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 2012/14

Communicating Subjective Evaluations

Matthias Lang

September 2009, this version April 2018

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 2018

Abstract

Consider managers evaluating their employees' performance. Should managers justify their subjective evaluations? To answer this question, I study justifications: Suppose a manager's evaluation is private information. To justify her evaluation, she can gather additional information that allows the agent to partially cross-check the evaluation. 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 discriminate poorly between good performances. This pattern was attributed to biased managers. I show that these 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, empirically, subjective evaluations

*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.

**University of Munich (LMU), Dept. of Economics, Kaulbachstr. 45, Munich, 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).

are distorted and wage dispersion for the best evaluations is low.[3] These empirical observations are referred to as biases – 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 adjusting wages and communication to the subjectivity of evaluations is optimal and does not require any bias. The wage and communication pattern results from optimal contracting with standard preferences.

In addition to explaining these empirical observations, 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. The principal has two options. Either she reports directly her evaluation or she justifies her evaluation by acquiring additional information, 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. 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., agents receive the highest wages more often than the best performance occurs, 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 suspicions, the principal justifies bad evaluations. Note that compared to common moral-hazard settings additional incentives are necessary: ex-ante the principal wants to justify bad evaluations

[3]Section 3.4 discusses these empirical observations in detail and provides references.

[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.

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, instead of reporting an evaluation yielding higher payments the principal would deviate and report an evaluation that does not require justification and yields 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 very limited. First, they have mass zero with respect to the principal's observations. Second, the agent cannot infer anything about the principal's evaluation from this information. Giving more information to the agent makes it easier for him to detect distorted reports by the principal. 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. Justifications apply 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 as defined above.

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

[5] Murphy (1993, p. 49) summarizes the reasoning as follows: Principals have "nonpecuniary costs [here, justification/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."

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), Pearce and Stacchetti (1998), 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 Proposition 1. 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 incentives – resulting from the contract and justification costs – to withhold and distort her evaluation that yields compression at the top.

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 technologies.

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. Thus, 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. Restricting contracts to use either the principal's offer or the verified output, 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 can make wages dependent on the fact whether verification occurred, and, if verified, the state of the world. In my setting, there is no verifiable output, but contracts can use any available information and there are no exogenous restrictions on the contracting space. Doornik (2010) has pooling at the bottom and at the top. As the agent triggers costly enforcement, optimal contracts are determined by two indifference conditions. The difference between the contractual payments and the settlement offer has to make the agent indifferent whether or not to accept the principal's offer. In addition, the agent's rejection rate makes the principal indifferent between offering the high wage and the low wage. Both indifference conditions are absent from my analysis. As with the previous papers, in my model, it is the better-informed party that makes the verification decision and optimal justification is deterministic.

Furthermore, leniency and wage compression or centrality 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 (2015) 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 Evaluating the Agent's Work

3.1 Subjective Evaluations at Work

As motivation for my information structure, in particular for the second technology, 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.

Suppose the principal can acquire a signal in each category. For example, she listens to costumers praising the employee. She observes the agent at work or the agent's output. She talks to the agent's colleagues to learn about his skills as a team worker. 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. Providing justification is costly, as it requires the principal to spend additional time and effort on the evaluation.[6] The next section formalizes these notions.

3.2 Setting

Consider a risk-averse agent (he) working for a risk-neutral principal (she). 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

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

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

Figure 1: Timing of the Model

justification, her message space is $\mathcal{M}_P^i$ depending on the technology $i \in \{1, 2\}$ specified below. If she provides justification, she pays costs κ to use the available justification technology $i \in \{1, 2\}$. Let $\beta \in \{0, 1\}$ denote the principal's justification decision. For $\beta = 1$, she justifies her evaluation of the agent's work. After observing the principal's choice β and her message m_P, the agent replies with an unverifiable message $m_A \in \mathcal{M}_A^i$ depending on the technology $i \in \{1, 2\}$. 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 (2018). 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. Figure 1 summarizes the timing.

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 $d'(0) = 0$ and 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.

I suppose that one of the following two justification technologies $i \in \{1, 2\}$ is available and determines message spaces $\mathcal{M}_P^i$, $\mathcal{M}_A^i$, and the set I^i that is yet to be defined.

3.2.1 Justification Technology 1

If technology 1 is available and the principal spends costs κ to use it, she sends a message $m_P \in [p, 1] \cap \mathcal{P}$. If she decides to forgo justification, she send a message $m_P \in \mathcal{M}_P^1 = \mathcal{P}$. The agent observes the principal's justification decision and her message and replies with a message $m_A \in \mathcal{M}_A^1 = \{0, 1\}$. To ease comparison to technology 2, denote by $I^1 = [0, p]$ for technology 1.

3.2.2 Justification Technology 2

If technology 2 is available and the principal spends costs κ to use it, she 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, $I(t) = 1$, with probability p or failure, $I(t) = 0$ with probability $1-p$ depending on the evaluation p specified above. The signals $I(t)$ are essentially pairwise independent as defined by Sun (2006, Definition 2.7).

Lemma 1. *There is a probability space that satisfies these assumptions and guarantees a law of large numbers.*

Independently of the principal's choice β, her message space is

$$m_P \in \mathcal{M}_P^2 = \mathcal{I} = \{X \subseteq [0,1] | X \text{ is } \lambda\text{-measurable and } \lambda(X) \in \mathcal{P}\}$$

with an extension λ of the Lebesgue measure on T as introduced by Sun and Zhang (2009, Theorem 1). Also independently of the principal's choice β, the agent observes one category S drawn uniformly from the successful categories $I^2 = \{t \in T | I(t) = 1\}$.[7] The principal does not know which category the agent observes. Note 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 $m_A \in \mathcal{M}_A^2 = T$. Define the subjective evaluation as the average $\lambda(I^2)$ of the principal's signals $I(t)$ in the categories T following the case study of Section 3.1. According to Sun (2006, Theorem 2.8), this average $\lambda(I^2)$ is well defined and equals p almost surely. Therefore, the evaluation $\lambda(I^2)$ is a sufficient statistics for the agent's effort.

3.3 Analysis

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

$$\max_{e \in [0,1)} B \sum_{p \in \mathcal{P}} p f(p|e) - \Pi(e).$$

Returning to the first step, I determine an optimal contract that implements effort e. Therefore, I consider contracts *feasible* if the agent accepts them, the contract incentivizes the agent to choose effort e and $W(m_P, m_A) \geq C(m_P, m_A)$ for any combination of messages.

Begin with a benchmark which provides a lower bound on the principal's expected costs for any feasible contract. If information I^i for justification technology $i \in \{1,2\}$ were observable and

[7] In Section 4 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.

contractible, only aggregate evaluations $\lambda(I^i)$ matter. In particular, the optimal contract neglects the detailed assessments in the different categories even if they were freely available.

Lemma 2. *Suppose technology $i \in \{1,2\}$ is available. If information I^i is observable and contractible, optimal wages $w_e^*(\cdot)$ depend only on the evaluation $\lambda(I^i)$. Wages $w_e^*(\cdot)$ are increasing for positive effort, $e > 0$. There is no justification.*

As Holmström (1979) shows, with contractible information, the wage depends only on the sufficient statistics $\lambda(I^i)$ instead of the entire information I^i. In addition, better evaluations imply higher wages, because the monotone likelihood ratio property implies that paying higher wages for better evaluations incentives the agent to exert work effort. This means that the principal shifts risk to the agent. To accept such a contract, the agent requires some compensation. To keep this compensation to a minimum, the principal reduces the variance in the wage payments. Reducing the variance in wage payments completely would eliminate the agent's incentives to exert effort. Therefore, the principal uses upward-sloping wage payments. For the remainder of this paper, I neglect the indices $i \in \{1,2\}$ for the information I^i and the message sets $\mathcal{M}_A^i$ and $\mathcal{M}_P^i$ as the relevant technology will be clearly specified in the context.

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.

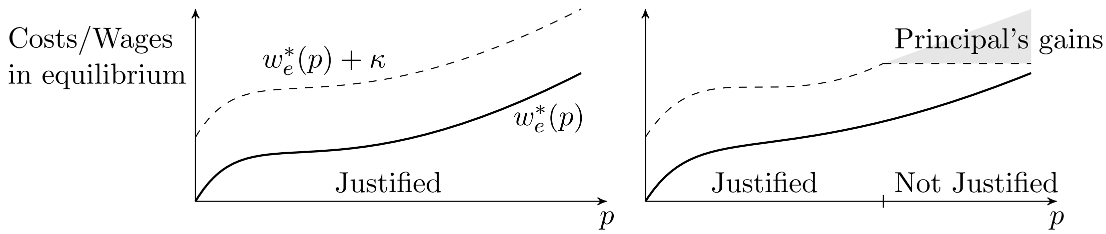
Proposition 1. *If there are no justification costs and $\kappa = 0$, there is an optimal contract in which all evaluations are justified. Equilibrium wages are the same as in Lemma 2. The contract*

$$C(m_P, m_A) = w_e^*\left(\lambda(m_P)\right) \quad \textit{and} \quad W(m_P, m_A) = \begin{cases} w_e^*\left(\lambda(m_P)\right) & \textit{if } m_A \in m_P \\ j w_e^*\left(p_{max}\right) & \textit{else} \end{cases}$$

is optimal for justification technology 2. The proof specifies the value of $j \in \mathbb{N}$ and provides the corresponding contract for technology 1.

The proposition constructs the optimal contract that is unique in terms of equilibrium utilities. The contract differs from the contract in Lemma 2 and depends on the justification technology. In particular, payments depend on whether messages agree. For technology 1, messages agree if the agent observes and reports justification. For technology 2, messages agree, if the category reported by the agent is contained in the principal's report of successful categories, i.e., $m_A \in m_P$. On the equilibrium path, messages agree. Therefore, the equilibrium wage is $w_e^*(p)$ – the same as before. In particular, equilibrium wages do not depend on the detailed information in the different categories. If messages agree, the principal's payment equals the agent's wage.

If the principal provides justification, the agent can (partially) detect deviations by the principal. The reporting incentives for the principal are straightforward for technology 1. For technology 2, the principal's incentives are as follows: If the principal reports more successful categories



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. For ease of exposition, I draw wages as continuous, although they are discrete.

Figure 2: Idea of the Proof of Lemma 3

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 almost surely 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 that differs from the agent's wage in this case. The proof shows that there is no profitable deviation for the principal. Hence, reporting $m_P = I$ is optimal for the principal.

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 Proposition 1. 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.

Lemma 3. *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 2. As demonstrated in the proof, the new contract is feasible. 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. It proves convenient to characterize contracts in terms of equilibrium utilities. In a first step, we derive a tighter lower bound for the principal's expected payments.

Lemma 4. *Program B is a relaxed version of the principal's problem:*

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

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

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

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

$$w(p) + \kappa\beta(p) \text{ non-decreasing in } p \tag{2}$$

$$w(p) \geq c(p) \qquad \forall p \in \mathcal{P} \tag{3}$$

For effort e, Program B determines equilibrium utilities and justification in an optimal contract that justifies evaluations according to $\beta(p)$. The objective is to minimize expected costs subject to five 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, any implementable equilibrium utilities satisfy constraints (1) and (2). 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. For the same reason, the principal's payment without justification has to be higher than payments with justification. Otherwise, the principal would deviate by not justifying the evaluation and paying the lower payment that requires no justification. The contract cannot detect such a deviation. The high pooling wage guarantees that this deviation is unprofitable. Finally, the principal's payments have to be higher than the agent's wage. As the next lemma shows, Program B points to threshold rules as optimal communication. Hence, the principal justifies only bad evaluations. Receiving a bad evaluation, the agent suspects a distortion by the principal who alleviates these fears by providing justification.

Lemma 5. *In any solution to Program B, there is a $\delta \in [0, 1)$ with $\beta(p) = 1$ if and only if $p \leq \delta$.*

Before turning to formal arguments, consider the following intuition. If optimal communication does not follow a threshold rule as in Lemma 5, 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 Lemma 4. 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. This slackness allows decreasing the principal's costs by cleverly modifying the contract. Therefore, the initial communication pattern cannot be optimal and the principal justifies only bad evaluations. Looking some steps ahead, Proposition 2 indeed shows that such a communication pattern, as depicted in Figure 3, is optimal for the principal. To see whether it is optimal to justify any evaluations at all, consider the optimal contract when the principal provides no justifications.

Lemma 6. *If the principal does not justify any evaluations, the following contract is optimal:*

$$C(m_P, m_A) = \begin{cases} u^{-1}\left(\bar{u} + d(e) - \frac{(1-f(p_{min}|e))d'(e)}{f^L(p_{min})-f^H(p_{min})}\right) & \text{if } m_P = p_{min} \text{ or } m_P = [0, p_{min}] \\ u^{-1}\left(\bar{u} + d(e) + \frac{f(p_{min}|e)d'(e)}{f^L(p_{min})-f^H(p_{min})}\right) & \text{else} \end{cases}$$

and $W(m_P, m_A) = u^{-1}\left(\bar{u} + d(e) + \frac{f(p_{min}|e)d'(e)}{f^L(p_{min})-f^H(p_{min})}\right)$ *for all* m_P *and all* m_A*. This contract is unique in terms of equilibrium utilities.*

Yet, the next lemma shows that such a contract is only optimal if justification costs are prohibitively high.

Lemma 7. *The principal optimally justifies some evaluations if she wants to implement positive effort* $e > 0$ *and justification costs* κ *are at most* $\bar{\kappa}$ *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* $\bar{\kappa}$*.*

If justification costs κ are prohibitively high, it is not optimal to use justification and, hence, $\delta = 0$. For moderate justification costs, the principal justifies some evaluations and the justification threshold is at least p_{min}. Before we can determine optimal contracts and justification thresholds δ for moderate costs, there is a final step missing to simplify Program B.

Lemma 8. *Suppose* $\kappa \leq \bar{\kappa}$*. The solution to Program C satisfies* $w \geq w(p)+\kappa$ *and* $w(p)$ *increasing in* p *for all* $p \in \{p' \in \mathcal{P} | p' \leq \delta\}$*:*

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

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

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

The lemma states that the monotone likelihood ratio property and optimal communication together imply constraints (1) and (2) in Program B. The wages that the principal justifies in equilibrium reflect just the participation constraint and the incentive compatibility. It is the monotone likelihood ratio property that ensures that $w(p)$ increases in p. The pooling wage is higher than the wages with justification because of the definition of $\bar{\kappa}$ and the communication pattern established in Lemma 5. Combining these results allows deriving optimal contracts.

Proposition 2. *Suppose* $\kappa \leq \bar{\kappa}$*. For technology 1, the following contract is optimal:*

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

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

Program C in Lemma 8 determines the values of δ, w^{**} and $w^{**}(p)$ for $p \leq \delta$. The corresponding contract for technology 2 is similar and implements the same equilibrium utilities.

The agent's wage depends only on the principal's message m_P. For good evaluations, the wage is constant and equals w^{**}. The agent's wage equals the principal's payments for good evaluations, which the principal does not justify. Varying the agent's wage in these cases means reducing the agent's wages. This reduction might ease 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 for good evaluations and to justify bad evaluations that is evaluations with low likelihood ratios. Therefore, the agent's wage equals the principal's payments for good evaluations without justification. For $\kappa > \bar{\kappa}$, there is no justification at all and Lemma 6 shows that the equality between the agent's wages and the principal's payments no longer holds.

The principal justifies any variation in the agent's wages. She does not justify the highest wage w^{**}. I postpone the discussion of this communication pattern to Corollary 1. If the principal reports an evaluation below δ, her payments depend also on the agent's message. In particular, her payments could differ from the agent's wage. On the equilibrium path, this case never occurs. The agent has the choice whether to confirm the principal's message. For technology 1, he does this by reporting $m_A = 1$. In equilibrium, he sends such a message whenever he observes a justification by the principal. We can interpret the agent's message as what the agent has observed regarding β.

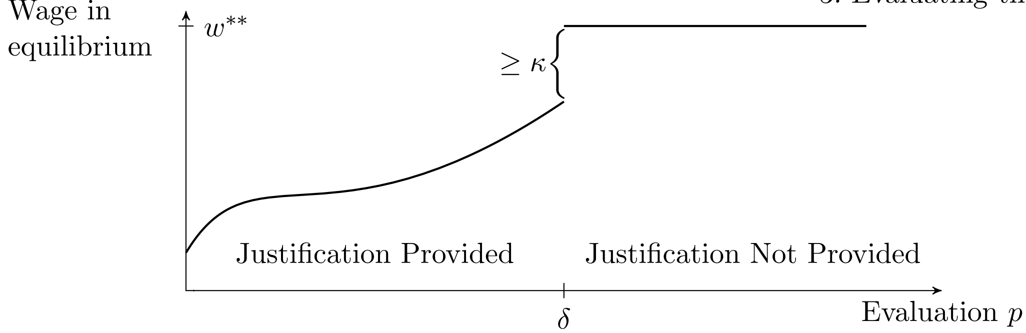
I provide a feasible contract that implements the equilibrium utilities of Program C. Note that Program C is a simplification of Program B for three reasons: Third-party payments can be avoided on the equilibrium path. Lemma 5 implies that justification follows a threshold rule. Constraint (1) in Program B specifies a constant wage for good evaluations. Moreover, Lemma 8 implies that any solution to Program C satisfies constraints (1) and (2) for $\kappa \leq \bar{\kappa}$. The next section discusses the main features of optimal contracts and strengthens the intuition.

3.4 Properties of Optimal Contracts

Now turn to the properties of optimal contracts determined in Proposition 2. Begin with the communication pattern.

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

The principal justifies only bad evaluations and low wages, she remains silent on good performance and the best wage. This is because the agent has reason to suspect a distortion only for bad evaluations. Remember that good evaluations imply higher wages and are, thus, more expensive for the principal. Therefore, the principal never gains by deviating to a better evaluation. Moreover, optimal contracts combine this communication pattern with rewarding evaluations at



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.

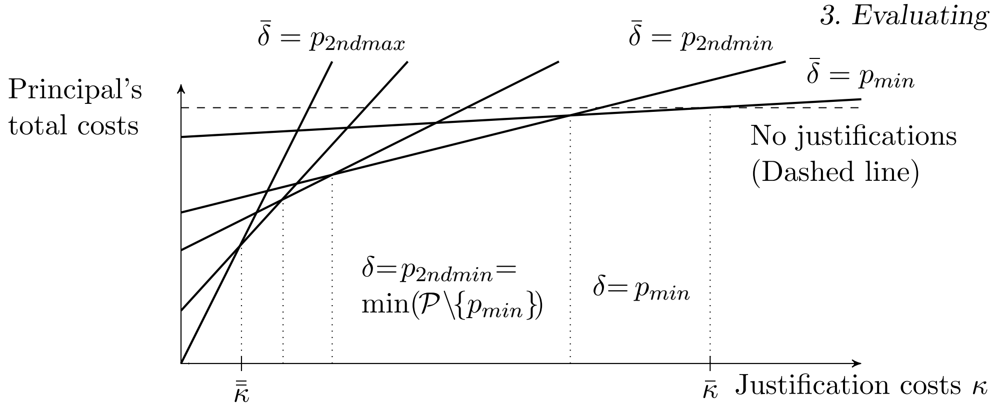
Figure 3: Equilibrium Wages and Communication

the top similarly. Thus, they eliminate wage differences that otherwise would call for justification. I describe this optimal communication pattern in the introduction and summarize it in Figure 3. Lemma 5 provides further intuition and, together with Proposition 2, also the formal arguments. If justification costs decrease, more evaluations are justified. Though recall from Lemma 3 that, unless justification is completely costless, it is never optimal to justify *all* evaluations.

Proposition 3. *Suppose $e > 0$. The justification threshold δ and the probability of providing justification decreases in justification costs κ. In particular, for sufficiently small costs, the principal justifies all, except the best evaluation. Hence, $\delta = p_{2ndmax} = \max(\mathcal{P} \setminus \{p_{max}\})$.*

The threshold δ depends on justification costs κ as depicted in Figure 4. The lower the justification costs, the higher the threshold. Hence, the ex-ante probability that the principal provides justification increases. Intuitively, the higher the justification costs, the more it becomes worthwhile to distort the wage payments by pooling more wages at the top to use fewer justifications. The proposition allows for rich comparative statics of the communication pattern to be tested across and within organizations. In particular, there is usually no binary or bang-bang solution for the optimal threshold. The communication pattern could, in principle, be found in time-use data or the length of written evaluations. Alternatively, one could explore differences in justification costs. These differences might arise from principals that are non-native speakers in multinational corporations, cultural distance following a takeover or a merger, or actual distance between principals and agents in companies with several locations.

Next, return to the empirical observations reviewed in the introduction. A large literature discusses distortions in wages set by subjective evaluations. The most prominent findings are leniency and centrality. Leniency means that agents receive the highest wages more often than the best performance occurs. Centrality means that variation in performance exceeds variation in wages, in particular, at the top. 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



The principal's total costs consist of wage payments and justification costs. The lines depict the principal's total costs for a given justification threshold $\bar{\delta}$. The pointwise minimum determines the optimal justification threshold δ. The optimal threshold is $\delta = p_{2ndmax}$ for low justification costs $\kappa \leq \bar{\bar{\kappa}}$ and the principal justifies all, except the best evaluations. With higher justification costs, the threshold decreases and the principal justifies less evaluations.

Figure 4: Optimal Justification Thresholds

raise just 3% higher than the median employee. Taylor and Wherry (1951, p. 39) were the first to find leniency and "a marked distortion . . . 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 wages set by subjective evaluations. This literature offers a plethora of approaches and data sources to test 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 wages. If the evaluation is 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 yield no straightforward explanation of this pattern, since those costs should apply to evaluations for all purposes similarly.

In terms of theoretical contributions to this literature, Suvorov and van de Ven (2009) study an intrinsically motivated agent who learns about his productivity from the principal's subjective evaluations. They show that leniency and centrality can be optimal to increase the probability that the agent continues to exert effort. Giebe and Gürtler (2012) show that leniency can be optimal if supervisors are altruistic. For this purpose, they scrutinize a three-tiered hierarchy. It can be optimal for the principal to allow the altruistic supervisor to be lenient, when evaluating the agent. Indirectly, also the papers I discuss in the related literature section under the headings of trust and fairness can be interpreted as explaining contractual features as the result of altruistic preferences. I do not assume altruistic preferences or intrinsic motivation, but limit myself to standard preferences. Nonetheless, my optimal contracts exhibit leniency and centrality.

Begin with centrality. Centrality means that, among top performers, variation in performance exceeds variation in wages: $\mathrm{Var}(p|p > \delta) > \mathrm{Var}(\text{equilibrium wages}|p > \delta)$. Alternatively, I can compare the variation of wages at the top in my model to the variation of wages for objective and verifiable performance measures as in Lemma 2. Both approaches yield the same result:

Corollary 2. *For $\kappa > \bar{\bar{\kappa}}$, wages set by subjective evaluations exhibit centrality.*

Hence, optimal contracting by a fully rational and unbiased principal results in centrality of wages and wage compression at the top. No biases or non-standard preferences are required to explain centrality in wages.

Now turn to leniency. Leniency means that agents receive the highest wages more often than the best performances occur: $\Pr(p = p_{max}) < \Pr(\max_{m_P, m_A} C(m_P, m_A) = \text{Equilibrium wages at } p)$. Alternatively, I can compare the probability of the highest wages in my model to the probability of the highest wages for objective and verifiable performance measures as in Lemma 2. Both approaches yield the same result:

Corollary 3. *For $\kappa > \bar{\bar{\kappa}}$, wages set by subjective evaluations exhibit leniency.*

The statement of Corollary 3 also remains valid if we extend the definition of leniency and compare the probabilities for the best two performances and the highest two wages, or for the best three performances and the highest three wages, and so on, for a sufficiently rich performance set $\mathcal{P}$. These results are in line with the aforementioned empirical observations of leniency and centrality: there is less variation in wages set by subjective evaluations than in the underlying performances, in particular at the top. Yet, this behavior is not the result of a bias, but implied by optimal contracting by a fully rational and unbiased principal, 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.

Finally, consider briefly leniency and centrality in evaluations directly. 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:

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

with a function

$$z(p) = \begin{cases} p & \text{if technology 1 is available} \\ \lambda(p) & \text{if technology 2 is available} \end{cases}$$

for any $p \in \mathcal{P}$ to ease the exposition and consider both justification technologies in one step. It is easy to check that the contract implements the same incentives and the same utilities for principal

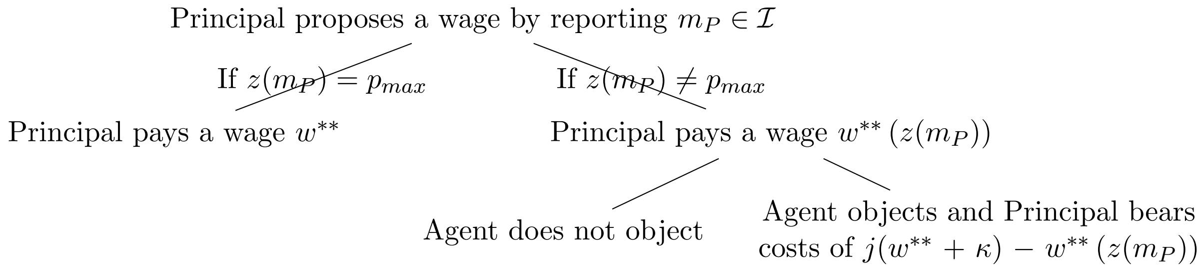


Figure 5: Indirect Implementation of Optimal Contracts

and agent as optimal contracts in Proposition 2. Indeed, the agent's wages are the same as in Proposition 2. Therefore, the agent chooses the same effort as before. The principal optimally reports $z(m_P) = p_{max}$ if the evaluation p is above the threshold δ. Otherwise, she justifies the true evaluation. The agent optimally replies with the message $m_A = 1$ to all messages of the principal that agree with her observation. Otherwise, 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. These costs can be interpreted as lawyers fees, mediation costs, or different kinds of conflicts. Note, however, that it is also possible to implement optimal contracts without such costs and with ex-post balance balance, as shown by Lang (2018). 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." Figure 5 sketches this implementation. These more realistic contracts imply leniency and centrality also in evaluations. For $\kappa > \bar{\bar{\kappa}}$, evaluations by a fully rational and unbiased principal exhibit centrality, $\mathrm{Var}(p|p > \delta) > \mathrm{Var}(z(m_P)|p > \delta)$, and leniency, $\Pr(p = p_{max}) < \Pr(z(m_P) = p_{max})$, also in terms of expectations $\mathbb{E}(z(m_P)) > \mathbb{E}(p)$.

Note, however, that the meaning of messages in messaging games occurs only in equilibrium. Therefore the optimal contract is not unique in terms of reporting strategies. My optimal contracts can account for leniency and centrality in evaluations, but it is also possible to write down an optimal contract without leniency and centrality in evaluations. Such a contract still exhibits leniency and centrality in wages as shown in above corollaries.

Next consider comparative statics. Sometimes, it may be empirically easier to measure differences in information than variation in justification costs. In these cases, costs are a random variable for the observer. Therefore assume that costs κ are drawn randomly and revealed to principal and agent in period -1.[8] Measuring information in moral hazard settings is not trivial. I consider two measures of information here: the worst performance becoming more likely; and more precise information for the principal. Begin with the first case. Formally, the worst

[8]I assume a continuous distribution on $(0, \infty)$ with positive density everywhere.

performance becoming more likely means adjusting the distribution $F(p|e)$ in the following way: For an $\epsilon \geq 0$, define $\bar{f}^H(p_{min}) = f^H(p_{min}) + \epsilon$ and $\bar{f}^L(p) = f^L(p)$ for all $p \in \mathcal{P}$, while reducing the probability measure $\bar{f}^H(p)$ for some better performances.[9] For sufficiently small ϵ, $\bar{f}^H(p)$ is a probability measure and the distribution $\bar{F}(p|e)$ satisfies the monotone likelihood ratio property.[10] Such a change in information yields more justifications and more variation in payments.

Corollary 4. *If the worst performance becomes more likely, i.e., ϵ increases, the probability increases that the principal uses contracts with justification. In these cases, variation in wages and payments by the principal increase.*

The intuition is that wages have to increase to compensate for the additional probability of receiving the lowest wage. The concavity of the agent's utilities and the changes in the probabilities implies that variation in wages has to increase to incentivize the agent to exert effort. More variation in wages makes providing justification more attractive. Therefore we observe more justifications.

The second measure of information is the precision of the principal's information in terms of a mean-preserving spread in the likelihood ratios of the agent's performance. Formally, this means adjusting the distribution $F(p|e)$ in the following way: Keep $\bar{f}^H(p) = f^H(p)$ for all $p < p_{2ndmax}$ and $\bar{f}^L(p) = f^L(p)$ for all $p \in \mathcal{P}$ unchanged and adjust $\bar{f}^H(p_{2ndmax}) = f^H(p_{2ndmax}) - \epsilon$ and $\bar{f}^H(p_{max}) = f^H(p_{max}) + \epsilon$ for an $\epsilon \geq 0$. This adjustment increases the variation in likelihood ratios improving the principal's information. For sufficiently small ϵ, $\bar{f}^H(p)$ is a valid probability measure and the distribution $\bar{F}(p|e)$ satisfies the monotone likelihood ratio property.[11] More precise information implies more justifications.

Corollary 5. *If the principal's information becomes more precise, i.e., ϵ increases, the probability increases that the principal provides justification.*

More precise information for the principal reduces her wage costs to incentivize the agent given enough justification. At the same time, the costs to incentivize the agent given little justification remain unchanged, because the principal cannot use her improved information without justifications. Therefore, the principal justifies more evaluations for more precise information.

Instead of varying the information, consider the case of the agent's work becoming more demanding. More demanding work means the disutility of exerting effort increases and there is an $\epsilon > 0$ so that $\bar{d}(e) = (1 + \epsilon)d(e)$ for all $e \in [0, 1)$. More demanding tasks yield more justifications and more variation in payments.

Corollary 6. *If the agent's work becomes more demanding, i.e., ϵ increases, the probability increases that the principal uses contracts with justification. In these cases, variation in wages and payments by the principal increase.*

[9] For example, $\bar{f}^H(p) = f^H(p) - \epsilon/(|\mathcal{P}| - 1)$ for all $p > p_{min}$.

[10] Note that the proof is more general than the setting defined here.

[11] Note that the proof is more general than the setting defined here.

The intuition is that the variation in wages has to increase to incentivize the agent for the more demanding work. More variation in wages makes providing justifications more attractive. Therefore we observe more justifications. Finally, consider the robustness of my results.

4 Robustness of the Results

With technology 2, 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 $m_A = -1$ if he did not learn any categories. Then message agree if $m_A \in m_P$ or $m_A = -1$. 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 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 2 still characterizes optimal contracts 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 my results 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 my analysis, so that all results remain valid.

So far, the agent is risk averse. Assume now that the agent is risk-neutral, i.e., $u(w) = w$ for all w, but has limited liability. To make the problem interesting, ensure that limited liability is binding by setting the reservation utility $\bar{u} = 0$. Finally, assume justification costs are not prohibitively expensive, i.e., $\kappa < \frac{d'(e)}{\sum_{\hat{p} \in \hat{P}} f^H(\hat{p}) - f^L(\hat{p})}$ with $\hat{P} = \{p \in \mathcal{P} | f^H(p) - f^L(p) > 0\}$. Otherwise, the model is the same as in Section 3.2 with technology 1. Then, my results are valid with two exceptions. Lemma 2 will only find weakly increasing wages and Proposition 3 is no longer valid. In particular, the optimal contract is

$$C(m_P, m_A) = \begin{cases} 0 & \text{if } m_P \notin \tilde{P} \\ \frac{d'(e)}{\sum_{\tilde{p} \in \tilde{P}} f^H(\tilde{p}) - f^L(\tilde{p})} & \text{if } p \in \tilde{P} \end{cases}$$

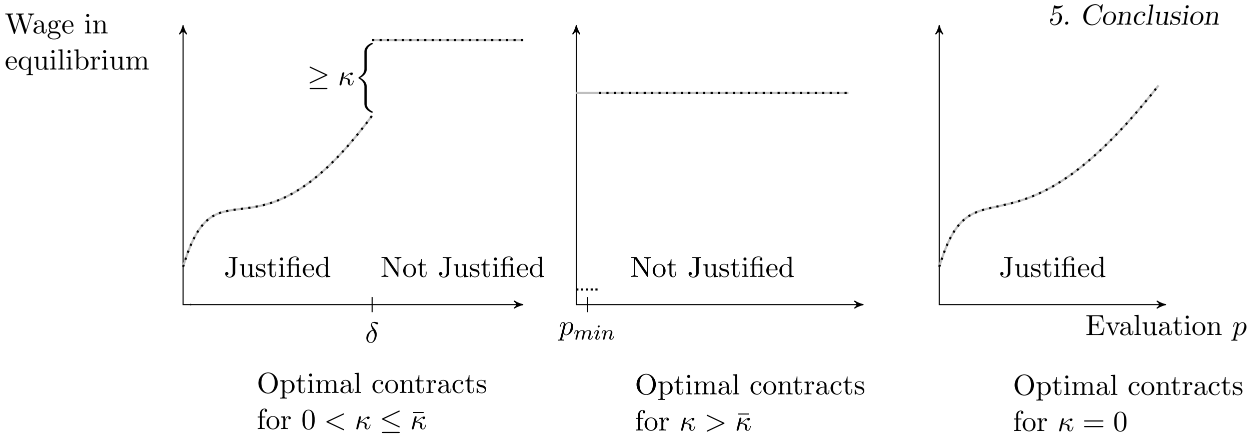
$$W(m_P, m_A) = \begin{cases} 0 & \text{if } m_P \notin \tilde{P} \text{ and } m_A = 1 \\ \frac{d'(e)}{\sum_{\tilde{p} \in \tilde{P}} f^H(\tilde{p}) - f^L(\tilde{p})} & \text{else} \end{cases}$$

for all $m_P \in \mathcal{P}$ and all $m_A \in \{0, 1\}$ with some set $\tilde{P} \subseteq \hat{P}$ and $\beta(p) = 0$ if and only if $p \in \tilde{P}$. The justification pattern is the same as with risk aversion. Evaluations are justified if and only if they are bad. Hence, my results with respect to the justification pattern are robust. Yet, the comparative statics of the communication pattern are more exciting with risk aversion than with risk neutrality, because the justification threshold δ varies more in the case of risk aversion.

5 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 prohibitively costly. This justification allows 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 6. 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 wages 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



Dotted lines denote the agent's wage in equilibrium, while gray lines depict the principal's payments in equilibrium. For ease of exposition, I draw wages as continuous, although they are discrete.

Figure 6: Comparison of optimal contracts

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.

Figure 6 allows relating to the literature. The most interesting case is for moderate justification costs. Then optimal contracts and the equilibrium wage pattern are new results and contribute to the literature. In the extreme cases of no justification costs and prohibitively large justification costs, equilibrium wages are similar to the previous literature, like Holmström (1979), Grossman and Hart (1983), MacLeod (2003) or Fuchs (2007), although optimal contracts differ due to my justifications. The middle graph depicts optimal equilibrium wages for prohibitively high justification costs. Unsurprisingly, the principal provides no justifications in this case and her payments are constant in the evaluation. The graph on the right-hand side depicts optimal equilibrium wages for zero justification costs. Unsurprisingly, the principal then provides justifications for all evaluations and the equilibrium wages and payments increase in the evaluation.

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 large 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

Proof of Lemma 1: 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. □

Lemma 2 characterizes the optimal contract if the principal's information is contractible. This contract is unique in terms of equilibrium utilities.

Proof of Lemma 2: Suppose justification technology 1 is available. Contractible information means $m_P = \lambda(I^1) = p$. 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 $p \in \mathcal{P}$. Then Program A_1 below determines how to optimally implement effort e by choosing wages $w_e^*(\cdot)$. The objective is to minimize expected costs subject to two 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.

$$\min_{w_e^*(\cdot)} \sum\nolimits_{p\in\mathcal{P}} w_e^*(p) f(p|e) \tag{A_1}$$

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

$$e \in \arg\max \sum\nolimits_{p\in\mathcal{P}} u(w_e^*(p)) f(p|e) - d(e) \tag{IC}$$

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

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

Note that the constraint set is nonempty. Take for example the contract

$$w(p) = \begin{cases} w_1 & \text{if } f^H(p) - f^L(p) \geq 0 \\ w_2 & \text{else} \end{cases}$$

with w_1 and w_2 determined below. Denote by $\mathcal{P}_c = \{p \in \mathcal{P} | f^H(p) - f^L(p) \geq 0\}$. The contract satisfies the agent's incentive compatibility IC if

$$\begin{aligned} d'(e) &= \sum\nolimits_{p\in\mathcal{P}_c} u(w_1)(f^H(p) - f^L(p)) + \sum\nolimits_{p\in(\mathcal{P}\setminus\mathcal{P}_c)} u(w_2)(f^H(p) - f^L(p)) = \\ &= (u(w_1) - u(w_2)) \sum\nolimits_{p\in\mathcal{P}_c} (f^H(p) - f^L(p)), \end{aligned}$$

because $\sum_{p\in\mathcal{P}} f^H(p) - f^L(p) = 0$ and, hence, $\sum_{p\in\mathcal{P}_c} f^H(p) - f^L(p) = -\sum_{p\in(\mathcal{P}\setminus\mathcal{P}_c)} f^H(p) - f^L(p)$. By definition of $\mathcal{P}_c$ and the monotone likelihood ratio property, the last sum is positive. Therefore the equation uniquely determines $u(w_1) - u(w_2)$. The contract satisfies the participation constraint PC if

$$\begin{aligned} d(e) + \bar{u} &= u(w_1) \sum\nolimits_{p\in\mathcal{P}_c} f(p|e) + u(w_2) \sum\nolimits_{p\in(\mathcal{P}\setminus\mathcal{P}_c)} f(p|e) = \\ &= u(w_2) + (u(w_1) - u(w_2)) \sum\nolimits_{p\in\mathcal{P}_c} f(p|e). \end{aligned}$$

because $\sum_{p\in\mathcal{P}} f(p|e) = 1$ and, hence, $\sum_{p\in\mathcal{P}_c} f(p|e) = 1 - \sum_{p\in(\mathcal{P}\setminus\mathcal{P}_c)} f(p|e)$. Plugging in above solution for $u(w_1) - u(w_2)$ uniquely determines w_2. Therefore, the constraint set of Program A_1 is nonempty. Moreover, the costs of an optimal contract are lower than $\max\{w_1, w_2\} < \infty$. According to Grossman and Hart (1983), the strict concavity of $u(\cdot)$ ensures that the solution to Program A_1 is unique. Denote the solution to Program A_1 by $w_e^*(p)$.

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

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

$$\frac{1}{u'(w_e^*(p))} = \nu_1 + \nu_2 \frac{f^H(p) - f^L(p)}{f(p|e)} = \nu_1 + \nu_2 \frac{\frac{f^H(p)}{f^L(p)} - 1}{e\frac{f^H(p)}{f^L(p)} + 1 - e} \tag{4}$$

The Lagrange multiplier ν_2 is positive for the following reason: If $\nu_2 = 0$, then Eq. (4) implies that $w(p)$ is constant in p violating 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 Eq. (4) 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 $p \in \mathcal{P}$.

Now suppose justification technology 2 is available instead. If signals I^2 and S are observable and contractible, the contract directly conditions on I^2 and S and messages do not matter. Similarly to above, Program A_2 below determines how to optimally implement effort e by choosing wages $w(I^2, S)$. The constraints are similar to the ones above in Program A_1. Denote by $P(I^2, S|e)$ the joint distribution of the principal's information I^2 and the agent's sample S conditional on the agent's effort e.

$$\inf_{w(\cdot,\cdot)} \int w(I^2, S)\mathrm{d}P(I^2, S|e) \tag{A_2}$$

$$\text{subject to} \int u(w(I^2, S))\mathrm{d}P(I^2, S|e) - d(e) \geq \bar{u} \tag{PC}$$

$$e \in \arg\max_e \left(\int u(w(I^2, S))\mathrm{d}P(I^2, S|e) - d(e) \right) \tag{IC}$$

Holmström (1979) shows that optimal wages only condition on aggregate evaluation $\lambda(I^2)$, because the average of the principal's information I^2 is a sufficient statistics for the agent's effort, $\Pr(I^2, S|e, \lambda(I^2)) = \Pr(I^2, S|\lambda(I^2)) = \Pr(I^2, S|p)$. Remember that $\lambda(I^2) = p$ almost surely. Therefore, only $\lambda(I^2)$ has to be observable and contractible. In particular, the detailed signals I^2 and S do not matter and Lemma 2 is also valid if S is private information and is not contractible. Therefore Program A_2 simplifies to Program A_1. To see this equivalence, define $\hat{w}(p) = \int u(w(I^2, S))\mathrm{d}P(I^2, S|p)$ and plug it into Program A_2. This step directly yields Program A_1, as

$$\int u(w(I^2, S))\mathrm{d}P(I^2, S|e) = \int \hat{w}(p)\mathrm{d}F(p|e) = \sum\nolimits_{p \in \mathcal{P}} \hat{w}(p) f(p|e).$$

The remainder of this proof is analogous to the case of technology 1 and therefore omitted. □

Proof of Proposition 1: The easiest way to study the principal's 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. Any feasible contract has to satisfy the following three conditions. 1) The agent accepts the contract if

$$\sum\nolimits_{p \in \mathcal{P}} u\big(c(p)\big) f(p|e) - d(e) \geq \bar{u}. \tag{5}$$

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

$$w(p) \geq c(p) \qquad\qquad \forall p \in \mathcal{P}. \tag{6}$$

3) The agent implements effort e if

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

By the same arguments as in Lemma 2, the first-order approach is valid here and focusing on the equilibrium utilities $c(p)$ is without loss of generality. Therefore, the agent implements effort e if

$$\sum\nolimits_{p \in \mathcal{P}} u\big(c(p)\big)(f^H(p) - f^L(p)) \geq d'(e). \tag{7}$$

Ensuring these equilibrium utilities might require additional constraints which I neglect for the moment. In particular, these additional constraints cannot make the contract cheaper for the principal. Therefore, the principal has to pay at least $\min_{w(\cdot)} \sum_{p \in \mathcal{P}} w(p) f(p|e)$ subject to the constraints (5), (7), (6). It is easy to see that in any solution of this program $w(p) = c(p)$ for all $p \in \mathcal{P}$. Hence, the program is equivalent to program A_1 in Lemma 2. Consequently, a lower bound for the principal's expected costs is given by $\sum_{p \in \mathcal{P}} w_e^*(p) f(p|e)$ with $w_e^*(p)$ defined in Lemma 2.

Suppose justification technology 1 is available. 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 induces the agent to report $m_A = \beta$ as the agent's wage is independent of her message m_A. Hence, we can interpret the agent's message as what the agent has observed regarding β. 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 2 ensures that these payments are bigger than $w_e^*(p)$, as $w_e^*(\cdot)$ is increasing. Hence, the contract induces the principal to report $m_P = p$ and $\beta(p) = 1$ for all $p \in \mathcal{P}$. It is easy to see that the principal's payments are always higher than the agent's wage. The contract proposed above implements the equilibrium wage $w_e^*(p)$. More importantly, the contract is optimal, because it is feasible and attains the lower bound on the principal's costs. Since it is impossible to incentivize the agent by a cheaper contract, this contract is optimal for $\kappa = 0$.

Now suppose justification technology 2 is available instead. Consider the contract

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

with an appropriately chosen $j \in \mathbb{N}$ that I determine below. The contract induces the agent to report $m_A = S$ as the agent's wage is independent of her message m_A. In addition, the contract induces the principal to report $m_P = I$ 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 2. 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 $j w_e^*(p_{max})$ 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 expected costs of such a deviation is

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

compared to costs of $w_e^*(\lambda(I))$ of reporting $m_P = I$. Such a deviation increases the principal's costs if

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

The last implication follows from $w_e^*(p_{max}) \geq w_e^*(\lambda(I))$ as established in Lemma 2. 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. Note 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 implements the equilibrium wage $w_e^*(\lambda(I))$ and is optimal, because it is feasible and attains the lower bound on the principal's costs for $\kappa = 0$. □

Proof of Lemma 3: Suppose to the contrary that the principal justifies all evaluations: $\beta(p) = 1$ for all $p \in \mathcal{P}$. Then, following the arguments in Proposition 1, the principal's expected costs are at least $\kappa + \mathbb{E}(w_e^*(p))$ with $w_e^*(\cdot)$ defined in Lemma 2. Yet, I am going to show that partial justification implements effort e even cheaper.

Suppose justification technology 1 is available. 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. It is easy to check that the principal's payments are always higher than the agent's wage. 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. As in Proposition 1, the contract induces the agent to report $m_A = \beta$ and the principal to report $\beta = \beta(p)$ and $m_P = p$. Hence, the agent's equilibrium wage equals $w_e^*(m_P)$. Therefore, by the definition of $w_e^*(\cdot)$, 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.

Now suppose justification technology 2 is available instead. 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 $j \in \mathbb{N}$. It is easy to check that the principal's payments are always higher than the agent's wage. 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. Similarly to Proposition 1, it is easy to show that the contract induces the agent to report $m_A = S$ and the principal to report $\beta = \beta(p)$ and $m_P = \begin{cases} [0, p_{max}] & \text{if } p = p_{max} \\ I & \text{else.} \end{cases}$. Hence, the agent's equilibrium wage equals $w_e^*(m_P)$. Analogously to the case of technology 1, this contract implements effort e of the agent and is cheaper than $\kappa + \mathbb{E}(w_e^*(p))$ with the same savings for the principal. This proves that the principal should not justify all evaluations for positive costs κ. □

Proof of Lemma 4: As in Proposition 1, we study the principal's program in terms of equilibrium utilities. Remember that expected payments for the principal given equilibrium strategies and evaluation p are denoted by $w(p)$. Similarly, the corresponding certainty equivalents of the agent's wages are $c(p)$. 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 $\beta(p)$. Now turn to the additional constraints which I neglected in Proposition 1.

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 principal 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, if available, and e, i.e., probabilities of p_2 conditionally on S and e are strictly positive. Hence, the agent does not detect 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 principal 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, if available, and e, i.e., probabilities of p_1 conditionally on S and e are strictly positive. Hence, the agent does not detect the principal's deviation from her message and follows his equilibrium strategy in reporting message m_A that cannot depend on p, because the agent's information is stochastically independent from p. 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

$$\bar{w} \geq w(p) + \kappa\beta(p) \geq (1 - \beta(p))\bar{w} \qquad \forall p \in \mathcal{P} \tag{8}$$

for a constant $\bar{w}$. 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 principal 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 $\beta(p_H) = 1$ and technology 1, she also uses the reporting strategy as if p_H had realized. For $\beta(p_H) = 1$ and technology 2, 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, if available, and e, i.e., probabilities of p_H and I' conditionally on S and e are strictly positive. Hence, the agent does not detect 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,

$$w(p) + \kappa\beta(p) \text{ is non-decreasing in } p. \tag{9}$$

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

$$\min_{w(p),\bar{w},c(p),\beta(p)} \sum\nolimits_{p\in\mathcal{P}} (w(p) + \kappa\beta(p))f(p|e)$$

subject to constraints (5), (6), (7), (8), and (9). This program equals Program B. □

Proof of Lemma 5: In order to show that only bad evaluations are justified, assume to the contrary that there is a solution $\mathcal{W}^{**}$ to Program B with $p_L, p_H \in \mathcal{P}$, such that $p_L < p_H$, $\beta^{**}(p_L) = 0$ and $\beta^{**}(p_H) = 1$. Constraints (1) and (2) imply that $w(p_H) + \kappa = w(p_L)$.

The next step constructs a $\mathcal{W}'$ that implements effort e cheaper than $\mathcal{W}^{**}$. Modify the solution $\mathcal{W}^{**}$ in the following way to get $\mathcal{W}'$: Set $\beta'(p_H) = 0$ 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 $\mathcal{W}'$ equals $\mathcal{W}^{**}$. The agent's certainty equivalent for an evaluation p_L decreases from $c(p_L)$ to $\tilde{c}$ in $\mathcal{W}'$, while the certainty equivalent for an evaluation p_H increases from $c(p_H)$ to $w(p_L)$, as $c(p_H) \le w(p_H) < w(p_L)$. The principal's expected costs remain unchanged. In $\mathcal{W}'$, 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 $\tilde{c} < c(p_L) \le w(p_L)$ and that the change in the agent's expected utilities ($\mathbb{E}[u(\mathcal{W}') - u(\mathcal{W}^{**})]$ in sloppy notation) equals

$$\begin{aligned}&\sum\nolimits_{i\in\{L,H\}} (u(c'(p_i)) - u(c(p_i)))f(p_i|e) =\\ =&(u(\tilde{c}) - u(c(p_L)))f(p_L|e) + (u(w(p_L)) - u(c(p_H)))f(p_H|e) = 0.\end{aligned} \tag{10}$$

Hence, $\mathcal{W}'$ 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. (10). 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 $p\in\mathcal{P}$, $\mathcal{W}'$ satisfies constraint (2). In addition, $\mathcal{W}'$ also satisfies constraints (1) and (3).

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. This procedure transforms solution $\mathcal{W}^{**}$ into $\mathcal{W}^m$ with the communication choice $\beta^m(p)$. Finally, I show that $\mathcal{W}^m$ can be made cheaper for the principal. There are two cases to consider: 1.) $\beta^m(p_{min})=1$; 2.) $\beta^m(p_{min})=0$. Begin with the first case and $\beta^m(p_{min})=1$. Above steps ensure that there is a $p_3\in\mathcal{P}$ with $c^m(p_3)<w^m(p_3)$ and $\beta^m(p_3)=0$. Hence, $p_3\neq p_{min}$. Increase $c^m(p_3)$ by a small $\epsilon>0$ to $c^m(p_3)+\epsilon$ and reduce $w^m(p_{min})$ and $c^m(p_{min})$ to $\tilde{c}^\epsilon$ with $\tilde{c}^\epsilon$ determined by

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

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 (3) due to $c^m(p_3)+\epsilon\leq w^m(p_3)$. It is easy to see that the modified contract satisfies constraints (1) and (2), because reducing $w^m(p_{min})$ never violates constraint (2). Moreover, $\beta^m(p_{min})=1$ and reducing $w^m(p_{min})$ does not violate constraint (1). The modified $\mathcal{W}^m$ implies lower expected costs for the principal compared to contract $\mathcal{W}^{**}$. The modified $\mathcal{W}^m$ satisfies all the constraints of Program B. Therefore, the modified $\mathcal{W}^m$ contradicts optimality of the solution $\mathcal{W}^{**}$.

Now turn to the second case with $\beta^m(p_{min})=0$. Note that $\beta^m(p_{min})=0$ implies $\beta^m(p)=0$ for all $p\in\mathcal{P}$. 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 $c^m(p_3)<w^m(p_3)$. Increase $c^m(p_3)$ by a small $\epsilon>0$ to $c^m(p_3)+\epsilon$ and reduce the remaining $c^m(p)$ and all $w^m(p)$ by $\hat{\epsilon}^\epsilon$ with $\hat{\epsilon}^\epsilon$ determined by

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

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 (3) due to $c^m(p_3)+\epsilon\leq w^m(p_3)-\hat{\epsilon}^\epsilon$. It is easy to see that the modified $\mathcal{W}^m$

satisfies constraints (1) and (2), because reducing all $w^m(p)$ equally never violates constraints (1) and (2). The modified $\mathcal{W}^m$ implies lower expected costs for the principal compared to solution $\mathcal{W}^{**}$. The modified $\mathcal{W}^m$ satisfies all the constraints of Program B. Therefore, the modified $\mathcal{W}^m$ contradicts optimality of the solution $\mathcal{W}^{**}$. Consequently, justification optimally follows a threshold rule and there is a $\delta \in [0,1)$ with $\beta(p) = 1$ if and only if $p \leq \delta$. □

Proof of Lemma 6: Assume $\beta(p) = 0$ for all $p \in \mathcal{P}$. Then, constraint (1) in Program B of Lemma 4 implies $w(p) = \bar{w}$ for all $p \in \mathcal{P}$. Hence, Program B simplifies to

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

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

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

$$f(p|e)(\bar{w} - c(p)) \geq 0 \qquad \forall p \in \mathcal{P}. \tag{11}$$

Define ν_1, ν_2 and $\chi(p)$ as the Lagrange multipliers of the conditions PC, IC and (11) 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

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

$$\nu_1 + \nu_2 \frac{f^H(p^*) - f^L(p^*)}{f(p^*|e)} = 0. \tag{12}$$

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 above program 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 (12) 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 (12), $\nu_2 > 0$, $p_{min} < p^*$ 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 (11) 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

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

According to Lemma 4, these equilibrium utilities solve the principal's relaxed problem. Consider the contract $W(m_P, m_A) = w$ and

$$C(m_P, m_A) = \begin{cases} c(p_{min}) & \text{if } m_P = p_{min} \text{ or } m_P = [0, p_{min}] \\ w & \text{else} \end{cases}$$

for all $m_P \in \mathcal{M}_P$ and all $m_A \in \mathcal{M}_A$. As the principal's payments are independent of her message, she provides no justification and reporting $m_P = p$ ($m_P = [0, p]$, for technology 2) is optimal for her. Hence, the contract implies the required equilibrium utilities. It remains to verify that this contract is feasible. By definition of w and $c(p_{min})$ as well as the principal's reporting strategy, the agent accepts the contract and chooses effort e. Finally, it is straightforward to check that the principal's payments $W(m_P, m_A)$ are always at least as high as the agent's wage $C(m_P, m_A)$. Consequently, if $e > 0$ and $\beta(p) = 0$ for all $p \in \mathcal{P}$, the contract is optimal for the principal, because it is feasible and attains the lower bound of Program B according to Lemma 4. $\square$

Proof of Lemma 7: 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 lemma. Second, I show that for lower costs κ and $e > 0$ there is a feasible contract 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 Lemma 6. Suppose $\kappa > \bar{\kappa}$. 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 $\beta'(\tilde{p}) = 1$. Again denote the principal's expected payments in contract $\mathcal{W}'$ given equilibrium strategies and evaluation p by $w'(p)$ and the agent's certainty equivalents of the wages given equilibrium strategies and evaluation p by $c'(p)$. Lemma 5 shows that $\beta'(\tilde{p}) = 1$ implies $\beta'(p_{min}) = 1$. If $p = p_{min}$, the principal's costs are at least $\kappa + c'(p_{min})$. Lemma 4 proves that $w'(p) + \kappa\beta'(p)$ is non-decreasing in p. Therefore, the principal's costs are at least $\kappa + c'(p_{min})$ for any $p \in \mathcal{P}$. Hence, the costs of contract $\mathcal{W}'$

are at least $\kappa + c'(p_{min}) > \bar{\kappa} + c'(p_{min}) = w - c(p_{min}) + c'(p_{min})$, because $\kappa > \bar{\kappa} = w - c(p_{min})$ with $c(p_{min})$ determined in Lemma 6. If $c(p_{min}) \leq c'(p_{min})$, these costs are above w. Then contract $\mathcal{W}'$ is more expensive than contract $\mathcal{W}$ in Lemma 6. Consequently, restrict attention to $c(p_{min}) > c'(p_{min})$. In contract $\mathcal{W}$, the agent's participation constraint PC was binding. Therefore, the agent's participation constraint PC requires for contract $\mathcal{W}'$:

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

Rearranging yields

$$u(\mathbb{E}(c'(p)|p \neq p_{min})) \geq u(w) + \frac{f(p_{min}|e)}{1 - f(p_{min}|e)}(u(c(p_{min})) - u(c'(p_{min}))) > u(w). \tag{13}$$

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

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

The ordering $c'(p_{min}) < c(p_{min}) < w < \mathbb{E}(c'(p)|p \neq p_{min})$ together with the concavity of u implies

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

This inequality contradicts the first inequality in (13). Hence,

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

$$\Rightarrow (1 - f(p_{min}|e))\mathbb{E}(c'(p)|p \neq p_{min}) + f(p_{min}|e)c'(p_{min}) > (1 - f(p_{min}|e))w + f(p_{min}|e)c(p_{min})$$
$$\Rightarrow (1 - f(p_{min}|e))\mathbb{E}(c'(p)|p \neq p_{min}) + f(p_{min}|e)(c'(p_{min}) + \kappa) >$$
$$> (1 - f(p_{min}|e))w + f(p_{min}|e)(c(p_{min}) + w - c(p_{min})) = w,$$

because $\kappa > \bar{\kappa} = w - c(p_{min})$. Therefore, contract $\mathcal{W}'$ is more expensive than contract $\mathcal{W}$ in Lemma 6. This contradiction proves that $\beta(p) = 0$ for all $p \in \mathcal{P}$ is optimal for $\kappa > \bar{\kappa}$.

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

for technology 1:

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

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

for technology 2:

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

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

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 all $p > p_{min}$. 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 feasible.

As the agent's wage in contract $\mathcal{W}'$ is independent of his message, he optimally reports $m_A = \beta$ for technology 1 and $m_A = S$ for technology 2. It is easy to verify that $W'(m_P, m_A) \geq C'(m_P, m_A)$ for any combination of messages. Next, consider the principal's reporting strategy. Begin with technology 1: If $p = p_{min}$, reporting $m_P = p_{min}$ implies costs of $c(p_{min}) + \kappa$ for the principal. Any deviation increases her costs to at least w. If $p > p_{min}$, reporting $m_P = p$ implies costs of w. Deviating to $\beta = 1$ and a message with $m_P \geq p$ increases her costs to $w + \kappa$. Deviating to $\beta = 1$ and a message with $m_P < p$ is impossible. Deviating to $\beta = 0$ and $m_P \neq p$ weakly increases the principal's costs to w. Next turn to technology 2: deviating to a message with $\lambda(m_P) > p$ weakly increases her costs to w or $w + \kappa$. Deviating to a message $\lambda(m_P) < p$ increases the principal's costs for the same reasons as in the proof of Proposition 1 for sufficiently large j. Deviating to a message with $\lambda(m_P) = p$ weakly increases her costs to w or $w + \kappa$, if $p > p_{min}$. Deviating to a message with $\lambda(m_P) = p$ weakly increases her costs for sufficiently large j, if $p = p_{min}$. Therefore, the considered reporting strategy is optimal for the principal. Finally, contract $\mathcal{W}'$ satisfies the agent's participation constraint PC and incentive compatibility IC by the definition of w and $c(p_{min})$. 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 Lemma 8: First, note that the monotone likelihood ratio property ensures

$$\frac{f^H(\delta)}{f^L(\delta)} = \frac{f^H(\delta)}{f^L(\delta)} \frac{\sum_{p'>\delta} f^L(p')}{\sum_{p'>\delta} f^L(p')} < \frac{\sum_{p'>\delta} \frac{f^H(p')}{f^L(p')} f^L(p')}{\sum_{p'>\delta} f^L(p')} = \frac{\sum_{p'>\delta} f^H(p')}{\sum_{p'>\delta} f^L(p')}.$$

Thus, for a given δ merging the performances above δ preserves the monotone likelihood ratio property. Hence, according to Grossman and Hart (1983), the strict concavity of $u(\cdot)$ ensures that for a given δ the solution to Program C is unique. Denote the Lagrange multiplier for the participation constraint by ν_1 and the one for the agent's incentive compatibility by ν_2. Then 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

$$\nu_1 + \nu_2 \frac{f^H(p) - f^L(p)}{f(p|e)} = \frac{1}{u'(w(p))} \quad \text{for all } p \leq \delta \tag{14}$$

$$\nu_1 + \nu_2 \frac{\sum_{p>\delta} (f^H(p) - f^L(p))}{1 - F(\delta|e)} = \frac{1}{u'(w)} \tag{15}$$

The monotone likelihood ratio property guarantees that

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

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 (14) implies that $w(p)$ increases in $p \leq \delta$. In addition, equations (14) and (15) as well as the monotone likelihood ratio property imply $w > w(p)$ for $p \leq \delta$. It remains to verify $w \geq w(\delta) + \kappa$. 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 $w < w(\delta) + \kappa$. This implies $w(\delta) \in (w - \kappa, w)$ by the preceding steps. Begin with $\delta = p_{min}$. I have already established that ν_2 is positive. Additionally, by standard arguments, ν_1 is positive. Therefore any solution to Program C has to satisfy both constraints PC and IC with equality. Yet, for $\delta = p_{min}$, the unique solution to constraints PC and IC holding with equality is $w(p_{min}) = C(p_{\min}, 0)$ and $w = C(p_{\max}, 0)$ with $C(\cdot,\cdot)$ from Lemma 6. By definition of $\bar{\kappa} = w - w(p_{min})$ and $\kappa \leq \bar{\kappa}$, $w(\delta) + \kappa \leq w(\delta) + \bar{\kappa} = w(p_{min}) + \bar{\kappa} = w$ contradicting $w < w(\delta) + \kappa$. Consequently, $w < w(\delta) + \kappa$ and $\delta = p_{min}$ is impossible for a solution of Program C.

If $\delta > p_{min}$, modify contract $\mathcal{W}$ by increasing $w(\delta)$ to w and 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

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

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 $w \geq w(\delta) + \kappa$. The first part of the proof shows that $w(p)$ increases in $p \leq \delta$. Consequently, any solution to Program C satisfies $w \geq w(p) + \kappa$ for all $p \leq \delta$. □

Proof of Proposition 2: In the first step, I characterize equilibrium behavior for the contract of Proposition 2. The second step shows that this contract is optimal.

Suppose justification technology 1 is available. 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 2. Suppose the principal uses the strategy $m_P = p$ and ($\beta(p) = 1$ if and only if $p \leq \delta$).

As the agent's wage does not depend on his message m_A, the strategy $m_A = \beta$ following the observed choice β of the principal 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 2, 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 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 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 $m_P = p$. Suppose the agent accepts the contract, exerts effort e and sends message $m_A = \beta$. If $p > \delta$, the principal's costs after reporting message m_P are

$$= \begin{cases} w^{**} & \text{if } \beta = 0 \\ w^{**} + \kappa & \text{if } \beta = 1 \end{cases}$$

If $p > \delta$, there is, hence, no profitable deviation for the principal. It remains to consider the case $p \leq \delta$. Then the principal's costs after reporting message m_P are

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

Lemma 8 ensures $w^{**} \geq w^{**}(p') + \kappa$ for all $p' \in \mathcal{P}$ and $p' \leq \delta$. Remember $m_P \geq p$ if $\beta = 1$. Therefore, there is no profitable deviation for the principal if $p \leq \delta$. Consequently, truthful reporting $m_P = p$ is optimal for the principal and the principal's strategy stated above is optimal.

Now turn to the second step. Lemma 5 proves that in any solution to Program B there is δ, so that $\beta(p) = 1$ if and only if $p \leq \delta$. Lemma 7 ensures that $\delta \geq p_{min}$ for $\kappa < \bar{\kappa}$ and $e > 0$. Lemma 8 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 2 is feasible and yields equilibrium utilities that solve Program C. Consequently, the contract in Proposition 2 is optimal.

Now suppose justification technology 2 is available instead. Return to the first step and consider the contract:

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

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

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. Suppose the principal uses the strategy:

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

$$m_P = \begin{cases} I & \text{if } p \in \mathcal{P} \text{ and } p \leq \delta \\ [0,p] & \text{else} \end{cases} \tag{16}$$

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 $e \in \arg\max \sum_{p\leq\delta} u\big(w^{**}(p)\big) f(p|e) + (1-F(\delta|e))u(w^{**}) - d(e)$, independently of his message. By the same arguments as in Lemma 2, 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 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 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 use the reporting strategy (16). Suppose the agent accepts the contract, exerts effort e and sends message $m_A = S$. If $p > \delta$, similarly to Proposition 1, the principal's costs after reporting message m_P are:

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

If $p > \delta$ and $j \geq \max_{q,q'\in\mathcal{P},q>q'} \frac{q}{q-q'}$, there is no profitable deviation for the principal, as in the proof of Proposition 1. It remains to consider the case $p \leq \delta$. Then the principal's costs after reporting message m_P are $w^{**}(p) + \kappa$ for $m_P = I$ and equal

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

Lemma 8 ensures $w^{**} \geq w^{**}(p') + \kappa$ for all $p' \in \mathcal{P}$ and $p' \leq \delta$. As in the proof of Proposition 1, there is no profitable deviation for the principal if $j \geq \max_{q,q'\in\mathcal{P},q>q'} \frac{q}{q-q'}$ and $p \leq \delta$. Consequently, the principal's reporting strategy stated in (16) is optimal. The second step is analogous to the case of technology 1 and therefore omitted. □

Proof of Corollary 1: Corollary 1 follows from Proposition 2 and Lemma 7. □

Proof of Proposition 3: Lemma 3 determines a contract $\mathcal{W}$ in the constraint set of Program B with $\delta = p_{2ndmax} = \max(\mathcal{P} \setminus \{p_{max}\})$. Lemma 3 shows that this contract yields lower costs than any contract with $\delta \geq p_{max}$. Given $\delta = p_{2ndmax}$, the definition of $w_e^*(\cdot)$ in Lemma 2 guarantees that contract $\mathcal{W}$ is optimal for sufficiently small costs κ. If $|\{p' \in \mathcal{P} | \beta(p') = 0\}| \geq 2$, Proposition 2 ensures that $\beta(p_{2ndmax}) = \beta(p_{max}) = 0$ in an optimal contract. According to

Proposition 2, the principal's equilibrium utilites are $w(p_{2ndmax}) = w(p_{max})$ in such a contract. Therefore, a feasible contract implies at least wage costs (excluding justification costs) of

$$\min_{w(\cdot)} \sum\nolimits_{p\in\mathcal{P}} w(p)f(p|e)$$

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

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

$$w(p_{2ndmax}) = w(p_{max}) \tag{17}$$

Lemma 2 proves that a solution to this program without the additional constraint (17) has $w(p_{2ndmax}) < w(p_{max})$. In addition, the solution to this program without the additional constraint (17) is unique according to Grossman and Hart (1983), because $u(\cdot)$ is strictly concave. The idea is to restate the program in terms of utilities. Then both constraints are linear and the objective function is strictly convex, since the inverse function $u^{-1}(\cdot)$ of the strictly concave and increasing utility function $u(\cdot)$ is strictly convex. Hence, the additional constraint is binding and the wage costs of a contract with $|\{p' \in \mathcal{P}|\beta(p') = 0\}| \geq 2$ are strictly higher than the wage costs of an optimal contract given $\delta = p_{2ndmax}$. The difference between these costs does not depend on κ. Thus, for sufficiently small κ, the additional justification costs κ in the contract given $\delta = p_{2ndmax}$ are lower than the reduction in wage costs. Therefore, $\delta = p_{2ndmax}$ is optimal for sufficiently small κ. Hence, there is a $\bar{\bar{\kappa}} > 0$, so that $\delta = p_{2ndmax}$ is optimal for $\kappa \leq \bar{\bar{\kappa}}$.

Without loss of generality, denote the elements of $\mathcal{P}$ in the following way: $\mathcal{P} = \{p_1, p_2, \ldots, p_{|\mathcal{P}|}\}$ with $p_i < p_{i+1}$ for $i = 1, 2, \ldots, |\mathcal{P}| - 1$. Then $p_{|\mathcal{P}|} = p_{max}$, $p_{|\mathcal{P}|-1} = p_{2ndmax}$ and $p_1 = p_{min}$. Consider for a given $\bar{\delta} \in \mathcal{P}$:

$$Y(\bar{\delta}) = \min_{w(\cdot)} \sum\nolimits_{p\in\mathcal{P}} w(p)f(p|e)$$

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

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

$$w(p') = w(p_{max}) \text{ for all } p' > \bar{\delta} \tag{18}$$

The proof of Lemma 8 shows that we can interpret all performances above $\bar{\delta}$ as one performance and the monotone likelihood ratio property is still satisfied. Then the same arguments as above ensure that the program has a unique solution. As in Lemma 2, the solution $w^{\bar{\delta}}(p)$ to above program is characterized by

$$\frac{1}{u'(w^{\bar{\delta}}(p))} = \nu_1 + \nu_2 \frac{f^H(p) - f^L(p)}{f(p|e)} \qquad \forall p \leq \bar{\delta}$$

Therefore, $w^{\bar{\delta}}(p)$ increases in p for all $p \leq \bar{\delta}$. Consequently, comparing $w^{p_i}(\cdot)$ to $w^{p_{i+1}}(\cdot)$ shows that constraint (18) is binding for $p' = \min\{p \in \mathcal{P}|p > \bar{\delta}\}$. As the solution $w^{\delta}(\cdot)$ is unique for each δ, this guarantees $Y(p_i) > Y(p_{i+1})$ for $i = 1, 2, \ldots, |\mathcal{P}| - 2$. Note that $Y(\bar{\delta})$ does not depend on justification costs κ and that $Y(p_{|\mathcal{P}|-1}) = Y(p_{|\mathcal{P}|})$. The principal's total costs are $Y(\bar{\delta}) + F(\bar{\delta}|e)\kappa$. Minimizing this sum with respect to $\bar{\delta} \in \mathcal{P}$, yields an optimal δ that is generically unique.[12]

[12]For any open set $K \subset [0, \bar{\kappa}]$ with a continuous distribution, the probability of $\kappa \in K$ with more than one optimal threshold δ is zero.

Figure 4 on page 16 depicts the principal's total costs depending on κ. Suppose to the contrary that the optimal δ increases in κ, i.e., there are $\kappa, \epsilon > 0$ and $\hat{\delta}, \tilde{\delta} \in \mathcal{P}$ such that $\hat{\delta} < \tilde{\delta}$ and

$$Y(\hat{\delta}) + F(\hat{\delta}|e)\kappa \leq Y(p') + F(p'|e)\kappa \qquad \forall p' \in \mathcal{P}$$
$$Y(\tilde{\delta}) + F(\tilde{\delta}|e)(\kappa+\epsilon) \leq Y(p'') + F(p''|e)(\kappa+\epsilon) \qquad \forall p'' \in \mathcal{P}$$

Set $p' = \tilde{\delta}$ and $p'' = \hat{\delta}$. Then adding both inequalities yields $F(\tilde{\delta}|e)\epsilon \leq F(\hat{\delta}|e)\epsilon$ contradicting $\hat{\delta} < \tilde{\delta}$. Therefore, no δ' above δ can be optimal if δ is optimal for costs κ and the justification costs increase.

Finally, denote by $Y(0) = u^{-1}\left(\bar{u} + d(e) + \frac{f(p_{min}|e)}{f^L(p_{min})-f^H(p_{min})}d'(e)\right)$ the costs of the optimal contract that does not justify any evaluations following Lemma 7. Remember that $Y(0) = Y(p_1) + F(p_1|e)\bar{\kappa}$. Analogously to Lemma 7, it is easy to show that $Y(0) < Y(p_i) + F(p_i|e)\kappa$ for all $\kappa \geq \bar{\kappa}$ and all $i \in \{2, \dots, |\mathcal{P}|\}$. The linearity in κ implies $\delta = p_1$ is optimal for $\kappa \in (\bar{\kappa}-\epsilon, \bar{\kappa}]$ with an $\epsilon > 0$. □

Proof of Corollary 2: For $\kappa > \bar{\bar{\kappa}}$ as defined in Proposition 3, Propositions 2 and 3 as well as Lemma 7 show that the optimal threshold $\delta < p_{2ndmax}$. Hence, wages for several evaluations are pooled as $|\{p \in \mathcal{P}|p > \delta\}| > 1$, because wages for evaluations above δ are constant. Yet, evaluations above δ have a positive variance. Similarly, wages for objective and verifiable performances have a positive variance given a performance above δ according to Lemma 2. Therefore, there is centrality and maximal wage compression at the top. □

Proof of Corollary 3: For $\kappa > \bar{\bar{\kappa}}$ as defined in Proposition 3, Propositions 2 and 3 as well as Lemma 7 show that the optimal threshold $\delta < p_{2ndmax}$. Hence, the principal pools several wages as $|\{p \in \mathcal{P}|p > \delta\}| > 1$. Denote the equilibrium wage in the optimal contract as

$$\hat{c}(p) = \begin{cases} w(p) & \text{if } p \leq \delta \text{ and } \kappa \leq \bar{\kappa} \\ w & \text{if } p > \delta \text{ and } \kappa \leq \bar{\kappa} \\ u^{-1}\left(\bar{u} + d(e) + \frac{f(p_{min}|e)d'(e)}{f^L(p_{min})-f^H(p_{min})}\right) & \text{if } p > p_{min} \text{ and } \kappa > \bar{\kappa} \\ u^{-1}\left(\bar{u} + d(e) - \frac{(1-f(p_{min}|e))d'(e)}{f^L(p_{min})-f^H(p_{min})}\right) & \text{if } p = p_{min} \text{ and } \kappa > \bar{\kappa} \end{cases}$$

for all $p \in \mathcal{P}$ with the values of $w(p)$ and w determined in Program C. Note that in the optimal contracts $\max_{m_P, m_A} C(m_P, m_A) = \hat{c}(p_{max})$. Therefore, optimal contracts imply leniency as

$$\Pr(\{p|w_e^*(p) = w_e^*(p_{max})\}) = \Pr(p = p_{max}) = f(p_{max}|e) < \sum_{p' \in \{p \in \mathcal{P}|p>\delta\}} f(p'|e) = \Pr(\{p|\hat{c}(p) = \hat{c}(p_{max})\}).$$

Thus, the principal pays the highest wage more often than observing the best performance or paying the highest wage in a setting with a verifiabel performance measure. □

Proof of Corollary 4: For a small $\epsilon \geq 0$ and $\alpha \in [0, (1-f^L(p_{min}))/(1-f^H(p_{min}))]$ define $\bar{f}^H(p_{min}) = f^H(p_{min}) + \epsilon$ and $\bar{f}^L(p_{min}) = f^L(p_{min}) + \alpha\epsilon$, while reducing the probability measures $\bar{f}^L(p)$ and $\bar{f}^H(p)$ for some better performances. For example, $\bar{f}^H(p) = f^H(p) - \epsilon/(|\mathcal{P}|-1)$ and $\bar{f}^L(p) = f^L(p) - \alpha\epsilon/(|\mathcal{P}|-1)$ for all $p > p_{min}$. Suppose ϵ is nonnegative, but sufficiently small, so that the probability measures $p^L(\cdot)$ and $p^H(\cdot)$ are well defined and the distribution $\bar{F}(p|e)$

satisfies the monotone likelihood ratio property. Note that

$$\frac{\bar f(p_{min}|e)}{\bar f^L(p_{min})-\bar f^H(p_{min})}=\frac{f(p_{min}|e)+\epsilon(e+\alpha(1-e))}{f^L(p_{min})-f^H(p_{min})-(1-\alpha)\epsilon}$$

Hence,

$$\partial\frac{\bar f(p_{min}|e)}{\bar f^L(p_{min})-\bar f^H(p_{min})}/\partial\epsilon=\frac{(e+\alpha(1-e))(\bar f^L(p_{min})-\bar f^H(p_{min}))+(1-\alpha)\bar f(p_{min}|e)}{(\bar f^L(p_{min})-\bar f^H(p_{min}))^2}=$$
$$=\frac{\bar f^L(p_{min})-\alpha\bar f^H(p_{min})}{(\bar f^L(p_{min})-\bar f^H(p_{min}))^2}>0,$$

because the monotone likelihood ratio property implies that $\bar f^L(p_{min})-\bar f^H(p_{min})>0$. Therefore, the principal's payments in the absence of justification increase in ϵ according to Lemma 6. At the same time, the agent's wage for good evaluations increases for $\delta\le p_{min}$ according to Proposition 2 as well as Lemmas 6 and 7. In addition,

$$\frac{1-\bar f(p_{min}|e)}{\bar f^L(p_{min})-\bar f^H(p_{min})}=\frac{1-f(p_{min}|e)-\epsilon(e+\alpha(1-e))}{f^L(p_{min})-f^H(p_{min})-(1-\alpha)\epsilon}$$

hence,

$$\partial\frac{1-\bar f(p_{min}|e)}{\bar f^L(p_{min})-\bar f^H(p_{min})}/\partial\epsilon=\frac{-(e+\alpha(1-e))(\bar f^L(p_{min})-\bar f^H(p_{min}))+(1-\alpha)(1-\bar f(p_{min}|e))}{(\bar f^L(p_{min})-\bar f^H(p_{min}))^2}=$$
$$=\frac{1-\alpha-\bar f^L(p_{min})+\alpha\bar f^H(p_{min})}{(\bar f^L(p_{min})-\bar f^H(p_{min}))^2}\ge 0,$$

because $\alpha\le(1-f^L(p_{min}))/(1-f^H(p_{min}))$. Therefore, the agent's wage for the worst evaluation decreases for $\delta\le p_{min}$ according to Proposition 2 as well as Lemmas 6 and 7. Hence, the threshold $\bar\kappa$ increases in ϵ, because the utility function $u(\cdot)$ is increasing, so that its inverse function is also increasing. Consequently, according to Propositions 2 and 3 as well as Lemma 7, there are some costs κ, in which for $\epsilon=0$ no justifications are used, but for small $\epsilon>0$ the principal offers a contract with justifications. Hence, due to the distribution of costs κ, the probability that the principal uses contracts with justification increases in ϵ. Above I also showed that such a change implies more variation in wages and payments by the principal. □

Proof of Corollary 5: Define $p_{3rdmax}=\max(\mathcal{P}\setminus\{p_{2ndmax},p_{max}\})$. Suppose $\bar f^H(p)=f^H(p)$ and $\bar f^L(p)=f^L(p)$ for all $p<p_{2ndmax}$ as well as

$$\bar f^H(p_{2ndmax})=f^H(p_{2ndmax})-\epsilon,\qquad \bar f^H(p_{max})=f^H(p_{max})+\epsilon,$$
$$\bar f^L(p_{2ndmax})=f^L(p_{2ndmax})-\rho,\qquad \bar f^L(p_{max})=f^L(p_{max})+\rho$$

for an $\epsilon\in[-\min\{f^H(p_{max},1-f^H(p_{2ndmax})\},\min\{f^H(p_{2ndmax}),1-f^H(p_{max}\}]$ and a $\rho\in[-\min\{f^L(p_{max},1-f^L(p_{2ndmax})\},\min\{f^L(p_{2ndmax}),1-f^L(p_{max}\}]$ with $\epsilon\ge\rho\frac{f^H(p_{max})}{f^L(p_{max})}$, $\epsilon\ge\rho\frac{f^H(p_{2ndmax})}{f^L(p_{2ndmax})}$, and $\epsilon<f^H(p_{2ndmax})-f^L(p_{2ndmax})\frac{f^H(p_{3rdmax})}{f^L(p_{3rdmax})}+\rho\frac{f^H(p_{3rdmax})}{f^L(p_{3rdmax})}$. First, note that the constraint set is not empty, because all constraints are satisfied for $\epsilon=\rho=0$ as the monotone likelihood ratio property ensures that $f^H(p_{2ndmax})-f^L(p_{2ndmax})\frac{f^H(p_{3rdmax})}{f^L(p_{3rdmax})}>0$. Indeed, any sufficiently small $\epsilon\ge 0$ satisfies the constraints for $\rho=0$. Second, $\frac{f^H(p_{max})}{f^L(p_{max})}<\frac{\bar f^H(p_{max})}{\bar f^L(p_{max})}$, because $\epsilon\ge\rho\frac{f^H(p_{max})}{f^L(p_{max})}$. Third, $\frac{f^H(p_{2ndmax})}{f^L(p_{2ndmax})}>\frac{\bar f^H(p_{2ndmax})}{\bar f^L(p_{2ndmax})}$, because $\epsilon\ge\rho\frac{f^H(p_{2ndmax})}{f^L(p_{2ndmax})}$.

Fourth, the second point and the observation that $\frac{f^H(p_{3rdmax})}{f^L(p_{3rdmax})} < \frac{\bar{f}^H(p_{2ndmax})}{\bar{f}^L(p_{2ndmax})}$ for $\epsilon < f^H(p_{2ndmax}) - f^L(p_{2ndmax})\frac{f^H(p_{3rdmax})}{f^L(p_{3rdmax})} + \rho\frac{f^H(p_{3rdmax})}{f^L(p_{3rdmax})}$ imply that the new distribution $\bar{F}(p|e)$ satisfies the monotone likelihood ratio property. Fifth, ϵ and ρ do not change the expectation of the likelihood ratios $\frac{\bar{f}^H(p)-\bar{f}^L(p)}{\bar{f}(p|e)}$. Hence, the second, third and fifth point imply that the likelihood ratio distribution of $\bar{F}(p|e)$ is a mean-preserving spread compared to the likelihood ratio distribution of $F(p|e)$.

Program C for a given δ satisfies the assumptions of Kim (1995, p. 90), as he writes "our results still hold in a discrete case as well." Kim (1995, Proposition 1) shows that the mean-preserving spread in the likelihood ratio distribution lowers the principal's wage costs. Therefore, following the proof of Proposition 3, $Y_{\bar{F}(p|e)}(p_{2ndmax}) < Y_{F(p|e)}(p_{2ndmax})$.

To compute $Y_{\bar{F}(p|e)}(p)$ for any $p < p_{2ndmax}$, consider the likelihood ratio if the two best performances are considered together:

$$\frac{\bar{f}^H(p_{max}) + \bar{f}^H(p_{2ndmax}) - \bar{f}^L(p_{max}) - \bar{f}^L(p_{2ndmax})}{\bar{f}(p_{max}|e) + \bar{f}(p_{2ndmax}|e)} =$$
$$= \frac{f^H(p_{max}) + f^H(p_{2ndmax}) - f^L(p_{max}) - f^L(p_{2ndmax})}{f(p_{max}|e) + f(p_{2ndmax}|e)},$$

as all ϵ and ρ terms cancel out. Hence, optimal contracts for $\delta \le p_{2ndmax}$ remain unchanged according to Proposition 2 as well as Lemmas 6 and 7. Hence, also the principal's total costs remain unchanged for a given $\delta \le p_{2ndmax}$. In the notation of Proposition 3, $Y_{\bar{F}(p|e)}(p) = Y_{F(p|e)}(p)$ for all $p < p_{2ndmax}$. Following the analysis of Proposition 3, there are justification costs for which the optimal δ increases to p_{2ndmax} and the principal justifies strictly more evaluations for the distribution $\bar{F}(p|e)$ than for the distribution $F(p|e)$. For such a κ, the probability of an evaluation with justification increases. Given the distribution of justification costs, the probability of providing justification increases for $\bar{F}(p|e)$ compared to $F(p|e)$. □

Proof of Corollary 6: Suppose there is an $\epsilon > 0$ such that the new disutility of exerting effort $\bar{d}(e) = (1+\epsilon)d(e)$ for all $e \in [0,1)$. Denote by $w_1 = \bar{u} + \bar{d}(e) + \frac{f(p_{min}|e)\bar{d}'(e)}{f^L(p_{min})-f^H(p_{min})}$, by $w_0 = \bar{u} + \bar{d}(e) - \bar{d}'(e)\frac{1-f(p_{min}|e)}{f^L(p_{min})-f^H(p_{min})}$ and by $(u^{-1})'[w]$ the derivative of the inverse function at w. Then

$$\frac{\partial \bar{\kappa}_{\bar{d}(e)}}{\partial \epsilon} = (u^{-1})'[w_1]\left(d(e) + \frac{f(p_{min}|e)d'(e)}{f^L(p_{min}) - f^H(p_{min})}\right) - (u^{-1})'[w_0]\left(d(e) - d'(e)\frac{1-f(p_{min}|e)}{f^L(p_{min}) - f^H(p_{min})}\right) >$$
$$> (u^{-1})'[w_0]\left(d(e) + \frac{f(p_{min}|e)d'(e)}{f^L(p_{min}) - f^H(p_{min})} - d(e) + d'(e)\frac{1-f(p_{min}|e)}{f^L(p_{min}) - f^H(p_{min})}\right) =$$
$$= (u^{-1})'[w_0]d'(e)\frac{1}{f^L(p_{min}) - f^H(p_{min})} > 0,$$

because $w_1 > w_0$ and the utility function $u(\cdot)$ is concave and increasing, so that its inverse function is increasing and convex. In addition, the monotone likelihood ratio property implies $f^L(p_{min}) - f^H(p_{min}) > 0$. For $\delta \le p_{min}$, Proposition 2 as well as Lemmas 6 and 7 imply that $\bar{\kappa}$ is the difference between the agent's wage for the worst evaluation and good evaluations. Therefore the variation in the agent's wages increases.

Moreover, according to Proposition 2 and 3 as well as Lemma 7, there are some costs κ, for which with $\epsilon = 0$ no justifications are used, but with small $\epsilon > 0$ the principal offers a contract with justifications. Hence, due to the distribution of costs κ, the probability that the principal uses contracts with justification increases in ϵ. Above I also showed that such a change implies more variation in wages and payments by the principal. □

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. (2015). Subjective Evaluations: Discretionary Bonuses and Feedback Credibility. *American Economic Journal: Microeconomics*, 7(1):99–108.

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. (2018). More Dispersion, Higher Bonuses? On Differentiation in Subjective Performance Evaluations. *Journal of Labor Economics*, 36(2):511–549.

Kim, S. K. (1995). Efficiency of an Information System in an Agency Model. *Econometrica*, 63(1):89–102.

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. (2018). Subjective Evaluations with Budget-Balanced Contracts. *Mimeo, http://lang.userweb.mwn.de/files/wp2018.pdf.*

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.

Pearce, D. G. and Stacchetti, E. (1998). The interaction of implicit and explicit contracts in repeated agency. *Games and Economic Behavior*, 23(1):75–96.

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.