortion

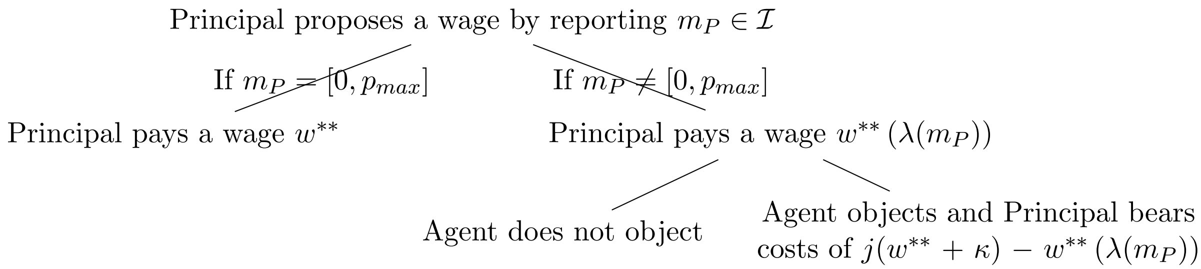

Figure 4: Indirect Implementation of Optimal Contracts

[of evaluations] ...with considerably poorer discrimination at the top."[9] Yet, this behavior is not the result of a bias, but implied by optimal contracting, which pools several evaluations and rewards them similarly. Thus, the contract eliminates wage differences that the principal would have to justify. For this reason, I refer to leniency and centrality instead of leniency bias and centrality bias.

Take this idea of an indirect implementation a step further and restrict the message spaces. Then the principal proposes a wage w by sending a message $m_P \in \mathcal{I}'$. If the principal pays a wage w^{**}, the game ends. If the principal pays a wage $w < w^{**}$, the agent has to opportunity to object to the proposed wage. If the agent objects, the principal has to pay additional costs of $j(w^{**}+\kappa) - w^{**}(\lambda(m_P))$. Figure 4 sketches this implementation. These costs can be interpreted as lawyers fees, mediation costs, or different kinds of conflicts. Notice, however, that it is even possible to implement optimal contracts without such costs and with ex-post balance balance, as shown by Lang (2017). Finally, consider the robustness of my results.

5 Robustness of the Results

In the model, the agent learns one successful category $S \in I$. Several extensions are possible. First, suppose the agent learns a finite subset $\mathcal{S} \subset I$ of successful categories. As long as the subset $\mathcal{S}$ is drawn randomly and is private information of the agent, the basic intuition remains valid. The agent cannot infer the realization of the evaluation p and the principal does not know the agent's subset. Extend the agent's message space to $m_A \in T^{|\mathcal{S}|}$ to let the agent report his information $\mathcal{S}$. Then messages agree if $m_A \subset m_P$. With this modification, all my previous results remain valid. Second, return to the initial setting, but suppose the agent learns one randomly drawn successful category $S \in I$ with probability $\rho < 1$ and nothing with probability $1 - \rho$. The agent could report zero if he did not learn any categories. Then message agree if $m_A \in m_P$ or $m_A = 0$. With this modification, all my previous results remain valid, as well. Third, return to the initial setting, but suppose the agent learns the result $I(S)$ in one randomly drawn category $S \in T$. Thus, his information is $(I(S), S)$. Extend the agent's message space to $\{0, 1\} \times T$. In this case, the agent has some, although very noisy information about the principal's evaluation p. Yet, the agent's wage cannot depend on the agent's message, because the agent would distort his message otherwise. Therefore, my results are still valid in this setting with the appropriate

[9] See also the references in Footnote 3.

definition of agreeing messages. Modify contracts so that messages agree if the agent reports a failed category or the agent reports a successful category that is included in the principal's message. Denote the agent's message as $m_A = (m_A^1, m_A^2)$. Hence, the principal pays

$$W^*(m_P, m_A) = \begin{cases} w^{**} & \text{if } \lambda(m_P) > \delta \\ w^{**}\left(\lambda(m_P)\right) & \text{if } \lambda(m_P) \leq \delta \text{ and } (m_A^1 = 0 \text{ or } (m_A^1 = 1 \text{ and } m_A^2 \in m_P)) \\ j(w^{**} + \kappa) & \text{else.} \end{cases}$$

It is easy to verify that Proposition 8 still characterizes an optimal contract and justification optimally follows a threshold rule. Therefore, my results are robust to alternative specifications of the agent's information.

Alternatively, assume that the agent is biased and systematically overestimates his performance. Hence, he understands some categories to report success, although they indeed report a failure. As long as the bias is systematical, however, it is possible to adapt the definition of agreeing messages. Then the results of this paper remain valid. Suppose, for example, that the agent learns two categories. One category is randomly drawn from the successful categories and the other category is randomly drawn from the unsuccessful categories. The agent is overoptimistic and believes both categories to be successful. Again extend the agent's message space to $m_A \in T^2$ and denote the agent's message as $m_A = (m_A^1, m_A^2)$. Then messages agree if $m_A^1 \in m_P$ or $m_A^2 \in m_P$. This extension does not change the analysis in this paper, so that all results remain valid.

6 Conclusion

This paper considers communicating a subjective performance measure in a principal-agent model. The principal can justify her evaluation of the agent's work. Providing justification is costly, does not convey additional information about the agent's effort, and does not serve a learning or instructing purpose. Nevertheless, the principal optimally justifies some evaluations if justification is not too costly. This justification allow the agent to detect distorted evaluations. Therefore, providing justification makes the incentives for the agent credible. The principal justifies only bad evaluations. Getting a good evaluation, the agent is happy to earn a high wage and does not suspect a distortion by the principal. Getting a bad evaluation, the agent wants to ensure that the principal evaluates him correctly and does not distort the evaluation downwards to save on wage costs. This communication pattern results in pooling and wage compression at the top, as illustrated in Figure 2 on page 10. These results fit well with empirical observations, often referred to as leniency bias and centrality bias, as discussed in Corollaries 2 and 3 as well as the introduction. The paper argues that this pattern of evaluations is a feature of optimal contracting with unbiased agents and no proof of biased behavior per se.

The principal's justification convinces the agent that the principal evaluates her appropriately. The expectation of an appropriate evaluation motivates the agent ex-ante to implement the specified work effort. Compare this to a naive contract that does not give the principal an incentive to provide justification. In this naive contract, the principal does not justify the evaluation and always reports the evaluation associated with the lowest wage. Anticipating this behavior the agent is unmotivated to implement any positive work effort. This partially explains the concern of the management literature to ensure credible feedback provision. In addition, the problem of credible evaluations provides a partial answer to Fuchs (2007, p. 1446), who emphasizes the importance of exploring "possible reasons for the existence of communication" between agent and principal. This paper shows that credibility problems cause communication.

The results of this paper are important for the design of incentives systems. First, the systems have to ensure the credible provision of appropriate feedback by institutionalizing the feedback process or using multi-source feedback. Second, pooling at the top could cause substantial costs for an incentives scheme if a high fraction of employees receives positive evaluations. Bernardin and Orban (1990, p. 199) provide the example of the Small Business Administration and NASA introducing a bonus scheme based on subjective evaluations. After more than 50% of eligible employees should receive a bonus, Congress responded with the requirement that no more than 25% of employees shall receive a bonus.

This paper assumes that the principal incurs costs for acquiring information. I would get similar results if the principal's costs instead concerned communicating the evaluation. In this case, the principal directly learns all the information I without incurring any costs. Then she decides whether to spend κ to communicate the entire information I by using the message space $\mathcal{I}$. If she does not spend the communications costs κ, her message space is restricted to $\mathcal{P}$. The costs of communication κ capture, for example, the opportunity costs of the principal having to spend time writing a report or talking to the agent instead of doing other tasks. Both settings have some merits; in reality, there could be a mixture of these two polar cases.

A Appendix

Lemma 6 shows that focusing on direct and truthful contracts is without loss of generality. The statement is similar to a revelation principle.

Lemma 6. *For every contract $\mathcal{W}$ there is a contract $\mathcal{W}'$, such that*

- *$\mathcal{W}'$ has the same costs for the principal as $\mathcal{W}$*
- *$\mathcal{W}'$ incentivizes the agent to accept the contract and to exert the same effort e as in $\mathcal{W}$*
- *$\mathcal{W}'$ incentivizes agent and principal to send truthful messages.*

Proof: Suppose there is a feasible contract $\mathcal{W}$. Denote the principal's expected payment conditional on evaluation p in contract $\mathcal{W}$ given equilibrium strategies by $\hat{w}(p)$ and the corresponding certainty equivalent of the agent's wages by $\hat{c}(p)$. Let $\hat{\beta}(p)$ denote the principal's equilibrium justification in contract $\mathcal{W}$. Determine contract $\mathcal{W}'$ by $C'(m_P, m_A) = \hat{c}(m_P)$ and

$$W'(m_P, m_A) = \begin{cases} \hat{w}(m_P) & \text{if } \hat{\beta}(m_P) = m_A \\ \max_{i \in \mathcal{P}, a \in \{0,1\}} W(i,a) & \text{else} \end{cases}$$

for all $m_P \in \mathcal{P}$ and all $m_A \in \{0, 1\}$. Denote by $\beta'(p)$ the principal's equilibrium justification in contract $\mathcal{W}'$. Given truth-telling in contract $\mathcal{W}'$ and $\beta'(p) = \hat{\beta}(p)$ for all $p \in \mathcal{P}$, equilibrium utilities in contract $\mathcal{W}$ equal the utilities in contract $\mathcal{W}'$ for the principal and for the agent. For the principal, we have:

$$W'(p, \beta'(p)) + \kappa\beta'(p) = W'(p, \hat{\beta}(p)) + \kappa\hat{\beta}(p) = \hat{w}(p) + \kappa\hat{\beta}(p)$$

for all $p \in \mathcal{P}$. For the agent, utilities are analogously the same in contract $\mathcal{W}$ and $\mathcal{W}'$. Next, feasibility in contract $\mathcal{W}$ and the agent's risk aversion together imply $\hat{w}(p) \geq \hat{c}(p)$. Therefore, payments are also feasible as $W'(p, \beta) \geq C'(p, \beta)$ for all $p \in \mathcal{P}$ and $\beta \in \{0, 1\}$. As the agent's equilibrium utilities do not change, contract $\mathcal{W}'$ incentivizes the agent to accept the contract and exert effort e as in contract $\mathcal{W}$. In addition, the agent's wage in contract $\mathcal{W}'$ does not depend on the agent's message. Hence, truth-telling is optimal for the agent in contract $\mathcal{W}'$. What about the principal?

By optimality of behavior in contract $\mathcal{W}$, equilibrium costs have to be optimal and

$$\kappa\hat{\beta}(p) + \hat{w}(p) \leq \kappa\hat{\beta}(m_P) + \hat{w}(m_P) = \kappa\hat{\beta}(m_P) + W'(m_P, \hat{\beta}(m_P))$$

for all $p \in \mathcal{P}$ and all $m_P \in \{p' \in \mathcal{P} | \hat{\beta}(p') = 0 \text{ or } p' \geq p\}$. For the same reasons,

$$\kappa\hat{\beta}(p) + \hat{w}(p) \leq \kappa\tilde{\beta} + \max_{i \in \mathcal{P}, a \in \{0,1\}} W(i, a) = \kappa\tilde{\beta} + W'(m_P, \tilde{\beta})$$

for all $p \in \mathcal{P}$, all $\tilde{\beta} \in \{0, 1\}$ and all $m_P \in \{p' \in \mathcal{P} | \hat{\beta}(p') \neq \tilde{\beta}\}$. Therefore,

$$\kappa\hat{\beta}(p) + W'(p, \hat{\beta}(p)) = \kappa\hat{\beta}(p) + \hat{w}(p) \leq \kappa\tilde{\beta} + W'(m_P, \tilde{\beta})$$

for all $p \in \mathcal{P}$ and all $(\tilde{\beta}, m_P) \in \big\{\{0\} \times \mathcal{P}, \{1\} \times ([p, 1] \cap \mathcal{P})\big\}$. Hence, it is optimal for the principal to follow $\beta'(p) = \hat{\beta}(p)$ for all $p \in \mathcal{P}$ and report $m_P = p$ in contract $\mathcal{W}'$. □

Denote Program A without the truth-telling constraints TT_P and TT_A by Program A*. Lemma 1 solves Program A* and characterizes the optimal contract if the principal's information is contractible. This contract is unique up to out-of-equilibrium payments.

Proof of Lemma 1: Contractible information means $m_P = p$. To fit contractible information into my setting, define contract $\mathcal{W}$ by $W(m_P, m_A) = C(m_P, m_A) = w_e^*(m_P)$ for all $m_P \in \mathcal{P}$ and all $m_A \in \{0, 1\}$ with $w_e^*(\cdot)$ defined below. Contract $\mathcal{W}$ implies $\beta(p) = 0$ for all $p \in \mathcal{P}$.

In order to implement no effort, $e = 0$, set $w_0^*(p) = u^{-1}(\bar{u} + d(0))$ for all p. If the principal wants positive effort, $e > 0$, the agent's incentive compatibility matters. The first-order approach is valid here, because $F(p|e)$ is a linear combination of distribution functions. This implies that the convex distribution function condition is satisfied. According to Grossman and Hart (1983) and Rogerson (1985), the convex distribution function condition in combination with the convexity of $d(\cdot)$ and the monotone likelihood ratio property guarantees that the first-order approach is valid. Therefore, the agent's incentive compatibility is equivalent to

$$\sum\nolimits_{p\in\mathcal{P}} u(w(p))(f^H(p) - f^L(p)) \geq d'(e). \tag{IC}$$

Notice that the constraint set is nonempty. Take for example any $\bar{w} > 0$ and

$$w(p) = \begin{cases} \bar{w} & \text{if } f^H(p) - f^L(p) \geq 0 \\ h(\bar{w}) & \text{else} \end{cases}$$

with $h(\bar{w})$ positive, but small enough, such that the agent's incentive compatibility IC is satisfied. This implicitly defines an increasing function $h(\cdot)$. Consequently, there is a $\bar{w}$ fulfilling the participation constraint PC with equality. Therefore, the constraint set of Program A* is nonempty. Moreover, the costs of the contract are lower than $\bar{w} < \infty$.

Optimization with the Lagrange multipliers of the participation constraint ν_1 and of the incentive compatibility ν_2 determines the optimal contract as

$$f(p|e) - \nu_1 u'(w(p))f(p|e) - \nu_2 u'(w(p))(f^H(p) - f^L(p)) = 0,$$

$$\frac{1}{u'(w(p))} = \nu_1 + \nu_2 \frac{f^H(p) - f^L(p)}{f(p|e)} = \nu_1 + \nu_2 \frac{\frac{f^H(p)}{f^L(p)} - 1}{e\frac{f^H(p)}{f^L(p)} + 1 - e} \tag{2}$$

The Lagrange multiplier ν_2 is positive for the following reason: The solution to problem Program A* without the incentive compatibility IC is a constant payment $u^{-1}(\bar{u}+d(e))$ that violates the incentive compatibility IC. Hence, $\nu_2 > 0$. Since the fraction $\frac{l-1}{el+1-e}$ increases in l, the right-hand side of above equation (2) increases in $p \in \mathcal{P}$ due to the monotone likelihood ratio property. Therefore, the concavity of $u(\cdot)$ implies that $w_e^*(p)$ increases in $p \in \mathcal{P}$. □

Proof of Lemma 2: Consider the following contract: $C(m_P, m_A) = w_e^*(m_P)$ and

$$W(m_P, m_A) = \begin{cases} w_e^*(m_P) & \text{if } m_A = 1 \\ w_e^*(p_{max}) & \text{else} \end{cases}$$

for all $m_P \in \mathcal{P}$ and all $m_A \in \{0,1\}$. The contract trivially satisfies TT$_A$ as the agent's wage is independent of her message m_A. Whenever the principal does not provide justification, the agent sends the message $m_A = 0$ and the principal's payments increase to $w_e^*(p_{max})$. If the principal deviates to a message $m_P > p$ with justification, her payments increase to $w_e^*(m_P)$. Lemma 1 ensures that these payments are bigger than $w_e^*(p)$, as $w_e^*(\cdot)$ is increasing. Hence, the truth-telling constraints TT$_P$ and TT$_A$ are satisfied. Constraint (1) is obviously satisfied, too. The contract proposed above is optimal, because it satisfies the additional constraints and implements the equilibrium wage $w_e^*(p)$. As $w_e^*(p)$ is a solution to the relaxed Problem A* according to Lemma 1, this contract is a solution to Problem A for $\kappa = 0$. □

Proof of Proposition 1: Suppose to the contrary that the principal justifies all evaluations: $\beta(p) = 1$ for all $p \in \mathcal{P}$. Then the principal's expected costs are at least $\kappa + \mathbb{E}(w_e^*(p))$ with $w_e^*(\cdot)$ defined in Lemma 1. Yet, partial justification implements effort e even cheaper. For this purpose, consider contract $\mathcal{W}$ with $C(m_P, m_A) = w_e^*(m_P)$ and

$$W(m_P, m_A) = \begin{cases} w_e^*(m_P) & \text{if } m_P \neq p_{max} \text{ and } m_A = 1 \\ \max\{w_e^*(p_{2ndmax}) + \kappa, w_e^*(p_{max})\} & \text{if } m_P = p_{max} \\ w_e^*(p_{max}) + \kappa & \text{else} \end{cases}$$

with $p_{2ndmax} = \max(\mathcal{P} \setminus \{p_{max}\})$ the second highest evaluation. The principal justifies all evaluations except the highest ones: $\beta(p_{max}) = 0$ and $\beta(p) = 1$ for all $p < p_{max}$. If the principal's information indicates a very good performance, $p = p_{max}$, justification would increase her costs by κ as the wage costs remain unchanged. In addition, constraints PC and IC are still satisfied, because the agent's wage remains unchanged compared to Lemma 2. As in Lemma 2, the truth-telling constraints TT_P and TT_A are satisfied. Again, it is easy to check that constraint (1) is also satisfied. Therefore, this contract implements effort e and is cheaper than $\kappa + \mathbb{E}(w_e^*(p))$. In particular, the principal's expected costs decrease by

$$\begin{aligned} &f(p_{max}|e)(w_e^*(p_{max}) + \kappa - \max\{w_e^*(p_{2ndmax}) + \kappa, w_e^*(p_{max})\}) = \\ =&f(p_{max}|e) \min\{w_e^*(p_{max}) - w_e^*(p_{2ndmax}), \kappa\}) > 0. \end{aligned}$$

This shows that the principal should not justify all evaluations for positive costs κ. □

Proof of Proposition 2: Suppose there is a contract $W(m_P, m_A)$ and $C(m_P, m_A)$ with equilibrium justification $\beta(p)$ that satisfies the constraints of Program A. If some evaluations are not justified in the contract and $\beta(p) = 0$ for a $p \in \mathcal{P}$, set $\bar{w} = W(p, 0)$ for a p with $\beta(p) = 0$. By Proposition 1, there is such a $p \in \mathcal{P}$ with $\beta(p) = 0$ in an optimal contract for $\kappa > 0$. If all evaluations are justified in the initial contract and $\beta(p) = 1$ for all p, set $\bar{w} = \kappa + \max_{p \in \mathcal{P}} W(p, 1)$. To simplify the exposition, define $w(p) = W(p, \beta(p))$ for all $p \in \mathcal{P}$.

The principal's truth-telling constraint TT_P in Program A deals with the following deviations: misreporting p with $m_p \neq p$ and/or justifying p for $\beta(p) = 0$ or not justifying p for $\beta(p) = 1$. There are three implications. First, constraint TT_P implies that $w(p)$ has to be constant in $\{p|\beta(p) = 0\}$, because the agent cannot detect a deviation by the principal in the absence of justification. Therefore, $W(p, 0) = w(p) = \bar{w}$ for all p with $\beta(p) = 0$. Second, constraint TT_P implies $w(p) + \kappa \leq W(p', 0)$ for all p' and all p with $\beta(p) = 1$. Otherwise, it would be profitable for the principal not to justify the evaluation p, but deviate to $m_P = p'$ and $\beta = 0$. Hence, constraint TT_P implies

$$\bar{w} \geq w(p) + \kappa\beta(p) \geq (1 - \beta(p))\bar{w} \qquad \forall p \in \mathcal{P}.$$

Third, constraint TT_P implies that $w(p) + \kappa\beta(p)$ is non-decreasing in $p \in \mathcal{P}$. Otherwise, there is a $p' > p$ with $w(p') + \kappa\beta(p') < w(p) + \kappa\beta(p)$ and the principal could reduce her costs by deviating to $m_P = p'$ and $\beta = \beta(p')$.

The agent's truth-telling constraint TT_A in Program A implies that $C(m_P, 0) = C(m_P, 1)$ for all $m_P \in \mathcal{P}$. Therefore, define $c(m_P) = C(m_P, 0)$ for all $m_P \in \mathcal{P}$. Lemma 1 proves that the first-order approach to the agent's incentive compatibility is valid here. Hence, IC in Program A

is equivalent to
$$\sum\nolimits_{p\in\mathcal{P}} u(c(p))(f^H(p) - f^L(p)) \geq d'(e). \tag{IC}$$

The preceding steps show that the following program is a relaxed version of Program A:

$$\min \sum\nolimits_{p\in\mathcal{P}} (w(p) + \kappa\beta(p))f(p|e), \tag{B}$$
$$\text{subject to } \sum\nolimits_{p\in\mathcal{P}} u\big(c(p)\big) f(p|e) - d(e) \geq \bar{u}, \tag{PC}$$
$$\sum\nolimits_{p\in\mathcal{P}} u\big(c(p)\big)(f^H(p) - f^L(p)) \geq d'(e), \tag{IC}$$
$$\bar{w} \geq w(p) + \kappa\beta(p) \geq (1-\beta(p))\bar{w} \qquad \forall p \in \mathcal{P} \tag{3}$$
$$w(p) + \kappa\beta(p) \text{ non-decreasing in } p \tag{4}$$
$$w(p) \geq c(p) \qquad \forall p \in \mathcal{P} \tag{5}$$

As before, the objective is to minimize expected costs, here justification costs and the principal's payments. The agent's participation constraint PC and incentive compatibility IC remain unchanged. The principal's truth-telling constraint TT_P implies constraints (3) and (4). Finally, constraint (1) implies constraint (5). Any contract in the constraint set of Program A is also in the constraint set of Program B. The remainder of this proof shows that the reverse is also true.

Suppose $w(p), \bar{w}$ and $c(p)$ with equilibrium justification $\beta(p)$ satisfy the constraints of Program B. Then construct a contract $\mathcal{W}^{**}$ by defining $\mathcal{C} = \{p \in \mathcal{P} | \beta(p) = 1\}$ and

$$C^{**}(m_P, m_A) = c\,(m_P)$$
$$W^{**}(m_P, m_A) = \begin{cases} w\,(m_P) & \text{if } m_P \in \mathcal{C} \text{ and } m_A = 1 \\ \bar{w} & \text{else} \end{cases}$$

and $\beta^{**}(p) = \beta(p)$ for the principal's justification choice. Contract $\mathcal{W}^{**}$ yields the same value for the objective of Program A as Program B assigns to $w(p), \bar{w}$ and $c(p)$. The participation constraint and the incentive compatibility in Program B guarantee that contract $\mathcal{W}^{**}$ satisfies the participation constraint PC and the incentive compatibility IC in Program A. In addition, the agent's wage is independent of his message m_A. Therefore, the agent's truth-telling constraint TT_A in Program A is trivially satisfied. Constraint (5) ensures constraint (1). It remains to verify the principal's truth-telling constraint TT_P.

Suppose $p \notin \mathcal{C}$ is realized. If the principal reports truthfully, i.e., $\beta = 0$ and $m_P = p$, her costs are $\bar{w}$ in contract $\mathcal{W}^{**}$. Deviating to $m_P \neq p$ and $\beta = 0$ causes the principal costs of $\bar{w}$. Deviating to $m_P \geq p$ and $\beta = 1$ causes the principal costs of

$$\begin{cases} w(m_P) + \kappa \overset{(3),(4)}{=} \bar{w} - \kappa + \kappa = \bar{w} & \text{if } m_P \in \mathcal{C} \\ \bar{w} + \kappa & \text{else} \end{cases}$$

due to constraints (3) and (4). Therefore, there are no profitable deviations for $p \notin \mathcal{C}$.

Suppose $p \in \mathcal{C}$ is realized. If the principal reports truthfully, i.e., $\beta = 1$ and $m_P = p$, her costs are $w(p) + \kappa$ in contract $\mathcal{W}^{**}$. Deviating to $\beta = 0$ causes costs of $\bar{w}$ for the principal. Constraint (3) implies $\bar{w} \geq w(p) + \kappa$. Deviating to $m_P > p$ and $\beta = 1$ causes costs of $w(m_P) + \kappa$

or $\bar{w}+\kappa$ for the principal. Constraints (3) and (4) imply that $\min\{w(m_P), \bar{w}\} \geq w(p)$. Therefore, there are no profitable deviations for $p \in \mathcal{C}$. Consequently, truthful reporting is optimal for the principal and contract $\mathcal{W}^{**}$ satisfies the principal's truth-telling constraint TT_P. Consequently, a solution to Program B also determines a solution to Program A. □

Proof of Proposition 3: In order to show that only bad evaluations are justified, assume to the contrary that there is an optimal contract $\mathcal{W}$ with $p_L, p_H \in \mathcal{P}$, such that $p_L < p_H$, $\beta(p_L) = 0$ and $\beta(p_H) = 1$. Construct a contract $\mathcal{W}^{**}$ based on contract $\mathcal{W}$ in the same way as in Proposition 2. The proof of Proposition 2 shows that contract $\mathcal{W}^{**}$ is a solution to Program B, yields the same value for the objective as contract $\mathcal{W}$ and satisfies all the constraints of Program A. Lemma 7 below shows that there is a contract $\mathcal{W}'$ that implements effort e cheaper than $\mathcal{W}$. The contract $\mathcal{W}'$ satisfies all the constraints of Program B and, hence, all the constraints of Program A according to Proposition 2. Therefore, the modified contract contradicts optimality of the initial contract $\mathcal{W}$. Consequently, optimal contracts do not require the principal to justify good evaluations, as the agent does not suspect a distortion of these evaluations. Nevertheless, the principal justifies bad evaluations, if any evaluations are justified at all. Hence, optimal justification follows a threshold rule. □

Lemma 7. *In any solution to Program B, there is a $\delta \in [0, 1)$ with $\beta(p) = 1$ if and only if $p \leq \delta$.*

Proof: In order to show that only bad evaluations are justified, assume to the contrary that there is an optimal contract $\mathcal{W}^{**}$ with $p_L, p_H \in \mathcal{P}$, such that $p_L < p_H$, $\beta(p_L) = 0$ and $\beta(p_H) = 1$. Program B and, in particular, constraints (3) and (4) imply that $w(p_H) + \kappa = w(p_L)$. Define $\mathcal{C}^{**} = \{p \in \mathcal{P} | \beta(p) = 1\}$.

The next step constructs a contract $\mathcal{W}'$ that implements effort e cheaper than $\mathcal{W}^{**}$. Modify contract $\mathcal{W}^{**}$ in the following way to get $\mathcal{W}'$: Define $\mathcal{C}' = \mathcal{C}^{**} \backslash \{p_H\}$ and $w'(p_H) = c'(p_H) = w(p_L)$ and $c'(p_L) = \tilde{c}$ with $\tilde{c}$ determined by

$$u(\tilde{c}) = u(c(p_L)) - (u(w(p_L)) - u(c(p_H))) \frac{f(p_H|e)}{f(p_L|e)}.$$

Otherwise contract $\mathcal{W}'$ equals contract $\mathcal{W}^{**}$. The agent's wage for an evaluation p_L decreases from $c(p_L)$ to $\tilde{c}$ in contract $\mathcal{W}'$, while the agent's wage for an evaluation p_H increases from $c(p_H)$ to $w(p_L)$, as $c(p_H) \leq w(p_H) < w(p_L)$. The principal's expected costs remain unchanged. In contract $\mathcal{W}'$, the principal does not justify p_L and p_H. Hence, $\beta'(p_L) = \beta'(p_H) = 0$. The definition of $\tilde{c}$ implies that $\tilde{c} < c(p_L) \leq w(p_L)$ and that the change in the agent's expected utilities ($\mathbb{E}[u(\mathcal{W}') - u(\mathcal{W}^{**})]$ in sloppy notation) equals

$$\begin{aligned} &\sum\nolimits_{i \in \{L,H\}} (u(c'(p_i)) - u(c(p_i))) f(p_i|e) = \\ =&(u(\tilde{c}) - u(c(p_L))) f(p_L|e) + (u(w(p_L)) - u(c(p_H))) f(p_H|e) = 0. \end{aligned} \tag{6}$$

Hence, contract $\mathcal{W}'$ satisfies the participation constraint PC in Program B, as the agent's expected utilities remain unchanged. Additionally, the left-hand side of the incentive compatibility IC is now bigger than the marginal cost of effort, $d'(e)$, because the left-hand side of the incentive compatibility IC changes by

$$\sum_{p\in\mathcal{P}}(f^H(p)-f^L(p))(u(c'(p))-u(c(p)))=\sum_{i\in\{L,H\}}(f^H(p_iL)-f^L(p_i))(u(c'(p_i))-u(c(p_i)))=$$
$$=(f^H(p_L)-f^L(p_L))(u(\tilde{c})-u(c(p_L)))+(f^H(p_H)-f^L(p_H))(u(w(p_L))-u(c(p_H)))=$$
$$=\frac{f^H(p_L)-f^L(p_L)}{f(p_L|e)}f(p_L|e)(u(\tilde{c})-u(c(p_L)))+\frac{f^H(p_H)-f^L(p_H)}{f(p_H|e)}f(p_H|e)(u(w(p_L))-u(c(p_H)))>$$
$$>\frac{f^H(p_L)-f^L(p_L)}{f(p_L|e)}\left(f(p_L|e)(u(\tilde{c})-u(c(p_L)))+f(p_H|e)(u(w(p_L))-u(c(p_H)))\right)=0.$$

The last equality follows from Eq. (6). The main inequality follows from the monotone likelihood ratio property, which ensures that

$$\frac{f^H(p)-f^L(p)}{f(p|e)}=\frac{\frac{f^H(p)}{f^L(p)}-1}{e\frac{f^H(p)}{f^L(p)}+1-e}$$

increases in $p\in\mathcal{P}$. As the principal's costs remain unchanged: $w'(p)+\kappa\beta'(p)=w(p)+\kappa\beta(p)$ for all $p\in\mathcal{P}$, contract $\mathcal{W}'$ satisfies constraint (4). Contract $\mathcal{W}'$ also satisfies constraints (3) and (5).

Apply this procedure repeatedly, until only low evaluations are justified and there are no more $p_L,p_H\in\mathcal{P}$ with the required properties. Finally, I show that the new contract is cheaper for the principal. There are two cases to consider. If $p_{min}\in\mathcal{C}'$, the first case applies. Otherwise, $p_{min}\notin\mathcal{C}'$ and the second case applies. Begin with the first case and $p_{min}\in\mathcal{C}'$. Above steps ensure that there is a $p_3\in\mathcal{P}$ with $c'(p_3)<w'(p_3)=w(p_L)$ and $p_3\notin\mathcal{C}'$. Hence, $p_3\neq p_{min}$. Increase $c'(p_3)$ by a small $\epsilon>0$ to $c'(p_3)+\epsilon$ and reduce $w'(p_{min})$ and $c'(p_{min})$ to $\tilde{c}^\epsilon$ with $\tilde{c}^\epsilon$ determined by

$$u(\tilde{c}^\epsilon)=u(c'(p_{min}))-(u(c'(p_3)+\epsilon)-u(c'(p_3)))\frac{f(p_3|e)}{f(p_{min}|e)}.$$

This modification does not affect the agent's participation constraint. I show above that constraint IC was slack. Hence, choosing ϵ sufficiently small ensures that the modified contract satisfies constraints IC and (5) due to $c'(p_3)+\epsilon\leq w'(p_3)$. It is easy to see that the modified contract satisfies constraints (3) and (4), because reducing $w'(p_{min})$ never violates constraint (4). Moreover, $\beta'(p_{min})=1$ and reducing $w'(p_{min})$ does not violate constraint (3). The modified contract implies lower expected costs for the principal compared to contract $\mathcal{W}^{**}$. The modified contract satisfies all the constraints of Program B. Therefore, the modified contract contradicts optimality of the initial contract $\mathcal{W}^{**}$.

Now turn to the second case with $p_{min}\notin\mathcal{C}'$. Notice that $p_{min}\notin\mathcal{C}'$ implies $\mathcal{C}'=\emptyset$. Otherwise, there were $p_L,p_H\in\mathcal{P}$ with the required properties. Again, above steps ensure that there is a $p_3\in\mathcal{P}$ with $c'(p_3)<w'(p_3)=w(p_L)$. Increase $c'(p_3)$ by a small $\epsilon>0$ to $c'(p_3)+\epsilon$ and reduce the remaining $c'(p)$ and all $w'(p)$ by $\hat{\epsilon}^\epsilon$ with $\hat{\epsilon}^\epsilon$ determined by

$$f(p_3|e)(u(c'(p_3)+\epsilon)-u(c'(p_3)))+\sum_{p\in\mathcal{P}\setminus\{p_3\}}f(p|e)(u(c'(p)-\hat{\epsilon}^\epsilon)-u(c'(p)))=0.$$

This modification does not affect the agent's participation constraint. I show above that constraint IC was slack. Hence, choosing ϵ sufficiently small ensures that the modified contract satisfies constraints IC and (5) due to $c'(p_3) + \epsilon \leq w'(p_3) - \hat{\epsilon}^{\epsilon}$. It is easy to see that the modified contract satisfies constraints (3) and (4), because reducing all $w'(p)$ equally never violates constraints (3) and (4). The modified contract implies lower expected costs for the principal compared to contract $\mathcal{W}^{**}$. The modified contract satisfies all the constraints of Program B. Therefore, the modified contract contradicts optimality of the initial contract $\mathcal{W}^{**}$. Consequently, justification optimally follows a threshold rule and there is a $\delta \in [0, 1)$ with $\beta(p) = 1$ if and only if $p \leq \delta$. $\square$

Lemma 8. *If $e > 0$ and the principal does not justify any evaluations, the optimal contract is*

$$C(m_P, m_A) = \begin{cases} u^{-1}\left(\bar{u} + d(e) - \frac{(1-f(p_{min}|e))d'(e)}{f^L(p_{min}) - f^H(p_{min})}\right) & \text{if } m_P = p_{min} \\ u^{-1}\left(\bar{u} + d(e) + \frac{f(p_{min}|e)d'(e)}{f^L(p_{min}) - f^H(p_{min})}\right) & \text{else} \end{cases}$$

and $W(m_P, m_A) = C(p_{max}, 0)$ for all $m_P \in \mathcal{P}$ and all $m_A \in \{0, 1\}$. This contract is unique up to out-of-equilibrium payments.

Proof: Assume $\beta(p) = 0$ for all $p \in \mathcal{P}$ is optimal. Proposition 2 shows that the principal's truth-telling constraint TT_P requires $W(p, 0)$ to be constant in $p \in \mathcal{P}$. In addition, the agent's truth-telling constraint IC_{m_A} requires $C(p, 0) = C(p, 1)$ for all $p \in \mathcal{P}$. Define $\bar{w} = W(p, 0)$ for a $p \in \mathcal{P}$ and $c(p) = C(p, 0)$ for all $p \in \mathcal{P}$. Lemma 1 shows that the first-order approach to the agent's incentive compatibility is valid here. Hence, Program A simplifies to

$$\min_{c(p), \bar{w}} \bar{w}, \tag{C}$$

$$\text{subject to} \sum_{p \in \mathcal{P}} u(c(p)) f(p|e) - d(e) \geq \bar{u}, \tag{PC}$$

$$\sum_{p \in \mathcal{P}} u(c(p))(f^H(p) - f^L(p)) \geq d'(e), \tag{IC}$$

$$f(p|e)(\bar{w} - c(p)) \geq 0 \qquad \forall p \in \mathcal{P}. \tag{7}$$

Define ν_1, ν_2 and $\chi(p)$ as the Lagrange multipliers of the conditions PC, IC and (7) respectively. If $c(p) = \bar{w}$ for all $p \in \mathcal{P}$, the contract violates incentive compatibility IC. Therefore, there is an evaluation $p^* \in \mathcal{P}$ with third-party payments, i.e., $c(p^*) < \bar{w}$. Then the complementary slackness condition yields $\chi(p^*) = 0$. Optimization of the Lagrangian with respect to $c(p^*)$ results in

$$-\nu_1 u'(c(p^*)) f(p^*|e) - \nu_2 u'(c(p^*))(f^H(p^*) - f^L(p^*)) = 0$$

$$\nu_1 + \nu_2 \frac{f^H(p^*) - f^L(p^*)}{f(p^*|e)} = 0. \tag{8}$$

The monotone likelihood ratio property ensures that $\frac{f^H(p) - f^L(p)}{f(p|e)}$ strictly increases in $p \in \mathcal{P}$. In addition, ν_2 is positive in an optimum, because the solution to Program C without the incentive compatibility IC is $\bar{w} = c(p) = u^{-1}(\bar{u} + d(e))$ for all $p \in \mathcal{P}$ and this solution violates constraint IC. Therefore, equation (8) can hold for at most one $p^* \in \mathcal{P}$. Hence, $c(p) = \bar{w}$ and $\chi(p) \geq 0$ for all $p \in \mathcal{P} \setminus \{p^*\}$. Assume to the contrary $p^* \neq p_{min}$. Optimization of the Lagrangian with respect to $c(p_{min})$ results in

$$-\nu_1 u'(c(p_{min}))f(p_{min}|e) - \nu_2 u'(c(p_{min}))(f^H(p_{min}) - f^L(p_{min})) + \chi(p_{min})f(p_{min}|e) = 0$$
$$\nu_1 + \nu_2 \frac{f^H(p_{min}) - f^L(p_{min})}{f(p_{min}|e)} = \frac{1}{u'(c(p_{min}))}\chi(p_{min}).$$

Equation (8), $\nu_2 > 0$, $p_{min} < p^*$ and the monotone likelihood ratio property imply that the left-hand side of the last equation is negative. The right-hand side is non-negative, because constraint (7) is binding for p_{min} and $\chi(p_{min}) \geq 0$ as $p^* \neq p_{min}$. This contradiction proves that $p^* = p_{min}$. Therefore, $c(p) = \bar{w}$ and $\chi(p) \geq 0$ for all $p \in \mathcal{P} \setminus \{p_{min}\}$.

Plugging these results into constraints PC and IC yields:

$$u(w)(1 - f(p_{min}|e)) + u(c(p_{min}))f(p_{min}|e) - d(e) = \bar{u},$$
$$(u(c(p_{min})) - u(w))(f^H(p_{min}) - f^L(p_{min})) \geq d'(e),$$

because $0 = \sum_{p\in\mathcal{P}} f^H(p) - f^L(p) = f^H(p_{min}) - f^L(p_{min}) + \sum_{p\in\mathcal{P}\setminus\{p_{min}\}} f^H(p) - f^L(p)$. Solving the first equation for $u(c(p_{min}))$ gives

$$u(c(p_{min})) = \frac{\bar{u} + d(e) - u(w)(1 - f(p_{min}|e))}{f(p_{min}|e)}.$$

Inserting this value for $u(c(p_{min}))$ into the second inequality leads to

$$(\bar{u} + d(e) - u(w))\left(f^H(p_{min}) - f^L(p_{min})\right) \geq f(p_{min}|e)d'(e)$$

and finally results in

$$u(w) = \bar{u} + d(e) + \frac{f(p_{min}|e)}{f^L(p_{min}) - f^H(p_{min})}d'(e) \qquad \text{and}$$
$$u(c(p_{min})) = \bar{u} + d(e) - \frac{1 - f(p_{min}|e)}{f^L(p_{min}) - f^H(p_{min})}d'(e).$$

Therefore, if $e > 0$ and $\beta(p) = 0$ for all $p \in \mathcal{P}$, the optimal contract is $W(m_P, m_A) = w$ and

$$c(m_P, m_A) = \begin{cases} c(p_{min}) & \text{if } m_P = p_{min} \\ w & \text{else} \end{cases}$$

for all $m_P \in \mathcal{P}$ and all $m_A \in \{0, 1\}$. Finally, it remains to verify that this contract is in the constraint set of Program A. The contract satisfies the agent's participation constraint PC and incentive compatibility IC by definition of w and $c(p_{min})$. As the agent's wage is independent of his message, his truth-telling constraint TT_A is satisfied. As the principal's payment is independent of her message, her truth-telling constraint TT_P is satisfied, too. The contract also satisfies constraint (1). □

Proof of Proposition 4: The proof proceeds in two steps. First, I show that using justification is not optimal if the costs κ are above the threshold in the proposition. Second, I show that for lower costs κ and $e > 0$ there is a contract in the constraint set of Program A with $\beta(p) = 1$ for a $p \in \mathcal{P}$ and lower costs for the principal than any contract without justification.

Begin with the first step. If $e = 0$, the optimal contract is $W(m_P, m_A) = C(m_P, c_A) =$

$u^{-1}(\bar{u}+d(0))$ for all m_P and m_A. Hence, it is not optimal to use justification for $\kappa > 0$. Therefore, restrict attention to $e > 0$. If the principal does not provide justification, her costs are

$$w = u^{-1}\left(\bar{u} + d(e) + \frac{f(p_{min}|e)}{f^L(p_{min}) - f^H(p_{min})} d'(e)\right)$$

according to contract $\mathcal{W}$ in Lemma 8. Suppose $\kappa > w - c(p_{min})$ with w and $c(p_{min})$ determined in Lemma 8. Assume to the contrary that there is an optimal contract $\mathcal{W}'$ with some justification, i.e., there is a $\tilde{p} \in \mathcal{P}$ with $\beta'(\tilde{p}) = 1$. Proposition 3 shows that $\beta'(\tilde{p}) = 1$ implies $\beta'(p_{min}) = 1$. If $p = p_{min}$, the principal's costs are $\kappa + C'(p_{min}, 1)$. Proposition 2 proves that constraint TT_P requires $W'(p, \beta'(p)) + \kappa\beta'(p)$ to be non-decreasing in p. Therefore, the principal's costs are at least $\kappa + C'(p_{min}, 1)$ for any $p \in \mathcal{P}$. Hence, the costs of contract $\mathcal{W}'$ are at least $\kappa + C'(p_{min}, 1) > w - c(p_{min}) + C'(p_{min}, 1)$. If $c(p_{min}) \leq C'(p_{min}, 1)$, these costs are above w. Hence contract $\mathcal{W}'$ is more expensive than contract $\mathcal{W}$. Consequently, restrict attention to $c(p_{min}) > C'(p_{min}, 1)$. In contract $\mathcal{W}$, the agent's participation constraint PC was binding. Therefore, the agent's participation constraint PC requires for contract $\mathcal{W}'$:

$$\begin{aligned}(1 - f(p_{min}|e))u(\mathbb{E}(C'(p, \beta'(p))|p \neq p_{min})) + f(p_{min}|e)u(C'(p_{min}, 1)) \geq \\ \geq (1 - f(p_{min}|e))u(w) + f(p_{min}|e)u(c(p_{min})) = \bar{u} + d(e).\end{aligned}$$

Rearranging yields

$$u(\mathbb{E}(C'(p, \beta'(p))|p \neq p_{min})) \geq u(w) + \frac{f(p_{min}|e)}{1 - f(p_{min}|e)}(u(c(p_{min})) - u(C'(p_{min}, 1))) > u(w). \quad (9)$$

Hence, $C'(p_{min}, 1) < c(p_{min}) < w < \mathbb{E}(C'(p, \beta'(p))|p \neq p_{min})$. Assume to the contrary that

$$\mathbb{E}(C'(p, \beta'(p))|p \neq p_{min}) - w \leq \frac{f(p_{min}|e)}{1 - f(p_{min}|e)}(c(p_{min}) - C'(p_{min}, 1)).$$

The ordering above together with the concavity of u imply

$$u(\mathbb{E}(C'(p, \beta'(p))|p \neq p_{min})) - u(w) < \frac{f(p_{min}|e)}{1 - f(p_{min}|e)}(u(c(p_{min})) - u(C'(p_{min}, 1))).$$

This inequality contradicts the last implication (9) of the agent's participation constraint. Hence

$$\begin{aligned}&\mathbb{E}(C'(p, \beta'(p))|p \neq p_{min}) - w > \frac{f(p_{min}|e)}{1 - f(p_{min}|e)}(c(p_{min}) - C'(p_{min}, 1)). \\ &\Rightarrow (1 - f(p_{min}|e))\mathbb{E}(C'(p, \beta'(p))|p \neq p_{min}) + f(p_{min}|e)C'(p_{min}, 1) > \\ &\qquad > (1 - f(p_{min}|e))w + f(p_{min}|e)c(p_{min}) \\ &\Rightarrow (1 - f(p_{min}|e))\mathbb{E}(C'(p, \beta'(p))|p \neq p_{min}) + f(p_{min}|e)(C'(p_{min}, 1) + \kappa) > \\ &\qquad > (1 - f(p_{min}|e))w + f(p_{min}|e)(c(p_{min}) + w - c(p_{min})) = w,\end{aligned}$$

because $\kappa > w - c(p_{min})$. Therefore, contract $\mathcal{W}'$ is more expensive than contract $\mathcal{W}$. This contradiction proves that $\beta(p) = 0$ for all $p \in \mathcal{P}$ is optimal for $\kappa > w - c(p_{min})$.

For the second step of the proof, again compare the principal's costs with and without justification. If she does not provide justification, the principal's costs are w according to Lemma 8.

If $e > 0$ and $\kappa < w - c(p_{min})$, the following contract implies lower costs for the principal:

$$C'(m_P, m_A) = \begin{cases} w & \text{if } m_P > p_{min} \\ c(p_{min}) & \text{if } m_P = p_{min}, \end{cases}$$

$$W'(m_P, m_A) = \begin{cases} w & \text{if } m_P > p_{min} \\ c(p_{min}) & \text{if } m_P = p_{min} \text{ and } m_A = 1 \\ j(w+\kappa) & \text{else} \end{cases}$$

with an appropriately chosen $j \in \mathbb{N}$. In this contract, the principal justifies evaluation p_{min}. Therefore, $\beta'(p_{min}) = 1$ and $\beta'(p) = 0$ for $p > p_{min}$. The costs of this contract are

$$f(p_{min}|e)(c(p_{min}) + \kappa) + (1 - f(p_{min}|e))w < f(p_{min}|e)w + (1 - f(p_{min}|e))w = w.$$

As justification costs are lower than third-party payments $w - c(p_{min})$, justifying low evaluations reduces the principal's costs. Finally, it remains to verify that contract $\mathcal{W}'$ is in the constraint set of Program A.

Contract $\mathcal{W}'$ satisfies the agent's participation constraint PC and incentive compatibility IC by definition of w and $c(p_{min})$. As the agent's wage is independent of his message, his truth-telling constraint TT_A is satisfied. Contract $\mathcal{W}'$ also satisfies constraint (1). Finally, consider the principal's truth-telling constraint TT_P. If $p = p_{min}$, the principal's equilibrium costs are $c(p_{min}) + \kappa$. Any deviation increases her costs to at least w. If $p > p_{min}$, her costs are w. Deviating to $\beta = 1$ and $m_P \geq p$ increases her costs to $w + \kappa$. Deviating to $\beta = 0$ and $m_P \neq p$ weakly increases her costs to w or $j(w + \kappa)$. Therefore, the principal's truth-telling constraint TT_P is satisfied. Hence, as long as justification costs are lower than this bound, the principal will justify some evaluations. If $e > 0$ and $\kappa = w - c(p_{min})$, above steps show that justification and no justification can be optimal. □

Proof of Proposition 5: The proof simplifies Program B from Proposition 2 and shows that the simplified program determines a contract in the constraint set of Program A. The proof proceeds in three steps. First, I show that $w(p) = c(p)$ for all $p \in \mathcal{C}$ in any solution of Program B. Second, I show that $w(p) = c(p)$ for all $p \in (\mathcal{P} \setminus \mathcal{C})$ in any solution of Program B. Third, I simplify Program B.

Begin with the first step. Suppose to the contrary that a solution to Program B includes a $p_4 \in \mathcal{C}$ with $w(p_4) > c(p_4)$. If $p_4 = p_{min}$, modify the contract by reducing $w(p_4)$ to $c(p_4)$. As the agent's wage remains unaffected, the modified contract satisfies PC and IC in Program B. It is easy to see that the modified contract also satisfies constraints (3), (4), and (5). Finally, the modified contract improves the principal's objective by $(w(p_4) - c(p_4))f(p_4|e)$. Now turn to the case $p_4 \neq p_{min}$. Proposition 3 implies $p_{min} \in \mathcal{C}$. Increase $c(p_4)$ by a small $\epsilon > 0$ to $c(p_4) + \epsilon$ and reduce $w(p_{min})$ and $c(p_{min})$ to $\tilde{c}^\epsilon$ with $\tilde{c}^\epsilon$ determined by

$$u(\tilde{c}^\epsilon) = u(c(p_{min})) - (u(c(p_4) + \epsilon) - u(c(p_4)))\frac{f(p_4|e)}{f(p_{min}|e)}.$$

This definition of $\tilde{c}^\epsilon$ ensures $\tilde{c}^\epsilon < c(p_{min}) \leq w(p_{min})$. This modification does not affect the agent's participation constraint. Additionally, the left-hand side of the incentive compatibility IC is now bigger than the marginal cost of effort, $d'(e)$, because the left-hand side of IC changes by

$$(f^H(p_{min}) - f^L(p_{min}))(u(\tilde{c}^\epsilon) - u(c(p_{min}))) + (f^H(p_4) - f^L(p_4))(u(c(p_4) + \epsilon) - u(c(p_4))) =$$
$$= \frac{f^H(p_{min}) - f^L(p_{min})}{f(p_{min}|e)} f(p_{min}|e)(u(\tilde{c}^\epsilon) - u(c(p_{min}))) + \frac{f^H(p_4) - f^L(p_4)}{f(p_4|e)} f(p_4|e)(u(c(p_4) + \epsilon) - u(c(p_4))) >$$
$$> \frac{f^H(p_{min}) - f^L(p_{min})}{f(p_{min}|e)} \Big(f(p_{min}|e)(u(\tilde{c}^\epsilon) - u(c(p_{min}))) + f(p_4|e)(u(c(p_4) + \epsilon) - u(c(p_4))) \Big) = 0.$$

The last equality follows from the definition of $\tilde{c}^\epsilon$. The main inequality follows from the monotone likelihood ratio property. Choosing ϵ sufficiently small ensures that the modified contract satisfies constraint (5) due to $c(p_4) + \epsilon \leq w(p_4)$. It is easy to see that the modified contract satisfies constraints (3) and (4), because reducing $w(p_{min})$ never violates constraint (4). Moreover, $\beta(p_{min}) = 1$ and reducing $w(p_{min})$ does not violate constraint (3). The modified contract implies lower expected costs for the principal compared to the initial contract. The modified contract satisfies all the constraints of Program B. Therefore, the modified contract contradicts optimality of the initial solution. Consequently, $w(p) = c(p)$ for all $p \in \mathcal{C}$.

In the second step, suppose to the contrary that a solution to Program B includes a $p_5 \in (\mathcal{P} \backslash \mathcal{C})$ with $w(p_5) > c(p_5)$. If $p_{min} \in \mathcal{C}$, $p_{min} \neq p_5$. Then increase $c(p_5)$ by a small $\epsilon > 0$ to $c(p_5) + \epsilon$ and reduce $w(p_{min})$ and $c(p_{min})$ to $\tilde{c}^\epsilon$ with $u(\tilde{c}^\epsilon) = u(c(p_{min})) - (u(c(p_5) + \epsilon) - u(c(p_5))) \frac{f(p_5|e)}{f(p_{min}|e)}$ as above. The remainder of the proof is analogous to the second part of the first step and, hence, omitted. If $p_{min} \notin \mathcal{C}$, Proposition 3 implies $\mathcal{C} = \emptyset$. Proposition 4 shows that $\mathcal{C} = \emptyset$ is not optimal for $\kappa < \bar{\kappa}$. This contradiction shows that $w(p) = c(p)$ for all $p \in \mathcal{P}$ with $\beta(p) = 0$.

Now turn to the third step. Proposition 4 shows that $\mathcal{C} \neq \emptyset$ is optimal for $\kappa < \bar{\kappa}$. Therefore, I restrict attention to $\mathcal{C} \neq \emptyset$ here. Then Proposition 3 implies that there is a $\delta \in \mathcal{P}$ such that $\mathcal{C} = \{p \in \mathcal{P} | p \leq \delta\}$. The first and the second step of this proof have shown that $w(p) = c(p)$ for all $p \in \mathcal{P}$. Hence, constraint (5) is satisfied. Plugging these results into Program B and neglecting constraints (3) and (4) yields the relaxed program:

$$\min_{w(p), w, \delta} \sum_{p \leq \delta} (w(p) + \kappa) f(p|e) + (1 - F(\delta|e)) w, \tag{C}$$

$$\text{subject to } \sum_{p \leq \delta} u\big(w(p)\big) f(p|e) + (1 - F(\delta|e)) u(w) - d(e) \geq \bar{u}, \tag{PC}$$

$$\sum_{p \leq \delta} u\big(w(p)\big)(f^H(p) - f^L(p)) + u(w) \sum_{p > \delta} (f^H(p) - f^L(p)) \geq d'(e) \tag{IC}$$

A solution to this program incentivizes the agent at minimal costs. Yet this solution might violate the principal's truth-telling constraint, namely, $w(p)$ non-decreasing and $w \geq \kappa + w(p)$. I show that a solution to Program C satisfies constraints (3) and (4). Therefore, a solution to the relaxed Program C is also a solution to the initial Program B.

Denote the Lagrange multiplier for the participation constraint by ν_1 and the one for the agent's incentive compatibility by ν_2. Program C yields the first-order conditions:

$$f(p|e) - \nu_1 u'(w(p))f(p|e) - \nu_2 u'(w(p))(f^H(p) - f^L(p)) = 0 \quad \text{for all } p \leq \delta$$
$$(1 - F(\delta|e)) - \nu_1 u'(w)(1 - F(\delta|e)) - \nu_2 u'(w) \sum_{p>\delta} (f^H(p) - f^L(p)) = 0$$

Rearranging yields

$$\nu_1 + \nu_2 \frac{f^H(p) - f^L(p)}{f(p|e)} = \frac{1}{u'(w(p))} \quad \text{for all } p \leq \delta \tag{10}$$
$$\nu_1 + \nu_2 \frac{\sum_{p>\delta} (f^H(p) - f^L(p))}{1 - F(\delta|e)} = \frac{1}{u'(w)} \tag{11}$$

The monotone likelihood ratio property guarantees that

$$\frac{\sum\limits_{p>\delta} (f^H(p) - f^L(p))}{1 - F(\delta|e)} = \frac{\sum\limits_{p>\delta} \frac{f^H(p) - f^L(p)}{f(p|e)} f(p|e)}{\sum\limits_{p>\delta} f(p|e)} > \frac{\frac{f^H(\delta) - f^L(\delta)}{f(\delta|e)} \sum\limits_{p>\delta} f(p|e)}{\sum\limits_{p>\delta} f(p|e)} = \frac{f^H(\delta) - f^L(\delta)}{f(\delta|e)}.$$

Additionally, as before, ν_2 is positive, because the solution to Program C without constraint IC is $\delta = 0$ and $w = u^{-1}(\bar{u} + d(e))$ violating IC. Then (10) implies that $w(p)$ increases in $p \leq \delta$ and, hence, (10) implies constraint (4) for $p \leq \delta$. In addition, equations (10) and (11) as well as the monotone likelihood ratio property imply $w > w(p)$ for $p \leq \delta$. It remains to verify $w \geq w(\delta) + \kappa$ to ensure constraints (3) and (4). If $\delta \notin \mathcal{P}$, set $\delta = \max\{p \in \mathcal{P} | p \leq \delta\}$ as a normalization.

Assume to the contrary that a solution $\mathcal{W}$ to Program C has $w < w(\delta) + \kappa$. Then $w(\delta) \in (w - \kappa, w)$ by the preceding steps. If $\delta = p_{min}$, the principal's costs are

$$(w(p_{min}) + \kappa)f(p_{min}|e) + w(1 - f(p_{min}|e)) > w.$$

Lemma 8 shows that there is a contract with lower costs. In particular, the principal could save at least $(w(p_{min}) + \kappa - w)f(p_{min}|e) > 0$. The reason is that the contract in Lemma 8 is the optimal contract satisfying PC and IC, while paying a constant wage to the agent for $p > p_{min}$. Therefore, w here is bigger than the principal's costs in Lemma 8 - a contradiction to Proposition 4 as $\kappa < \bar{\kappa}$. If $\delta > p_{min}$, modify contract $\mathcal{W}$ by increasing $w(\delta)$ to w reducing $w(p_{min})$ to $\tilde{w}$ determined by

$$u(\tilde{w}) = u(w(p_{min})) - (u(w) - u(w(\delta)))\frac{f(\delta|e)}{f(p_{min}|e)}.$$

As the agent's expected utility does not change, his participation constraint PC is satisfied. Additionally, the left-hand side of IC is now bigger than the marginal cost of effort, $d'(e)$, because

$$(f^H(p_{min}) - f^L(p_{min}))(u(\tilde{w}) - u(w(p_{min}))) + (f^H(\delta) - f^L(\delta))(u(w) - u(w(\delta))) =$$
$$= \frac{f^H(p_{min}) - f^L(p_{min})}{f(p_{min}|e)} f(p_{min}|e)(u(\tilde{w}) - u(w(p_{min}))) + \frac{f^H(\delta) - f^L(\delta)}{f(\delta|e)} f(\delta|e)(u(w) - u(w(\delta))) >$$
$$> \frac{f^H(p_{min}) - f^L(p_{min})}{f(p_{min}|e)} (f(p_{min}|e)(u(\tilde{w}) - u(w(p_{min}))) + f(\delta|e)(u(w) - u(w(\delta)))) = 0.$$

The last equality follows from the definition of $\tilde{w}$. The main inequality follows from the monotone likelihood ratio property. The modified contract satisfies all constraints of Program C and decreases the principal's costs by

$$(w(\delta) + \kappa - w)f(\delta|e) + (w(p_{min}) - \tilde{w})f(p_{min}|e) > 0.$$

This contradiction shows that $w \geq w(\delta) + \kappa$ and any solution to the relaxed problem also satisfies constraints (3) and (4). Therefore, any solution to Program C is in the constraint set of Program B. Proposition 2 shows that any contract in the constraint set of Program B determines a contract in the constraint set of Program A. Consequently, equations (10) and (11) in combination with the participation constraint and the agent's incentive compatibility determine an optimal contract. □

Proof of Proposition 6: Proposition 1 determines a contract $\mathcal{W}$ in the constraint set of Program A with $\mathcal{C} = \mathcal{P} \setminus \{p_{max}\}$ or equivalently $\delta = p_{2ndmax} = \max(\mathcal{P} \setminus \{p_{max}\})$. Proposition 1 shows that this contract yields lower costs than any contract with $\mathcal{C} = \mathcal{P}$. Given $\mathcal{C} = \mathcal{P} \setminus \{p_{max}\}$, the definition of $w_e^*(\cdot)$ in Lemma 1 guarantees that contract $\mathcal{W}$ is optimal for sufficiently small costs κ. If $|\mathcal{C}| < |\mathcal{P}| - 1$, Proposition 3 ensures that $\beta(p_{2ndmax}) = \beta(p_{max}) = 0$ in an optimal contract. According to Proposition 2, the principal's truth-telling constraint TT_P implies $W(p_{2ndmax}, 0) = W(p_{max}, 0)$ in such a contract. Therefore, an optimal contract implies at least wage costs (excluding justification costs) of a solution to Program A* with the additional constraint $W(p_{2ndmax}, 0) = W(p_{max}, 0)$. Lemma 1 proves that a solution to Program A* without this additional constraint has $W(p_{2ndmax}, 0) < W(p_{max}, 0)$. Consequently, the additional constraint is binding and the wage costs of an optimal contract with $|\mathcal{C}| < |\mathcal{P}| - 1$ are strictly higher than the wage costs of an optimal contract with $|\mathcal{C}| = |\mathcal{P}| - 1$. The difference between these costs is independent of κ. Hence, for sufficiently small κ, the additional justification costs κ in the contract with $\mathcal{C} = \mathcal{P} \setminus \{p_{max}\}$ are lower than the reduction in wage costs. Therefore, $\mathcal{C} = \mathcal{P} \setminus \{p_{max}\}$ is optimal for sufficiently small κ. □

Proof of Lemma 3: A probability space for the random variable p is constructed in the common way using the power set. For a given p, I require a suitable probability space for I. For this purpose, I follow the approach by Sun (2006). Hence, I consider a Fubini extension instead of the usual continuum product based on the Kolmogorov construction. Sun and Zhang (2009, Theorem 1 and Corollary 2) prove that there exist a set Ω, a probability space on Ω, an extension λ of the Lebesgue measure $\bar{\lambda}$ on $T = [0, 1]$, a Fubini extension on $T \times \Omega$ and a process $g\colon T \times \Omega \mapsto \mathbb{R}$, such that the random variables $g(t, \cdot)$ are essentially pairwise independent with the required distribution: $\Pr(\{\omega \in \Omega | g(t, \omega) = 1\}) = p$ for $t \in T$ almost surely. By definition of a Fubini extension, the integral $\int_T g(t, \omega) \mathrm{d}\lambda$ is well defined for all $\omega \in \Omega$. In addition, Sun (2006, Theorem 2.8) proves that the integral equals p almost surely. □

Proof of Lemma 4: If signals I and S are observable and contractible, the contract directly conditions on I and S and messages do not matter. Program $\hat{A}^*$ below determines how to optimally implement effort e by choosing payments $W(I, S)$ by the principal, $C(I, S)$ for the agent, and implicitly which evaluations to justify, $\beta(p)$. The objective is to minimize expected costs subject to several conditions. The participation constraint PC makes the agent accept the proposed contract. The agent's incentive compatibility IC guarantees that the agent chooses the desired level of effort. Finally, the principal's payment has to be higher than the wage for the agent.

$$\inf \int W(I,S) + \kappa\beta(\lambda(I))\mathrm{d}P(I,S|e) \tag{$\hat{A}^*$}$$

$$\text{subject to} \int u(C(I,S)) - d(e)\mathrm{d}P(I,S|e) \geq \bar{u} \tag{PC}$$

$$e \in \arg\max \int u(C(I,S)) - d(e)\mathrm{d}P(I,S|e) \tag{IC}$$

$$W(I,S) \geq C(I,S) \qquad \forall I \in \mathcal{I}, \forall S \in T. \tag{12}$$

Holmström (1979) shows that optimal wages only condition on aggregate evaluation $\mu = \lambda(I)$, because the average of the principal's information I is a sufficient statistics for the agent's effort, $\Pr(I,S|e,\lambda(I)) = \Pr(I,S|\lambda(I))$. Therefore, only $\lambda(I)$ has to be observable and contractible. In particular, the detailed signals I and S do not matter and Lemma 4 is valid if S is private information and is not contractible. The remainder of this proof is analogous to the proof of Lemma 1 and therefore omitted. □

Proof of Lemma 5: Consider the following contract $C(m_P, m_A) = w_e^*(\lambda(m_P))$ and

$$W(m_P, m_A) = \begin{cases} w_e^*(\lambda(m_P)) & \text{if } m_A \in m_P \\ jw_e^*(p_{max}) & \text{else} \end{cases}$$

with an appropriately chosen $j \in \mathbb{N}$ that I determine below. The contract induces truth-telling by the agent as the agent's wage is independent of her message m_A. In addition, the contract induces truth-telling by the principal and $\beta(p) = 1$ for all $p \in \mathcal{P}$. Consider deviations by the principal. Any deviation m_P with $\lambda(m_P) \geq \lambda(I)$ weakly increases the principal's costs, because $w_e^*(\cdot)$ is increasing by Lemma 4. Therefore, these deviations are unprofitable. If the principal deviates to m_P with $\lambda(m_P) < \lambda(I)$, her costs decrease to $w_e^*(\lambda(m_P))$ for $m_A \in m_P$, but increase to $jw_e^*(p_{max})$ for $m_A \notin m_P$. As the principal does not know the agent's sample S and, hence, his message m_A, she uses her prior knowing that S is uniformly distributed on I. The probability of decreasing her costs is thus at most $\frac{\lambda(m_P)}{\lambda(I)}$, i.e., the probability that a uniformly distributed random variable on I is within a set of mass $\lambda(m_P)$. Therefore, a lower bound on the change in the principal's costs is

$$\left(1 - \frac{\lambda(m_P)}{\lambda(I)}\right) jw_e^*(p_{max}) + \frac{\lambda(m_P)}{\lambda(I)} w_e^*(\lambda(m_P)) - w_e^*(\lambda(I))$$

Such a deviation increases the principal's costs if

$$\begin{aligned} &\frac{\lambda(I) - \lambda(m_P)}{\lambda(I)} jw_e^*(p_{max}) > w_e^*(\lambda(I)) - \frac{\lambda(m_P)}{\lambda(I)} w_e^*(\lambda(m_P)) \\ \Leftrightarrow \quad & j > \frac{\lambda(I)}{\lambda(I) - \lambda(m_P)} \left(\frac{w_e^*(\lambda(I))}{w_e^*(p_{max})} - \frac{\lambda(m_P)}{\lambda(I)} \frac{w_e^*(\lambda(m_P))}{w_e^*(p_{max})} \right) \\ \Leftarrow \quad & j > \frac{\lambda(I)}{\lambda(I) - \lambda(m_P)} \end{aligned}$$

The last implication follows from $w_e^*(p_{max}) \geq w_e^*(\lambda(I))$ as established in Lemma 4. Hence if

$$j > \max_{p,p' \in \mathcal{P}, p > p'} \frac{p}{p - p'},$$

any deviation m_P with $\lambda(m_P) < \lambda(I)$ is unprofitable. Notice that $\max_{p,p' \in \mathcal{P}, p>p'} \frac{p}{p-p'}$ is finite and bounded. Therefore, any deviation $m_P \neq I$ by the principal is unprofitable for an appropriately chosen $j \in \mathbb{N}$. Moreover, $\beta(p) = 1$ for all $p \in \mathcal{P}$.

The proposed contract is optimal, because it implements the equilibrium wage $w_e^*(\mu)$. As $w_e^*(\mu)$ is a solution to the relaxed Problem $\hat{A}^*$ according to Lemma 4, this contract is also a solution for $\kappa = 0$. Hence, it is impossible to incentivize the agent by a cheaper contract. □

Proof of Proposition 7: Suppose to the contrary that the principal justifies all evaluations, $\beta(p) = 1$ for all p. Then the principal's expected costs are at least $\kappa + \mathbb{E}(w_e^*(p))$ with $w_e^*(\cdot)$ defined in Lemma 1. It is possible, however, to implement effort e by the agent even cheaper by partial justification. For this purpose, consider the following contract with $C(m_P, m_A) = w_e^*(\lambda(m_P))$ and

$$W(m_P, m_A) = \begin{cases} w_e^*(\lambda(m_P)) & \text{if } \lambda(m_P) \neq p_{max} \text{ and } m_A \in m_P \\ \max\{w_e^*(p_{2ndmax}) + \kappa, w_e^*(p_{max})\} & \text{if } \lambda(m_P) = p_{max} \\ j w_e^*(p_{max}) & \text{else} \end{cases}$$

with an appropriately chosen $j \in \mathbb{N}$ and $p_{2ndmax} = \max(\mathcal{P} \backslash \{p_{max}\})$ the second highest evaluation. In this contract, for sufficiently high j the principal justifies all evaluations except the highest one and $\beta(p) = 1$ if and only if $p < p_{max}$. If the principal's information indicates very good performance, $p = p_{max}$, justification would increase her justification costs as the wage costs remain unchanged. Analogously to Proposition 1, this contract implements effort e of the agent and is cheaper than $\kappa + \mathbb{E}(w_e^*(p))$. In particular, the principal's expected costs decrease by

$$f(p_{max}|e)(w_e^*(p_{max}) + \kappa - \max\{w_e^*(p_{2ndmax}) + \kappa, w_e^*(p_{max})\}) > 0.$$

Similarly to Lemma 5, it is easy to show that the proposed contract induces truth-telling. This proves that the principal should not justify all evaluations for positive costs κ. □

Proof of Proposition 8: In the first step, I characterize equilibrium behavior for the contract of Proposition 8. The second step shows that this contract is optimal.

Begin with the first step. It is easy to see that $C^*(m_A, m_P) \leq W^*(m_A, m_P)$ for any messages m_A and m_P in the contract of Proposition 8. Suppose the principal uses the strategy

$$\beta(p) = 1 \text{ if and only if } p \in \mathcal{P} \text{ and } p \leq \delta$$

$$m_P = \begin{cases} I & \text{if } p \in \mathcal{P} \text{ and } p \leq \delta \\ [0, p] & \text{else} \end{cases}$$

As the agent's wage does not depend on his message m_A, the strategy $m_A = S$ is optimal for the agent. The agent's optimal effort is $e \in \arg\max \sum_{p \leq \delta} u\big(w^{**}(p)\big) f(p|e) + (1 - F(\delta|e)) u(w^{**}) - d(e)$, independent of his message. By the same arguments as in Lemma 1, the first-order approach is valid here. Therefore, the agent's incentive compatibility is equivalent to

$$\sum\nolimits_{p\leq\delta} u\big(w^{**}(p)\big)(f^H(p)-f^L(p)) + u(w^{**})\sum\nolimits_{p\leq\delta}(f^H(p)-f^L(p)) \geq d'(e).$$

Hence, the definition of w^{**}, $w^{**}(p)$ and δ in Proposition 5 ensures that the desired effort level e is optimal for the agent. The agent accepts the contract if

$$\sum\nolimits_{p\leq\delta} u\big(w^{**}(p)\big)f(p|e) + (1-F(\delta|e))u(w^{**}) - d(e) \geq \bar{u}.$$

Again the definition of w^{**}, $w^{**}(p)$ and δ in Proposition 5 guarantees that this condition is satisfied and the agent accepts the contract. To conclude the first step, I show that the contract makes the principal report truthfully. Suppose the agent accepts the contract, exerts effort e and sends message $m_A = S$. If $p > \delta$, the principal's costs after reporting message m_P are similar to Lemma 5 and equal

$$\begin{cases} = w^{**} & \text{if } \beta = 0 \text{ and } \lambda(m_P) > \delta \\ = w^{**} + \kappa & \text{if } \beta = 1 \text{ and } \lambda(m_P) > \delta \\ \geq (1-\frac{\lambda(m_P)}{p})j(w^{**}+\kappa) + \frac{\lambda(m_P)}{p}w^{**}(\lambda(m_P)) & \text{if } \lambda(m_P) \leq \delta. \end{cases}$$

If $p > \delta$ and $j > \max_{q,q'\in\mathcal{P},q>q'} \frac{q}{q-q'}$, there is no profitable deviation for the principal, as in the proof of Lemma 5. It remains to consider the case $p \leq \delta$. Then the principal's costs after reporting message m_P are

$$\begin{cases} \geq w^{**} & \text{if } \lambda(m_P) > \delta \\ \geq (1-\frac{\lambda(m_P)}{p})j(w^{**}+\kappa) + \frac{\lambda(m_P)}{p}w^{**}(\lambda(m_P)) & \text{if } m_P \neq I \text{ and } \lambda(m_P) \leq \delta \\ = w^{**}(p) + \kappa & \text{if } m_P = I. \end{cases}$$

Proposition 5 ensures $w^{**} \geq w^{**}(p') + \kappa$ for all $p' \in \mathcal{P}$ and $p' \leq \delta$. As in the proof of Lemma 5, there is no profitable deviation for the principal if $j > \max_{q,q'\in\mathcal{P},q>q'} \frac{q}{q-q'}$ and $p \leq \delta$. Consequently, truthful reporting is optimal for the principal and the principal's strategy I stated in the beginning of this proof is optimal.

Now turn to the second step. The easiest way to study the problem is in terms of equilibrium utilities and certainty equivalents, respectively. Define expected payments for the principal given equilibrium strategies and evaluation p by $w(p)$. Similarly, define the certainty equivalent of the agent's wages given equilibrium strategies and evaluation p by $c(p)$. As the agent is unaware of p, $c(p)$ is a purely theoretical concept to analyze the contract. Finally denote the principal's equilibrium justification strategy by $\beta(p)$. The agent only accepts the contract if

$$\sum\nolimits_{p\in\mathcal{P}} u\big(c(p)\big)f(p|e) - d(e) \geq \bar{u}. \tag{13}$$

The agent implements effort e if

$$e \in \arg\max \sum\nolimits_{p\in\mathcal{P}} u\big(c(p)\big)f(p|e) - d(e).$$

By the same arguments as in Lemma 1, the first-order approach is valid here. Therefore, the agent implements effort e if

$$\sum\nolimits_{p\in\mathcal{P}} u\big(c(p)\big)(f^H(p)-f^L(p)) \geq d'(e). \tag{14}$$

For $\beta(p) = 0$, the agent cannot check whether the principal reports her information p truthfully. Hence, optimal equilibrium play requires that equilibrium payments are constant in the principal's message as I prove here: Suppose to the contrary that there are $p_1, p_2 \in \mathcal{P}$ with $\beta(p_1) = \beta(p_2) = 0$ and $w(p_1) \neq w(p_2)$. Without loss of generality assume $w(p_1) > w(p_2)$. If p_1 has realized, the prinicpal could deviate from equilibrium strategies and use the reporting strategy as if p_2 had realized. The agent is unaware of this deviation and, in addition, p_2 and the resulting message are consistent with his information S and e, i.e., probabilities of p_2 conditionally on S and e are strictly positive. Hence, the agent does not notice the principal's deviation from her message and follows his equilibrium strategy in reporting message m_A. Therefore, the principal expects a wage $w(p_2)$ following this deviation. By assumption, $w(p_1) > w(p_2)$ yielding a contradiction to optimality of equilibrium play. Consequently, $w(p_1) = w(p_2)$ for all $p_1, p_2 \in \mathcal{P}$ with $\beta(p_1) = \beta(p_2) = 0$. Similarly, optimal equilibrium play requires that equilibrium costs to the principal are lower with justification than without justification as I prove here: Suppose to the contrary that there are $p_1, p_2 \in \mathcal{P}$ with $\beta(p_1) = 0$ and $\beta(p_2) = 1$ and $w(p_1) < w(p_2) + \kappa$. If p_2 has realized, the prinicpal could deviate from equilibrium strategies and use the reporting strategy with $\beta = 0$ and m_P as if p_1 had realized. The agent is unaware of this deviation and, in addition, p_1 and the resulting message are consistent with his information S and e, i.e., probabilities of p_1 conditionally on S and e are strictly positive. Hence, the agent does not notice the principal's deviation from her message and follows his equilibrium strategy in reporting message m_A. Therefore, the principal expects costs of $w(p_1)$ following this deviation. By assumption, $w(p_1) < w(p_2) + \kappa$ yielding a contradiction to optimality of equilibrium play. Consequently, $w(p_1) \geq w(p_2) + \kappa$ for all $p_1, p_2 \in \mathcal{P}$ with $\beta(p_1) = 0$ and $\beta(p_2) = 1$. Together, optimal equilibrium play implies

$$\bar{w} \geq w(p) + \kappa\beta(p) \geq (1 - \beta(p))\bar{w} \qquad \forall p \in \mathcal{P}. \tag{15}$$

Next, optimal equilibrium play requires that the principal's equilibrium costs are non-decreasing in p. Suppose to the contrary that there are $p_L, p_H \in \mathcal{P}$ with $w(p_L) + \kappa\beta(p_L) > w(p_H) + \kappa\beta(p_H)$ and $p_L < p_H$. If p_L has realized, the prinicpal could deviate from equilibrium strategies and use the reporting strategy $\beta(p_H)$. For $\beta(p_H) = 0$, she uses the reporting strategy as if p_H had realized. For $\beta(p_H) = 1$, she constructs a new set I' by adding additional categories to the true I until $\lambda(I') = p_H$. Then she reports as if I' had realized. The agent is unaware of this deviation and the resulting messages are in both cases consistent with his information S and e, i.e., probabilities of p_H and I' conditionally on S and e are strictly positive. Hence, the agent does not notice the principal's deviation from her message and follows his equilibrium strategy in reporting message m_A. Therefore, the principal expects costs of $w(p_H) + \kappa\beta(p_H)$ following this deviation. By assumption, $w(p_L) + \kappa\beta(p_L) > w(p_H) + \kappa\beta(p_H)$ yielding a contradiction to optimality of equilibrium play. Consequently,

$$w(p) + \kappa\beta(p) \text{ is non-decreasing in } p. \tag{16}$$

Finally, the principal's payment has to be at least the agent's wage and the agent is risk averse. Therefore,

$$w(p) \geq c(p) \qquad \forall p \in \mathcal{P}. \tag{17}$$

Then we can write down a relaxed version of the principal's initial problem. The principal is minimizing her expected costs:

$$\min \sum\nolimits_{p \in \mathcal{P}} (w(p) + \kappa\beta(p)) f(p|e) \tag{B*}$$

subject to constraints (13), (14), (15), (16) and (17). This program equals Program B. Lemma 7 in this appendix proves that in any solution to Program B there is δ, so that $\beta(p) = 1$ if and only if $p \leq \delta$. Proposition 4 proves that $\delta < p_{max}$ for $\kappa < \bar{\kappa}$ and $e > 0$. Proposition 5 proves that the solution to Program B can be determined by analyzing Program C. The first step of this proof shows that the contract in Proposition 8 is feasible and yields equilibrium utilities that solve Program C. Consequently, Proposition 8 characterizes optimal contracts. □

Proof of Corollary 1: Corollary 1 is a direct consequence of Proposition 8. □

Proof of Corollary 2: For sufficiently large κ, Propositions 4 and 6 show that wages for several evaluations are pooled as $|\{p \in \mathcal{P}|p > \delta\}| > 1$: wages for evaluations above δ equal w^{**} and are constant. Therefore, there is maximal wage compression at the top and centrality. □

Proof of Corollary 3: For sufficiently large κ, Propositions 4 and 6 show that the principal pools several evaluations in the indirect mechanism as $|\{p \in \mathcal{P}|p > \delta\}| > 1$. Therefore, optimal contracts imply leniency as

$$\mathbb{E}(\lambda(m_P)) = \sum\nolimits_{p \leq \delta} pF(p|e) + p_{max} \sum\nolimits_{p > \delta} F(p|e) = \mathbb{E}(p) + \sum\nolimits_{p > \delta} (p_{max} - p) F(p|e) > \mathbb{E}(p),$$

because $p_{max} \geq p$ with a strict inequality for a $p > \delta$. In addition, optimal contracts imply centrality as $\mathrm{Var}(p|p > \delta) > 0 = \mathrm{Var}(\lambda(m_P)|p > \delta)$. □

Acknowledgements: Financial support from the Max Planck Institute for Research on Collective Goods, the German Research Foundation (DFG) through the Bonn Graduate School of Economics (BGSE), Collaborative Research Centre (SFB/TR 15), Research Training Group (GRK 629), and from the Leibniz Association through Leibniz Science Campus - Berlin Centre for Consumer Policies (BCCP), and Leibniz Competition - Berlin Economics Research Associates (BERA) is gratefully acknowledged.

References

Addison, J. and Belfield, C. R. (2008). The Determinants of Performance Appraisal Systems: A Note (Do Brown and Heywood's Results for Australia Hold Up for Britain?). *British Journal of Industrial Relations*, 46(3):521–531.

Al-Najjar, N. I., Anderlini, L., and Felli, L. (2006). Undescribable Events. *Review of Economic Studies*, 73(4):849–868.

Al-Najjar, N. I. and Casadesus-Masanell, R. (2001). Discretion in Agency Contracts. *Harvard Business School Working Paper*, 02–015.

Anderlini, L. and Felli, L. (1994). Incomplete Written Contracts: Undescribable States of Nature. *Quarterly Journal of Economics*, 109(4):1085–1124.

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

Bernardin, H. J. and Orban, J. A. (1990). Leniency Effect as a Function of Rating Format, Purpose for Appraisal, and Rater Individual Differences. *Journal of Business and Psychology*, 5(2):197–211.

Bretz, R. D., Milkovich, G. T., and Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions, and implications. *Journal of Management*, 18(2):321–352.

Brown, M. and Heywood, J. S. (2005). Performance Appraisal Systems: Determinants and Change. *British Journal of Industrial Relations*, 43(4):659–679.

Compte, O. (1998). Communication in Repeated Games with Imperfect Private Monitoring. *Econometrica*, 66(3):597–626.

Dessler, G. (2008). *Human resource management*. Pearson Prentice Hall, 11th edition.

Dewatripont, M. and Tirole, J. (2005). Modes of Communication. *Journal of Political Economy*, 113(6):1217–1238.

Doornik, K. (2010). Incentive contracts with enforcement costs. *Journal of Law, Economics, and Organization*, 26(1):115–143.

Fuchs, W. (2007). Contracting with Repeated Moral Hazard and Private Evaluations. *American Economic Review*, 97(4):1432–1448.

Fuchs, W. (2014). Subjective Evaluations: Discretionary Bonuses and Feedback Credibility. *American Economic Journal: Microeconomics*, forthcoming.

Gale, D. and Hellwig, M. (1985). Incentive-Compatible Debt Contracts: The One-Period Problem. *Review of Economic Studies*, 52(4):647–663.

Gibbons, R. (1998). Incentives in organizations. *Journal of Economic Perspectives*, 12(4):115–132.

Gibbs, M. J., Merchant, K. A., Van Der Stede, W. A., and Vargus, M. E. (2009). Performance Measure Properties and Incentive System Design. *Industrial Relations*, 48(2):237–264.

Giebe, T. and Gürtler, O. (2012). Optimal contracts for lenient supervisors. *Journal of Economic Behavior & Organization*, 81(2):403–420.

Goldluecke, S. and Kranz, S. (2013). Renegotiation-proof relational contracts. *Games and Economic Behavior*, 80:157–178.

Golman, R. and Bhatia, S. (2012). Performance evaluation inflation and compression. *Accounting, Organization and Society*, 37:534–543.

Grossman, S. J. and Hart, O. D. (1983). An analysis of the principal-agent problem. *Econometrica*, 51(1):7–45.

Hall, B. J. and Madigan, C. (2000). Compensation and Performance Evaluation at Arrow Electronics. *Harvard Business School Case*, 800-290.

Hart, O. and Moore, J. (1998). Default and Renegotiation: A Dynamic Model of Debt. *Quarterly Journal of Economics*, 113(1):1–41.

Holmström, B. (1979). Moral Hazard and Observability. *Bell Journal of Economics*, 10(1):74–91.

Jawahar, I. M. and Stone, T. H. (1997). Appraisal Purpose versus Perceived Consequences: The Effects of Appraisal Purpose, Perceived Consequences, and Rater Self-Monitoring on Leniency of Ratings and Decisions. *Research and Practice in Human Resource Management*, 5(1):33–54.

Kampkötter, P. and Sliwka, D. (2017). More Dispersion, Higher Bonuses? The Role of Differentiation in Subjective Performance Evaluations. *Journal of Labor Economics*, forthcoming.

Krasa, S. and Villamil, A. P. (2000). Optimal Contracts when Enforcement Is a Decision Variable. *Econometrica*, 68(1):119–134.

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

Lang, M. (2017). Subjective Evaluations require Stochastic Contracts for Budget Balance. *Mimeo*.

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

Li, J. and Matouschek, N. (2013). Managing Conflicts in Relational Contracts. *American Economic Review*, 103(6):2328–2351.

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

MacLeod, W. B. (2007). Reputations, Relationships, and Contract Enforcement. *Journal of Economic Literature*, 45(3):595–628.

MacLeod, W. B. and Malcomson, J. M. (1989). Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment. *Econometrica*, 57(2):447–480.

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

MacLeod, W. B. and Tan, T. Y. (2016). Optimal Contracting with Subjective Evaluation: The Effects of Timing, Malfeasance and Guile. *NBER Working Paper*, 22156.

Malcomson, J. M. (2012). Relational Incentive Contracts. In Gibbons, R. and Roberts, J., editors, *Handbook of Organizational Economics*. Princeton University Press. ch. 25:1014–1065.

Milkovich, G. T., Newman, J. M., and Milkovich, C. (2008). *Compensation*. McGraw-Hill/Irwin, 9th edition.

Murphy, K. J. (1993). Performance measurement and appraisal: Merck tries to motivate managers to do it right. *Employment Relations Today*, 20(1):47–62.

Porter, C., Bingham, C., and Simmonds, D. (2008). *Exploring human resource management*. McGraw-Hill Education, London.

Puhani, P. A. and Yang, P. (2017). Does increased accountability decrease leniency in performance ratings? *Working Paper*.

Rahman, D. (2012). But Who will Monitor the Monitor? *American Economic Review*, 102(6):2767–97.

Rogerson, W. P. (1985). The First-Order Approach to Principal-Agent Problems. *Econometrica*, 53(6):1357–1367.

Rotemberg, J. J. and Saloner, G. (1993). Leadership Style and Incentives. *Management Science*, 39(11):1299–1318.

Spence, J. R. and Keeping, L. (2011). Conscious Rating Distorting in Performance Appraisal: A Review, Commentary, and Proposed Framework for Research. *Human Resource Management Review*, 21:85–95.

Sun, Y. (2006). The exact law of large numbers via Fubini extension and characterization of insurable risks. *Journal of Economic Theory*, 126(1):31–69.

Sun, Y. and Zhang, Y. (2009). Individual risk and Lebesgue extension without aggregate uncertainty. *Journal of Economic Theory*, 144(1):432–443.

Suvorov, A. and van de Ven, J. (2009). Discretionary rewards as a feedback mechanism. *Games and Economic Behavior*, 67(2):665–681.

Taylor, E. K. and Wherry, R. J. (1951). A study of leniency in two rating systems. *Personnel Psychology*, 4(1):39–47.

Townsend, R. M. (1979). Optimal Contracts and Competitive Markets with Costly State Verification. *Journal of Economic Theory*, 21(2):265–293.

Zábojník, J. (2014). Subjective evaluations with performance feedback. *RAND Journal of Economics*, 45(2):341–369.