Lang, Matthias

Working Paper

Communicating Subjective Evaluations

CESifo Working Paper, No. 4830

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Lang, Matthias (2014) : Communicating Subjective Evaluations, CESifo Working Paper, No. 4830, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/102244

The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

CESifo Working Papers

www.cesifo.org/wp

Communicating Subjective Evaluations

Matthias Lang

CESIFO WORKING PAPER NO. 4830
CATEGORY 11: INDUSTRIAL ORGANISATION
JUNE 2014

Presented at CESifo Area Conference on Applied Microeconomics, February 2014

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 4830*

Communicating Subjective Evaluations

Abstract

Should principals explain and justify their evaluations? Suppose the principal's evaluation is private information, but she can provide justification by sending a costly cheap-talk message. If she does not provide justification, her message space is restricted, but the message is costless. I show that the principal justifies her evaluation to the agent if the evaluation indicates bad performance. The justification assures the agent that the principal has not distorted the evaluation downwards. In equilibrium, the wage increases in the agent's performance, when the principal justifies her evaluation. For good performance, however, the principal pays a constant high wage without justification.

JEL-Code: D820, D860, J410, M520.

Keywords: communication, justification, subjective evaluation, stochastic contracts, disclosure.

Matthias Lang
Humboldt University Berlin
Institute of Economic Theory I
Spandauer Str. 1
Germany – 10099 Berlin
lang@uni-bonn.de

May 2014
I thank Helmut Bester, Christoph Engel, Leonardo Felli, Ludivine Garside, Thomas Giebe, Alia Gizatulina, Olga Gorelkina, Dominik Grafenhofer, Paul Heidhues, Martin Hellwig, Sergei Izmalkov, Johannes Koenen, Daniel Krähmer, Marco Ottaviani, Régis Renault, Ilya Segal, Caspar Siegert, Roland Strausz, Dezsö Szalay, and Stefan Terstiege for very helpful discussions, and the audiences at the CESifo conference 2014, World Congress of the Game Theory Society 2012, SFB/TR-15 2012 fall meeting, Econometric Society NASM and EM 2011, VfS 2010, EEA 2010 meetings, and seminars in Berlin (Humboldt), Bonn (MPI Econ and University), Maastricht and Munich (LMU) for comments.

This paper analyzes communication in a principal-agent model in which the principal's performance measure is nonverifiable by third parties and unobservable by the agent. To make contracting feasible, the principal communicates with the agent. Such subjective or nonverifiable measures of performance are widely used, as verifiable, i.e., objective, performance measures are often unavailable.[1] Examples for subjective measures of performance are evaluations by supervisors, co-workers, and others. Their subjectivity, however, allows the principal to choose whether and how to disclose and to justify her evaluation of the agent's work. An endogenous hold-up problem arises. This problem explains some of the emphasis personnel policies place on feedback and communication.

In the model, an agent (he) works for the principal (she) who privately receives information about the agent's performance, like reports from colleagues, observations of the agent at work or of the agent's output. By random encounters or joint observations, the agent learns a very small fraction of the principal's information. These shared signals, however, are uninformative about the evaluation of the agent's work by the principal. The principal has two options. Either she reports only the aggregated result of her evaluation or she justifies her evaluation by telling the agent about the information she collected.[2] Her message is not necessarily truthful, and providing justification is costly. The agent replies with a cheap-talk message about the shared signals. As the messages are the only third-party enforceable information, the contract just depends on these messages. The paper studies the resulting communication pattern: in equilibrium the principal justifies only bad evaluations. In this case, the wage is increasing 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.

The intuition for this communication pattern is the following. First, it is never optimal to justify all evaluations, because communication is costly. Second, the agent cannot verify the evaluation without justification. Hence, without justification the principal has an incentive to choose the evaluation yielding the lowest wage payment. Therefore, no wage dispersion is feasible and there is pooling if the principal provides no justification. Third, providing justification reduces the wage. Otherwise, not providing justification would allow the principal to save on wage and communication costs. Hence, the highest wages lack justification. Finally, the monotone likelihood ratio property of the agent's performance ensures that, with regard to justifications, a threshold strategy is optimal. The principal justifies bad evaluations, but does not justify good evaluations. If performance is bad, the principal has to bear the communication costs, but pays a lower wage. If performance

[1] The extensive use of subjective performance measures is confirmed by Dessler (2008, p. 339), Porter et al. (2008, p. 148), 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] Justifications of subjective evaluations are a common HR practice: "92% require a review and feedback session as part of the appraisal process." (Dessler, 2008, p. 366)

is good, she pays a higher wage instead of justifying her evaluation.[3] Compared to a common moral-hazard setting additional incentives are necessary: ex-ante the principal wants to justify some evaluations to the agent ex-post. Nevertheless, ex-post she wants to save on wage and communication costs by not providing justification. 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 her evaluation.

The second contribution of this paper is to model a 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 paper, the agent learns his evaluation by the principal when receiving the principal's message. In addition, the agent can to some extent verify the principal's message although the result of the evaluation and the agent's information are uncorrelated and stochastically independent. In contrast to previous literature I do not assume an exogenous verification technology, type-dependent messages spaces, or that messages are verifiable by a third party. All messages are cheap-talk. Hence, a third party cannot tell whether a message is truthful.

The mechanism uses the fact that the principal and the agent share some observations of the environment and the processes that lead to the evaluation. These shared observations are uninformative about the result of the evaluation and have mass zero with respect to the principal's information resulting in the evaluation. Nevertheless, the principal recalls all observations to justify the evaluation. 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 meaningful distortion. The reason is that the principal does not know which observations the agent has learned.

As an example consider a chef cooking for the principal. The health-conscious chef prepares the food using a number of ingredients, but without salt. Suppose the principal pretends not to like the food and justifies her assessment saying that the food was too salty. Although the chef is unaware of the principal's real evaluation of the food, he knows that the principal is lying. Nonetheless, the mechanism used here is not limited to employment relations. It applies more generally to moral hazard and hold-up settings whenever the contracting parties interact and share some information. The mechanism only requires that the informed party is able to provide justification.

Moreover, the optimal contract fits well with empirical observations that evaluations are lenient and wage dispersion for the best evaluations is low.[4] Those observations

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

[4] According to Bretz et al. (1992), usually 60–70% of all employees get an evaluation from the best or second-best category. Moreover, "Medoff and Abraham (1980) found in two companies that, among the 99% of employees in the same position who received the top three performance ratings, the difference in salary between the highest and lowest rated employees was about 5%." (Gibbs, 1991, pp. 4-5) Similarly, Murphy (1993, p. 56) reports that the top 1% of employees at the pharmaceutical company Merck receive a pay raise just 3% higher than the median employee in 1985.

are typically referred to as leniency bias and centrality bias. This paper argues that this pattern can be understood as a feature of the optimal contract instead of biased behavior. In addition, many studies show that principals evaluating for developmental or feedback purposes are more likely to differentiate among subordinates than they are when the evaluation is used for administrative purposes, like merit increases or promotions.[5] In the latter case, evaluations are more compressed and show less variation between employees. The finding goes back to Taylor and Wherry (1951, p. 39). They find "a marked distortion [of evaluations] ... with considerably poorer discrimination at the top." This observation is in line with the predictions of this paper. The principal must be given explicit incentives to report her evaluation truthfully. These incentives cause pooling of the best evaluations. If the evaluation is for developmental purposes, 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 wage payments. Managers at Merck, for example, experienced that "the salary link made discussions on performance improvement difficult." (Murphy, 1993, p. 58) Psychological costs of supervisors to give bad evaluations to their subordinates yield no straightforward explanation of this pattern, since those costs apply to evaluations for all purposes similarly.

Finally, this paper shows that the optimal contract can be ex-post budget-balanced. Hence, the contract requires no payments to third parties in contrast to previous models. Instead, stochastic contracts use differences in the risk preferences of the contracting parties to implement the required incentives.

The remainder of the paper is organized as follows. Section 1 discusses the related literature. Section 2 illustrates the setting using a case study. Section 3 sets up the model and characterizes the optimal communication pattern. Section 4 points out a more familiar implementation of the optimal contract by an indirect mechanism. Section 5 makes the optimal contract ex-post budget-balanced. Section 6 examines the robustness of the model. Section 7 contains the concluding remarks. All proofs are relegated to the appendix.

1 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

[5] This effect is found in Dessler (2008, p. 356), Milkovich et al. (2008, p. 351), a meta-study by Jawahar and Williams (1997), Jawahar and Stone (1997), Harris et al. (1995), McDaniel et al. (1994), Milkovich and Wigdor (1991, pp. 3, 72), and Landy and Farr (1980).

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 Compte (1998), Kandori and Matsushima (1998), Baker et al. (1994), MacLeod and Malcomson (1989), and Bull (1987). Then reputation effects created by the continuation value for both contracting parties allow subjective performance measures to gain credibility and to be used as the basis for the agent's incentives. Levin (2003) drops the assumption that the subjective performance measure is perfectly observable by both contracting parties. In this case optimal contracts often have a termination form, i.e., the contract ends after observing bad performance. 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 and introduces a message game. Each party reports their information by sending a public message. This enables the parties to condition their contract on these messages. As the information structure is exogenously given, the principal cannot decide, depending on the performance measure, whether to justify her evaluation. Thus, the 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. This corresponds to obligatory or costless justification in my model as in Lemma 2. If the signals are uncorrelated, the optimal contract in MacLeod (2003) resembles the case of prohibitively expensive justification in my model. The case of imperfect correlation in combination with a binding upper limit on wage payments shares some features with the optimal contract here, but the reasoning and the proofs are different. First, I do not assume an upper limit on payments. Second, the agent receives no private signals telling him that he received no information. Instead, it is the principal's incentive – resulting from the contract and the communication costs – to withhold and distort her evaluation that yields the compression at the top result. Economically, the main difference between this paper and MacLeod (2003) is that I endogenize communication. In addition, I discuss the resulting communication pattern.

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 communication is costly. As justification allows 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 by a costly audit. They show the optimality of debt contracts, which are similar to the optimal contract in my paper, as there are no audits for high payments. In this literature, however, the firm learns its performance, while the investor chooses whether to perform an audit. In my model, due to the nature of a justification it is the better-informed party that makes the communication decision. This is also the reason why mixed strategies with respect to the communication strategy are not optimal in my setting. In addition, the communication need not be truthful and cannot be verified directly by one of the contracting parties, while the result of an audit is truthful and verifiable.

In Rahman (2012), the principal instructs the agent to shirk with strictly positive probability. These instructions create shared observations between the principal and the monitor. The shared observations allow the principal to verify the monitor's report if the probability of shirking is strictly positive. In my model of justification, it is sufficient that there are some shared observations, but they can be uninformative and have mass zero.

Following truthful communication, the performance measure becomes observable, but unverifiable – similarly to a hold-up setting. Aghion et al. (2012), Hart and Moore (1988), and Grossman and Hart (1986) discuss solutions to this problem. In my model, preferences are independent of the evaluation, while in the hold-up setting the preferences depend on types or invested efforts. Therefore I cannot replicate the solutions of these models. In contrast to the literature on informed principals, the principal's information arises during the principal-agent relationship and is unavailable at the contracting stage.

Furthermore, credibility of promised incentives is sometimes discussed under the notion of fairness and 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.

Finally, the present article concerns stochastic contracts and ex-post budget balance. Previous literature, like MacLeod (2003) or Fuchs (2007), requires payments to third parties. This allows the contracting parties to renegotiate in order to avoid paying money to an outsider – as already discussed by Hart and Moore (1988). If stochastic contracts

are possible, I show how to establish ex-post budget balance. Maskin and Tirole (1999) use a similar mechanism to implement incomplete contracts in an investment setting. Rasmusen (1987) shows that stochastic wage payments ensure ex-post budget balance in a team-production setting. He does not consider differences in risk aversion between the principal and the agent, as the principal's payment is complete deterministic, only the sharing rule between the agents is stochastic. In my model, the principal's payment has to be stochastic to guarantee budget balance.

2 Subjective Evaluations at Work

As an example, consider the performance evaluation 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, and skills as a team worker. In each category, they receive a rating on a scale from one to five. The average rating across the seven categories yields the result of the evaluation that is used for compensation purposes.

Suppose that the principal receives a signal in each category. For example, she listen to costumers complaining or praising the employee. Then, she talks to the agent's colleagues to learn about his skills as a team worker. Finally, she observes the agent at work or the agent's output. This closely captures a practical evaluation process, as "an appraiser would use evidence from direct observation of the employee, or by reports from others, to make judgment about the appraisee's performance." (Porter et al., 2008, p. 149) These signals are subjective and private information of the principal. The agent, however, sometimes gets direct feedback from customers or is told by colleagues about their reports to the principal. Hence, he observes a small number 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. A justification tells the agent all signals. Providing justification is costly, as it requires the principal to spend additional time on the evaluation.[6] If the principal provides justification, the agent can partially verify the evaluation. In the absence of a justification, the agent cannot detect any distortions in the evaluation. The next section formalizes these notions.

3 Justify Bad Evaluations

3.1 Model

Consider a risk-averse agent working for a risk-neutral principal. The principal proposes a contract $\mathcal{W}$ to the agent. The contract $\mathcal{W}$ specifies the agent's wage depending on

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

any information that is available at the time of the wage payments and enforceable by a third party. After signing such a contract, the agent chooses his work effort $e \in [0,1)$, which is unobservable by the principal. Then the principal receives signals $I(t) \in \{0,1\}$ in different categories $t \in T = [0,1]$. Define the result of the subjective evaluation as the average $\mu = \int I(t)\mathrm{d}t$. In each category $t \in T$, the signal is independent and identically distributed and declares success, $I(t) = 1$, with probability p and failure, $I(t) = 0$, with probability $1 - p$. The probability p is once drawn from $F(p|e) = eF^H(p) + (1-e)F^L(p)$ depending on the agent's effort e. The distributions $F^H(p)$ and $F^L(p)$ admit continuous densities $f^H(p), f^L(p) > \epsilon > 0$. In addition, the distribution $F(p|e)$ satisfies the monotone likelihood ratio property, i.e., the ratio $f^H(p)/f^L(p)$ is strictly increasing in p. Notice that the evaluation μ is a sufficient statistics for the agent's effort, as $\mu = \int_T I(t)\mathrm{d}t = p$.[7] The monotone likelihood ratio property ensures that a higher evaluation μ indicates higher work effort.

The agent observes only the signals in a finite subset $S \subset \{t \in T | I(t) = 1\}$ of successful categories with $|S| = n$. In Section 6 the agent also learns the signals in failed categories. Assume that the agent's sample S is large and consists of a random draw with full support over the successful categories, no atoms and a density bounded away from zero.[8] The set S is private information of the agent, while its cardinality n is common knowledge. Hence, the principal knows how many categories, but not which categories the agent observes. Notice that the evaluation μ and the agent's information S are stochastically independent. For later reference, denote by $P(S|I,e)$ the conditional distribution and by $P(I,S|e)$ the joint distribution of the agent's sample S and the principal's information I: $t \mapsto I(t)$ conditional on the agent's effort e.

Now, turn to communicating the evaluation. $\beta \in \{0,1\}$ denotes the principal's justification decision. For $\beta = 0$, she does not provide justification and her message space is restricted to $R = [0,1]$. For $\beta = 1$, she justifies her evaluation of the agent's work at costs κ by sending a message m_P from $\mathcal{I} = \{0,1\}^T$.[9] Independently of the principal's choice, the agent replies with a cheap-talk message $m_A \in T^n$. Both parties can lie and send any message from the corresponding messages spaces. Section 6 discusses general message spaces. Finally, the contract $\mathcal{W}$ is performed according to the available and enforceable information, the messages m_P and m_A. The contract $\mathcal{W}$ 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 Proposition 2 in MacLeod (2003) and Proposition 1 in Fuchs (2007) demonstrate, some surplus has to be destroyed in this kind of model to implement pos-

[7] I assume a law of large number here. Judd (1985) constructs a probability measure that allows avoiding measurability problems in formulating a law of large numbers for a continuum of random variables. Sun (2006) proves such a law of large numbers assuming essential pairwise independence.

[8] All results are such that $\exists \bar{N} \in \mathbb{N}$ and the results are valid for all $n > \bar{N}$. If n is small, e.g., $n = 1$, the results also hold, but require higher off-equilibrium payments.

[9] $\{0,1\}^T$ denotes the set of all functions $T \to \{0,1\}$. For $\beta = 0$, I identify a message $m_P \in R = [0,1]$ in the restricted message space R with the step function: $T \to \{0,1\}$ with 1 for $t \leq m_p$ and 0 otherwise.

- At $t = 0$, principal proposes a contract $\mathcal{W}$ to the agent.
- At $t = 1$, agent can accept the contract offer and determine his work effort e.
- At $t = 2$, principal collects information I, while agent learns S.
- At $t = 3$, principal either provides justification $m_P \in \mathcal{I}$ at costs κ or sends only a short message $m_P \in R$.
- At $t = 4$, agent sends a message $m_A \in T$.
- At $t = 5$, principal makes payments according to $W(m_P, m_A)$ and $c(m_P, m_A)$.

Figure 1: Timing of the Model

itive effort of the agent. To account for this, I allow for $W(m_P, m_A) \geq c(m_P, m_A)$. As the message spaces are sufficiently rich to encode β, e and μ, it is unnecessary to include them explicitly in the contract. The principal has no commitment power other than the contract.

The agent is represented by a utility function $u(c(m_P, m_A)) - d(e)$ if he chooses effort e and receives a wage $c(m_P, m_A)$. The function u is strictly concave with the limit $\lim_{w \to 0} u(w) = -\infty$ and derivatives $u' > \epsilon > 0$ and $u'' < \hat{\epsilon} < 0$. The function d, the disutility of performing effort, is also increasing, but strictly convex with the limit $\lim_{e \to 1} d(e) = \infty$. Both functions are twice continously differentiable. The agent receives a reservation utility $\bar{u}$ if he rejects the principal's offer. There is a $B > 0$, such that the principal's benefit is

$$\begin{cases} B\mu - W(m_P, m_A) - \kappa & \text{if she provides justification} \\ B\mu - W(m_P, m_A) & \text{if she does not provide justification} \end{cases}$$

after paying a wage $W(m_P, m_A)$.

3.2 Analysis

It is crucial here that the principal has to pay communication costs κ to transmit the information I. The model ensures this by making it impossible to encode the information I in a message from the restricted message space R. The reason is that the cardinality of $\mathcal{I}$ and R differs. Due to the restriction of the message space for $\beta = 0$, the classical revelation principle does not apply. See Green and Laffont (1986) for an example. Hence, it is unclear whether truthful revelation is optimal in my setting. Define truthful revelation by the principal as

$$\theta(I, \beta) = \begin{cases} I & \text{if } \beta = 1 \\ \int_T I(t) \mathrm{d}t & \text{if } \beta = 0. \end{cases}$$

Lemma 3 in the appendix proves that truthful revelation is optimal if an optimal contract exists. Thus, in the optimal contract principal and agent send a truthful message that reveals their private information, the principal's information I or its average μ and the

agent's sample S respectively.[10] Grossman and Hart (1983) prove that the model can be solved in two steps. First, for every level of effort e, the optimal contract $\mathcal{W}$ and its expected costs $C(e)$ for the principal are computed. The second step determines the optimal effort level e by

$$\max_{e\in[0,1)} B\int p\mathrm{d}F(p|e) - C(e).$$

Now, returning to the first step, Program A below determines the optimal contract that implements effort e by choosing payments $W(m_P, m_A)$ by the principal, $c(m_P, m_A)$ for the agent, and which evaluations to justify, $\beta(I)$. The objective is to minimize the expected costs subject to several conditions. The participation constraint (PC) makes the agent willing to accept the proposed contract. The agent's incentive compatibility (IC_A) guarantees that the agent chooses the desired level of effort. The principal's incentive compatibility (IC_P) gives her an incentive to justify her evaluation if justification is desired. In addition, sending a truthful message has to be incentive compatible for the principal (IC_P) and the agent (IC_{m_A}). Finally, the principal's payment has to be higher than the wage received by the agent. To simplify the exposition, define equilibrium payments $W_{eq}(I,S) = W\big(\theta(I,\beta(I)),S\big)$ and $c_{eq}(I,S) = c\big(\theta(I,\beta(I)),S\big)$.

$$C(e) = \inf \int W_{eq}(I,S) + \kappa\beta(I)\mathrm{d}P(I,S|e), \tag{A}$$

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

$$e \in \arg\max \int u\big(c_{eq}(I,S)\big)\mathrm{d}P(I,S|e) - d(e), \tag{IC_A}$$

$$\kappa\beta(I) + \int W_{eq}(I,S)\mathrm{d}P(S|I,e) \leq \tilde{\kappa} + \int W(\hat{I},S)\mathrm{d}P(S|I,e)$$
$$\forall I \in \mathcal{I}, \forall(\tilde{\kappa},\hat{I}) \in \big\{\{0\}\times R, \{\kappa\}\times\mathcal{I}\big\}, \tag{IC_P}$$

$$\int u(c_{eq}(I,S))\mathrm{d}P(I|S,e) \geq \int u(c_{eq}(I,\hat{S}))\mathrm{d}P(I|S,e) \qquad \forall S,\hat{S}\in T^n, \tag{IC_{m_A}}$$

$$W(I,S) \geq c(I,S) \qquad \forall I\in\mathcal{I}, \forall S\in T^n. \tag{1}$$

If the principal's information I were observable and contractible, the principal's and the agent's incentive for sending truthful messages (IC_P) and (IC_{m_A}) can be neglected. Denote this problem by A* and the solution, the optimal complete wage, by $w_e^*(\mu)$.

Lemma 1. *If the principal's information I is contractible, the optimal wage $w_e^*(\mu)$ depends only on the evaluation μ, i.e., the average of the principal's information I. $w_e^*(\mu)$ is almost surely continuous and increasing in μ for positive effort, $e > 0$.*

As Holmström (1979) shows, the wage depends only on the sufficient statistics μ instead of the entire information I and a better evaluation implies a higher wage.

If the principal's information is subjective and justification is a choice variable of

[10] Yet it is impossible to have the agent reveal his work effort e truthfully and make the wage dependent on his message about the effort in order to implement $e > 0$. Therefore it is without loss of generality for the messages to contain only the information the contracting parties collected at $t = 2$.

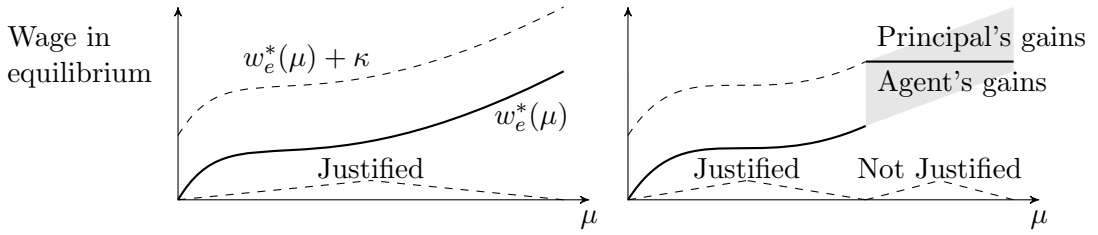


Figure 2: Idea of the Proof of Proposition 1

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

the principal, the additional incentive constraints for the messages do matter. Yet these incentive constraints do not change the equilibrium wage in the absence of communication costs when $\kappa = 0$.

Lemma 2. *If communication costs are $\kappa = 0$, all evaluations are justified.*

The equilibrium wage is $w_e^*(\mu)$, the same as in Lemma 1. Whenever communication is costly and $\kappa > 0$, however, the equilibrium wage differs from $w_e^*(\mu)$. To gain some intuition, suppose for the moment that the principal justifies all evaluations. Then the optimal contract implements wage payments $w_e^*(\mu)$ in equilibrium as in Lemma 2. Yet, the principal can modify this contract to save on communication costs. The reason is that the agent does not suspect a distorted evaluation by the principal and does not demand an explanation for the highest wages. Therefore it is suboptimal to justify all evaluations. For ease of exposition, denote the justification set by $\mathcal{I}_C = \{I \in \mathcal{I} | \beta(I) = 1\}$.

Proposition 1. *If $\kappa > 0$, justifying all evaluations is not optimal, i.e., $\Pr(\mathcal{I}_C) < 1$.*

The proof in the appendix shows that the principal's total costs decrease if the principal refrains from justifying the best evaluations. By paying a constant wage that is not justified, it is possible to reduce communication costs. These efficiency gains go partly to the principal and partly to the agent as indicated by the gray ares in Figure 2. The proof shows that the new contract satisfies all constraints in Program A. To determine the justification set, it is necessary to know more about the optimal contract. I call the messages m_P and m_A agreeing if $m_P(t) = 1$ for all $t \in m_A$. Otherwise, they disagree.

Proposition 2. *Suppose $\Pr(\mathcal{I}_C) > 0$. Then optimal wages are constant and equal w^{**} if the principal's message is not in the justification set. Otherwise, wages depend on the average of the principal's message, i.e., the evaluation μ, if the messages agree.*

$$c^{**}(m_P, m_A) = \begin{cases} w^{**} & \text{if } m_P \notin \mathcal{I}_C \\ w^{**}\left(\int m_P(t)\mathrm{d}t\right) & \text{if } m_P \in \mathcal{I}_C \end{cases}$$

$$W^{**}(m_P, m_A) = \begin{cases} w^{**} & \text{if } m_P \notin \mathcal{I}_C \\ w^{**}\left(\int m_P(t)\mathrm{d}t\right) & \text{if } m_P \in \mathcal{I}_C \text{ and } m_P(t) = 1 \text{ for all } t \in m_A \\ w^{**} + \kappa & \text{otherwise.} \end{cases}$$

The principal justifies low payments, as $w^{**}\left(\int m_P(t)\mathrm{d}t\right) \leq w^{**} - \kappa$ for all $m_P \in \mathcal{I}_C$.

There is pooling of wages, but evaluations are unbiased in the optimal contract. Yet, this is for ease of exposition only. It is easy to introduce also pooling of evaluations for $I \notin \mathcal{I}_C$. This requires only the principal sending a pooling message for $I \notin \mathcal{I}_C$. In the contract the agent's wage is either constant or depends only on the average of the principal's message m_P. Additionally, the principal's payments are higher in the absence of justification. Otherwise the principal would deviate and would not justify the evaluation, because the agent cannot verify such a deviation. Furthermore, the agent's wage equals the principal's payments in the absence of a justification. Varying the agent's wage in the absence of a justification eases the agent's incentive compatibility (IC_A), but makes it more difficult to satisfy the agent's participation constraint (PC). The principal's costs remain unchanged. The proof shows that the second effect dominates the first one due to the monotone likelihood ratio property. Therefore the agent's wage equals the principal's payments in the absence of a justification.

If the principal provides justification, the agent can detect deviations by the principal, as deviations trigger a mismatch in the messages. Therefore payments can vary in the principal's message m_P. A disagreement in the messages is punished by making the principal pay the highest wage. If the messages agree, the principal's payment equals the agent's wage. This structure of the contract allows characterizing the optimal justification strategy, that follows a threshold rule. The principal justifies only bad evaluations. In these cases, the agent suspects a distortion by the principal and insists on a justification of bad evaluations.

Proposition 3. *Suppose* $\Pr(\mathcal{I}_C) > 0$. *In the optimal contract, the principal justifies evaluations* μ *below a threshold* $\delta \in (0,1)$, *while she does not justify evaluations above* δ.

The proposition exhibits the communication pattern described in the introduction and summarized in Figure 3. There is partial justification. The principal justifies only bad evaluations and low wages, while she remains silent on good performance. This confirms empirical observations, like the leniency bias and the centrality bias that there is less distinction in subjective evaluations than in the underlying performance measure, in particular at the top. Yet this behavior is not the result of a bias, but part of the optimal contract, which pools several evaluations and rewards them with the same wage. Thus, the contract eliminates wage differences that the principal would have to justify.

Suppose that the optimal communication pattern does not satisfy Proposition 3. Then there would be a threshold μ^* such that the principal justifies some evaluations above μ^*, but does not provide justification for some evaluations below μ^*. In particular, such a communication pattern implies that the equilibrium wage is locally decreasing in the evaluation μ. The monotone likelihood ratio property ensures that a decreasing equilibrium wage cannot be optimal.

Formally, such a communication pattern implies a constant wage for the evaluations above μ^* in the justification set. To show this, the proof establishes that the Lagrange

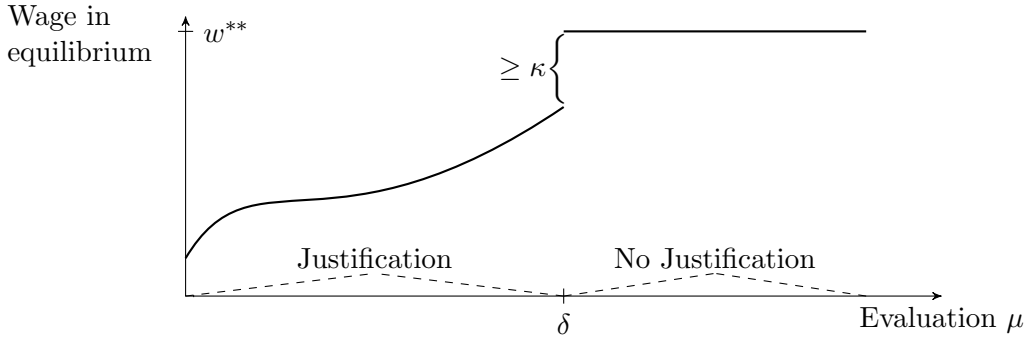


Figure 3: Proposition 3 and the Equilibrium Wage

In the optimal contract there is pooling of wages, but evaluations are unbiased. It is easy, however, to have pooling of evaluations, too. Then the principal sends $m_P = 1$ if $\mu \geq \delta$. No change to the optimal contract is required.

multiplier of the modified principal's incentive compatibility (IC_P) is positive on this set of evaluations. Split the set of the evaluations above μ^* in the justification set into a lower and an upper half. Then adjust the contract so that the principal justifies evaluations in the lower half, but not in the upper half. At the same time the agent's wage is decreased in the lower half and increased in the upper half such that the agent's expected utility remains constant. Hence, the agent's participation constraint (PC) remains satisfied. Finally, I show that the agent's incentive compatibility (IC_A) becomes slack by this contract modification due to the monotone likelihood ratio property. Yet, in this case the initial communication pattern cannot be optimal. Proposition 2 characterized the optimal contract. It remains to determine the values of w^{**} and $w^{**}(\mu)$.

Proposition 4. *Suppose $\delta > 0$. The wages w^{**} and $w^{**}(\mu)$ for $\mu \in [0, \delta]$ are determined by*

$$\inf_{w(\mu),w,\delta} \int_0^\delta w(\mu) + \kappa dF(\mu|e) + (1 - F(\delta|e))w, \tag{D}$$

$$\text{subject to} \int_0^\delta u\big(w(\mu)\big)\, dF(\mu|e) + (1 - F(\delta|e))u(w) - d(e) \geq \bar{u}, \tag{PC}$$

$$\int_0^\delta u\big(w(\mu)\big)(f^H(\mu) - f^L(\mu))\, d\mu + u(w) \int_\delta^1 f^H(\mu) - f^L(\mu)\, d\mu = d'(e) \tag{IC$_A$}$$

Justification is indeed optimal if communication costs are not prohibitively high as specified in the next proposition. Then the optimal contract makes the principal justify some evaluations.

Proposition 5. *In the optimal contract, justification occurs with positive probability, $\Pr(\mathcal{I}_C) > 0$, if the principal wants to implement positive effort $e > 0$ and the communication costs are not too high,*

$$\kappa < u^{-1}\left(\bar{u} + d(e) + \frac{f(0|e)}{f^L(0) - f^H(0)} d'(e)\right). \tag{2}$$

No justifications are provided for higher communication costs.

Justification is beneficial, although it is costly and conveys no additional information about the agent's effort. Justification also does not provide the principal with information for her decision making nor gives the agent instructions in the sense of learning or which

tasks to perform. Instead, justification makes the principal's promise of incentives to the agent credible and allows the principal to assure the agent that her evaluation is not distorted. Thus, it is in the principal's interest to be transparent about her evaluations, even if justification is costly and takes place after the agent's effort choice. Next, consider the optimal amount of justification.

Proposition 6. *For $e > 0$, the justification threshold δ is determined by*

$$u(w^{**}) - u(w^{**}(\delta)) = u'(w^{**}(\delta))(w^{**} - w^{**}(\delta) - \kappa).$$

For small communication costs, the contract converges to complete justification and to the optimal complete contract from Lemma 1: $\delta \nearrow 1$ for $\kappa \searrow 0$.

Finally, consider two extensions to simplify the contract and make it ex-post budget balanced.

4 Indirect Mechanism

It is possible to simplify the optimal contract by changing the reporting strategies. Let the principal report $m_P = 1$ if the evaluation μ is above the threshold δ. Otherwise, justify the truth, i.e., send message $m_P = I$. Reduce the agent's message to the binary decision whether to accept or reject the evaluation by the principal. If he accepts the evaluation $\mu = \int m_P(t)\mathrm{d}t$, the principal pays him the corresponding wage $w^{**}(\mu)$. If he rejects, the principal has to pay $w^{**} + \kappa$, while the agent receives $w^{**}(\mu)$. Thus, the agent has the possibility to object to the principal's evaluation. This conflict resolution might be quite 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."

The indirect mechanism leaves the incentives of the contracting parties unchanged. If the principal receives information in the justification set, any deviation makes her worse off, as the deviation increases her payments to at least w^{**}. For evaluations outside the justification set, it is also unprofitable to deviate, as any evaluation yielding a lower wage will be rejected. For the agent, on the other hand, the following strategy is a best reply: accept an evaluation if and only if the principal proposed $\mu \geq \delta$ or she provides justification that matches the agent's information. Consequently, the modified setting also implements effort e of the agent at optimal costs. Hence, the relevant part of the model is the principal's decision to justify her evaluation to the agent at $t = 3$.

5 Stochastic Contracts

For stochastic payments, the expected value for the principal is higher than the agent's certainty equivalent. Therefore it is possible to replace payments to a third party by

stochastic payments. This does not require a risk-neutral principal. As long as the degree of risk aversion differs between the principal and the agent, the optimal contract can achieve ex-post budget balance.

Proposition 7. *Stochastic contracts allow making the optimal contract ex-post budget-balanced:*

$$\bar{W}^{**}(m_P, m_A) = \begin{cases} w^{**}(\int m_P(t)\mathrm{d}t) & \text{if } m_P \in \mathcal{I}_C \text{ and } m_P(t) = 1 \text{ for all } t \in m_A \\ w^{**} & \text{if } m_P \notin \mathcal{I}_C \\ w^{**} + \Lambda_{z(\int m_P(t)\mathrm{d}t)} & \text{otherwise.} \end{cases}$$

The lotteries Λ_z have $\mathbb{E}(\Lambda_z) = \kappa$ and a certainty equivalent for the agent of $w^{**}(\mu) = u^{-1}(\mathbb{E}[u(w^{**} + \Lambda_{z(\mu)})])$. The values of w^{**} and $w^{**}(\mu)$ are determined in Proposition 4.

Stochastic contracts ensure ex-post budget balance. Examples are stock options whose valuation is influenced by external shocks to the financial sector or uncertain arbitration procedures. The contracting parties might be uncertain how a disagreement is interpreted and which wage payment is appropriate.

6 Robustness of the Results

So far I have assumed specified message spaces. The only restriction on messages spaces is with respect to cardinality, however. Denote the agent's message space by $\mathcal{M}_A$ and the principal's one by $\mathcal{M}_P$ if she provides justification. Remember the principal's message space is R if she does not provide justification. The optimal contract, in particular, the discussion in Section 4, implies that $\mathcal{M}_A$ can be replaced by any set with at least two elements. R needs to be such that messages in $\mathcal{M}_P$ cannot be encoded or approximated in R. Therefore R can be replaced by any set with a cardinality of at most $|T|$ as discussed in the proof of Lemma 3. Finally, $\mathcal{M}_P$ has to be sufficiently rich, i.e., $|\mathcal{M}_P| \geq |\mathcal{I}|$. As long as these assumptions on the cardinalities are satisfies, the analysis in this paper remains valid.

In the model, the agent learns the reports of a finite subset S of positive signals. It is possible, however, to extent the model to include negative signals into S. For this purpose assume, that there are informative and uninformative categories. A category is informative with probability $q \in (0, 1)$. An informative category reports a success with probability p as before. With probability $1 - q$ the category is uninformative and reports statistical noise, i.e., success or failure with probability $1/2$ each. The quality of a category is unobservable. Figure 4 depicts the distribution in each category. The agent's sample S is drawn randomly from the uninformative categories. Then the agent learns

$$\begin{cases} I(t) & \text{if } t \in S \\ 0 & \text{if } t \notin S. \end{cases}$$

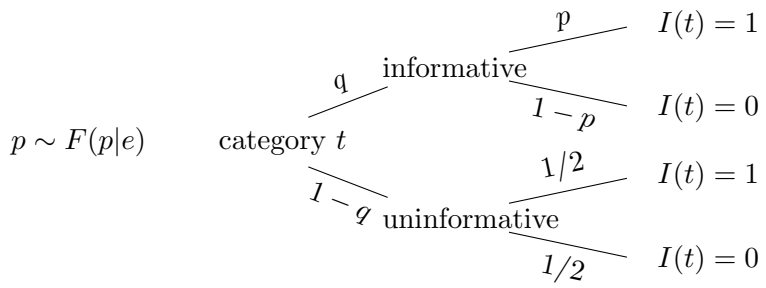


Figure 4: Alternative Distribution of the Signal in a Category $t \in T$

In this case my results remain valid even if S is unobservable by the agent. There are some adaptations to the optimal contract, however. In particular, the messages agree, if $m_P(t)m_A(t) = m_A(t)$ for all $t \in T$.

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. However, as long as the bias is systematical, it is possible to adapt the definition of an agreement in the messages. Then the results of this paper remain valid. Suppose, for example, that the agent perceives a finite number of additional categories S' to report success independently of their true value. If S' is known at the time the contract is proposed, the contract just neglects these elements of the agent's message. If S' is unknown at the time the contract is proposed, agreement is defined as $m_P(t) = 1$ for all $t \in m_A$ with $|S'|$ exceptions.

So far there is only one messaging stage, i.e., both the agent and the principal are allowed to send just one message. This assumption ensures that pure (communication) strategies are optimal. If there were an alternating sequence of messages, random communication might become optimal. In this case, the principal first sends the result of the evaluation as a costless message. Then the agent with some small probability requests justification. After that the principal has the possibility to reply with a justification which has to match a subsequent message by the agent. This sequence allows saving communication costs but requires high punishments out of equilibrium.

In the model, the agent can learn the principal's justification decision β from the message, as the restricted message space R has mass zero in $\mathcal{I}$. This is unnecessary, however. Assume that a message in $m_P \in R = [0, 1]$ is identified with a uniformly distributed lottery over $\{I \in \mathcal{I} | \int I(t)\mathrm{d}t = m_P\}$, i.e., all the functions $I(t)$ with an average of m_P. The contract proposed in the paper still remains optimal in this setting.

7 Conclusion

This article discusses 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, in the optimal contract the

principal justifies some evaluations. This justification allows the agent to detect distorted evaluations. Therefore providing a justification makes the incentives for the agent credible. In the optimal contract, the principal justifies only bad evaluations. This communication pattern results in pooling and wage compression at the top, as illustrated in Figure 3 on page 13. These results fit well with empirical observations, often referred to as leniency bias and centrality bias.[11] The paper argues that this pattern of evaluations is a feature of the optimal contract with unbiased agents and no proof of biased behavior per se.

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

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, the pooling at the top could cause the costs of an incentives scheme to be substantial if there is a bonus attached to receiving a positive evaluation and many employees receive a positive evaluation due to the compression at the top.[12]

This paper assumes that the principal incurs costs for communicating with the agent. I would get similar results if the principal's costs instead concerned the acquisition of information. In this case, the principal only learns the result of the evaluation, i.e., the average $\mu = \int I(t)\mathrm{d}t$ of the information I. Then she decides whether to spend κ to acquire the entire information I. Independently of her choice of information acquisition, the principal can send a cheap-talk message in the unrestricted message space $\mathcal{I}$. Both settings have some merits; in reality, there could be a mixture of these two polar cases.

[11] "The distribution of ratings is typically both concentrated and biased." (Gibbs, 1991, p. 5)

[12] 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.

A Appendix

A.1 The Optimal Complete Contract

Lemma 1 characterizes the optimal complete contract and states the solution to Program A* if the principal's information I is public and verifiable. Additionally, the lemma shows that every effort $e \in [0,1)$ is implementable at finite costs $C^*(e)$.

Proof of Lemma 1: Holmström (1979) shows that the optimal wage only conditions on $\mu = \int I(t)\mathrm{d}t$, because the average of the principal's information I is a sufficient statistics for the agent's effort, $\Pr(I, S|e, \int I(t)\mathrm{d}t) = \Pr(I, S| \int I(t)\mathrm{d}t)$. In order to implement no effort, $e = 0$, set $w_0^*(\mu) = u^{-1}(\bar{u} + d(0))$ for all μ.

If the desired effort is positive, $e > 0$, 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 equals

$$\int u(w(\mu))(f^H(\mu) - f^L(\mu))\mathrm{d}\mu = d'(e). \tag{IC$_A$}$$

In addition, the constraint set is nonempty. Take for example any $\bar{w} > 0$ and the contract

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

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

Holmström (1979) proves that the Lagrange multipliers of the participation constraint λ_1 and of the incentive compatibility λ_2 are positive. Pointwise optimization[13] determines the optimal contract as

$$\begin{aligned} &f(\mu|e) - \lambda_1 u'(w(\mu))f(\mu|e) - \lambda_2 u'(w(\mu))(f^H(\mu) - f^L(\mu)) = 0 \quad \text{a.s.}, \\ &\frac{1}{u'(w(\mu))} = \lambda_1 + \lambda_2 \frac{f^H(\mu) - f^L(\mu)}{f(\mu|e)} = \lambda_1 + \lambda_2 \frac{\frac{f^H(\mu)}{f^L(\mu)} - 1}{e\frac{f^H(\mu)}{f^L(\mu)} + 1 - e} \quad \text{a.s.} \end{aligned} \tag{3}$$

Since the fraction $\frac{l-1}{el+1-e}$ is increasing in l, the right-hand side of above equation is increasing in μ due to the monotone likelihood ratio property. Therefore the concavity

[13] This technique allows for piecewise continuous functions, as Kamien and Schwartz (1991) show in Part II, Section 12. Therefore bonus wages are possible and there is no restriction to continuous contracts.

of $u(\cdot)$ implies that $w_e^*(\mu)$ is increasing in $\mu \in [0,1]$ almost surely. Moreover, $w_e^*(\mu)$ is continuous almost surely and any discontinuity is removable, because the densities f^H and f^L are continuous. □

A.2 The Optimal Contract

All evaluations are justified in the absence of communication costs, $\kappa = 0$.

Proof of Lemma 2: The following contract satisfies the additional incentive constraints (IC_P) and (IC_{m_A}):

$$c(m_P, m_A) = \quad w_e^* \left(\int m_P(t) \mathrm{d}t \right)$$

$$W(m_P, m_A) = \begin{cases} w_e^* \left(\int m_P(t) \mathrm{d}t \right) & \text{if } m_P(t) = 1 \text{ for all } t \in m_A \\ 2w_e^*(1) & \text{otherwise.} \end{cases}$$

As the agent's wage is independent of m_A, (IC_{m_A}) is trivially satisfied. Now consider a deviation by the principal. For this purpose, consider a subset $D \subset T$ with $\int_D 1 \mathrm{d}t > 0$. Then there is a $\bar{N} \in \mathbb{N}$, such that a deviation on D is unprofitable for the principal for all $n > \bar{N}$, because the probability of a mismatch in the messages is sufficiently close to 1. Additionally, deviations on larger subsets of T are also unprofitable given the assumption of full support and no atoms for the distribution of the agent's sample S. What about deviating on a subset D with infinitesimal mass $\int_D 1 \mathrm{d}t$? Consider a parameterization of such deviations $D(v)$, which maps $[0,1]$ to the Borel algebra on T, with $v = \int_{D(v)} 1 \mathrm{d}t$. In the limit for $v \searrow 0$ the potential gain of a deviation *and* the probability of detection converge to 0. Yet, for $v \searrow 0$ such a deviation is unprofitable if

$$\frac{\partial w_e^*(\mu)}{\partial \mu} < w_e^*(1) \left. \frac{\partial \Pr(S \cap D(v) \neq \emptyset | I, e)}{\partial v} \right|_{v=0} \qquad \text{for all } \mu \in [0,1].$$

The left-hand side of this inequality captures the potential gain of a deviation and is determined by Eq. (3). If the deviation is detected, the principal's payment increases by $2w_e^*(1) - w_e^*(\mu) \geq w_e^*(1)$. The right-hand side of this inequality equals this increase multiplied with the density of at least one category from the agent's sample being an element of $D(0)$. This density is increasing in the size n of the agent's sample. Therefore there is a $\bar{N} \in \mathbb{N}$, such that a deviation on D is unprofitable for the principal for all $n > \bar{N}$. Hence, the additional incentive constraints (IC_P) and (IC_{m_A}) are satisfied.

The contract proposed above is the optimal contract, because it satisfies the additional incentive constraints (IC_P) and (IC_{m_A}) and implements the equilibrium wage $w_e^*(\mu)$. As $w_e^*(\mu)$ is a solution to Problem A* according to Lemma 1, this contract is a solution to Problem A for $\kappa = 0$. □

It is not optimal to provide justification almost surely for $\kappa > 0$.

Proof of Proposition 1: Suppose the principal justifies almost surely, $\Pr(\mathcal{I}_C) = 1$.

Then the expected communication costs are $\kappa\mathbb{E}(\beta(I)) = \kappa$ and it just remains to minimize the wage costs. Yet it is possible to implement payments $w_e^*(\mu)$ defined in Program A* by the following contract:

$$\bar{W}(m_P, m_A) = \begin{cases} w_e^*\big(\int m_P(t)\mathrm{d}t\big) & \text{if } m_P(t) = 1 \text{ for all } t \in m_A \\ w_e^*(1) + \kappa & \text{if } m_P(t) \neq 1 \text{ for some } t \in m_A \end{cases}$$

and $\bar{c}(m_P, m_A) = w_e^*(\int m_P(t)\mathrm{d}t)$. For the same reasons as in the proof of Lemma 2, the additional incentive constraints (IC_P) and (IC_{m_A}) are satisfied. Furthermore, the equilibrium wage is $w_e^*(\mu)$, as the parties' messages agree in equilibrium.

It is possible, however, to implement a certain work effort e of the agent even cheaper by partial justification. For this purpose, modify the contract $\bar{\mathcal{W}}$ to

$$\bar{W}'(m_P, m_A) = \begin{cases} w_e^*\big(\int m_P(t)\mathrm{d}t\big) & \text{if } \int m_P(t)\mathrm{d}t < \delta \text{ and } m_P(t) = 1 \text{ for all } t \in m_A \\ w_e^*(1) + \kappa & \text{if } \int m_P(t)\mathrm{d}t < \delta \text{ and } m_P(t) \neq 1 \text{ for some } t \in m_A \\ w_e^*(\delta) + \kappa & \text{if } \int m_P(t)\mathrm{d}t \geq \delta \end{cases}$$

with a $\delta < 1$, such that

$$f^H(\delta) - f^L(\delta) \geq 0 \quad \text{and} \quad w_e^*(\delta) + \kappa \geq w_e^*(1). \tag{4}$$

Lemma 1 proves that $w_e^*(\mu)$ is almost surely continuous and any discontinuity is removable. Consequently, there exists a continuous function that almost surely equals $w_e^*(\mu)$. Replacing $w_e^*(\mu)$ by that function in the definition of $\bar{W}'$ also yields a solution to Program A. This procedure guarantees that the conditions (4) on δ are feasible.

In the contract $\bar{\mathcal{W}}'$, for $n > \bar{N}$, the principal justifies all evaluations except the highest ones and the justification set is

$$\mathcal{I}_C = \left\{ I \in \mathcal{I} \,\middle|\, \int I(t)\mathrm{d}t \leq \delta \right\}.$$

If the principal's information indicates a very good performance, $\mu > \delta$, justification would increase her communication costs as the wage costs remain unchanged. In addition, the conditions in (4) guarantee that constraints (PC) and (IC_A) are still satisfied by choosing the agent's wage appropriately. Therefore contract $\bar{\mathcal{W}}'$ implements effort e of the agent and is cheaper than the contract $\bar{\mathcal{W}}$. This shows that the principal will not justify all evaluations for positive κ. □

Without loss of generality, I can concentrate on truthful messages.

Lemma 3. *For every contract $\mathcal{W}$ there is a contract $\mathcal{W}'$, such that*

- *$\mathcal{W}'$ gives the agent incentives to implement the same effort e as in $\mathcal{W}$,*
- *$\mathcal{W}'$ has (weakly) lower costs for the principal than $\mathcal{W}$ and*
- *$\mathcal{W}'$ gives the agent and the principal incentives to send truthful messages.*

In addition, contract $\mathcal{W}'$ has the following structure

$$c'(m_P, m_A) = \begin{cases} c(m_P) & \text{if } m_P \in \mathcal{I}'_C \\ \bar{w} & \text{if } m_P \notin \mathcal{I}'_C \end{cases} \tag{5}$$

$$W'(m_P, m_A) = \begin{cases} w(m_P) & \text{if } m_P \in \mathcal{I}'_C \text{ and } m_P(t) = 1 \text{ for all } t \in m_A \\ \bar{w} + \kappa & \text{if } m_P \in \mathcal{I}'_C \text{ and } m_P(t) \neq 1 \text{ for some } t \in m_A \\ \bar{w} & \text{if } m_P \notin \mathcal{I}'_C \end{cases} \tag{6}$$

with the justification set $\mathcal{I}'_C = \{I \in \mathcal{I} | \beta'(I) = 1\}$.

Proof: The proof consists of four parts. The first part characterizes the equilibrium utilities in contract $\mathcal{W}$. The second part guarantees that the contract $\mathcal{W}'$ fulfills the incentive compatibilities for the messages (IC_P) and (IC_{m_A}). The third part analyzes the agent's incentive compatibility (IC_A). The fourth part ensures that the new contract $\mathcal{W}'$ satisfies the agent's incentive compatibility and participation constraint.

Step 1 Denote the expected payment conditional on information I given equilibrium strategies in contract $\mathcal{W}$ by $w(I)$ for the principal and the corresponding certainty equivalent by $c(I)$ for the agent. If in contract $\mathcal{W}$ the principal provides no justification in equilibrium, $I \notin \mathcal{I}_C$, the agent cannot verify the evaluation. The reason is that the cardinality of $\mathcal{I}$ and R differ. According to Cantor's theorem, the set of all subsets of a set A has a strictly greater cardinality than the set A itself. Here, the cardinality of $\{0,1\}^T$, i.e., the set of all functions $T \to \{0,1\}$, equals the cardinality of the power set of T and is bigger than the cardinality of T and the one of R. Hence, it is impossible to encode the information I in R. No matter which information the principal tells in her message $m_P \in R$, the probability that it matches the agent's information S is 0. Consequently, the agent cannot verify the principal's message. Therefore the principal's payments have to be constant or $w(I) = w(I')$ for all $I, I' \notin \mathcal{I}_C$. Moreover, they have to be higher than the principal's payments in the justification set including communication costs.

$$w(I) \geq \kappa + \sup_{I' \in \mathcal{I}_C} w(I') \qquad \forall I \notin \mathcal{I}_C$$

Otherwise, the principal would not provide justification and act as if $I \notin \mathcal{I}_C$, because the agent could not observe this deviation.

Suppose there are $I, I' \in \mathcal{I}$ with $I(t) \leq I'(t)$ for all $t \in T$ and $w(I) > w(I')$. Then the principal could decrease her wage payments by reporting more successful categories. The agent cannot detect such a deviation. Therefore $w(I) \leq w(I')$ for all $I, I' \in \mathcal{I}$ with $I(t) \leq I'(t)$ for all $t \in T$. Finally, $w(I) = w(I')$ for all $I, I' \in \mathcal{I}$ with $I = I'$ a.s.

Step 2 Consider the contract $\mathcal{W}'$ given by equations (5) and (6) with $\mathcal{I}'_C = \mathcal{I}_C \setminus R$ and $\bar{w} = w(I)$ for an $I \notin \mathcal{I}_C$.[14] The agent's wage does not depend on his message.

[14] This requires identifing a message $m_P \in R$ with the step function: $T \to \{0,1\}$ with 1 for $t \leq m_p$ and 0 otherwise. Additionally, if $\mathcal{I}_C = \mathcal{I}$, set $\bar{w} = \kappa + \sup_{\hat{I} \in \mathcal{I}_C} w(\hat{I})$.

Therefore he is indifferent between sending any message and the incentive compatibility for his message is satisfied. If the principal should justify, $I \in \mathcal{I}_C$, any disagreement in the messages shows a deviation by the principal and the payment $W'(I,S)$ with $I \in \mathcal{I}_C$ and $I(t) \neq 1$ for some $t \in S$ matters only for the right-hand side of the principal's incentive compatibility (IC_P). Therefore I can increase this payment to satisfy (IC_P) without affecting any other constraint or the objective function. Accordingly, there will be a penalty for $I \in \mathcal{I}_C$ and $I(t) \neq 1$ for some $t \in S$. By setting the principal's payment in this case to

$$W(I,S) = \bar{w} + \kappa$$

there is a $\bar{N} \in \mathbb{N}$ such that for $n > \bar{N}$ the principal will never deviate to another message in the justification set $\mathcal{I}_C$. The reason is that a deviation at least weakly increases payments as

$$\bar{w} + \kappa > \int W'(\hat{I},S)\mathrm{d}P(S|I,e) \geq \bar{w} > w(I) = \int W'(I,S)\mathrm{d}P(S|I,e)$$

for all $I, \hat{I} \in \mathcal{I}_C$ with $\int_{\{t \in T | \hat{I}(t) \neq I(t) \text{ and } \hat{I}(t)=0\}} 1\mathrm{d}t > 0$. The reasoning is the same as in the proof of Proposition 1. For large deviations the probability of a mismatch in the messages is sufficiently close to 1. For small deviations increasing the number of categories in the agent's sample increases the density of one of these categories being in the deviation set. On the margin, the density matters and not the probability. Deviations with $\int_{\{t \in T | \hat{I}(t) \neq I(t) \text{ and } \hat{I}(t)=1\}} 1\mathrm{d}t > 0$ might remain undetected, but increase the wage as $w(I) \leq w(\hat{I})$. Hence, a deviation is unprofitable.

Step 3 Denote by $M(\mu) = \{I \in \mathcal{I} | \int I(t)\mathrm{d}t = \mu\}$ the set of all information with an average μ. For the agent's incentives only the expected wage in $M(\mu)$ matters, because the agent's sample, S, does not depend on her effort choice and the average of the principal's information is a sufficient statistics for the agent's effort, $\Pr(I,S|e,\int I(t)\mathrm{d}t) = \Pr(I,S|\int I(t)\mathrm{d}t)$. In addition, Lemma 1 shows that the first-order approach is valid here. Therefore, the agent's incentive compatibility (IC_A) equals

$$\int \left(\int u(c(I))\mathrm{d}P\left(I,S\,|e,I \in M(\mu)\right) \right) \left(f^H(\mu) - f^L(\mu)\right)\mathrm{d}\mu = d'(e) \tag{7}$$

Consequently, the agent's incentives remain unchanged in contract $\mathcal{W}'$ if

$$\int u(c(I))\mathrm{d}P(I,S|e,I \in M(\mu)) = \int u\big(c'(\theta(I,\beta(I)),S)\big)\mathrm{d}P(I,S|e,I \in M(\mu)) \tag{8}$$

for all $\mu \in [0,1]$. By the definition of c' in (5), the right-hand side equals

$$\int \beta'(I)u\big(c(I)\big)\mathrm{d}P(I,S|e,I \in M(\mu)) + u(\bar{w}) \int 1 - \beta'(I)\mathrm{d}P(I,S|e,I \in M(\mu))$$

Hence, the equality in (8) is not guaranteed for all μ, as $c(I) \leq c'(I,m_A) = \bar{w}$ for $I \notin \mathcal{I}'_C$.

Step 4 If (8) is not satisfied, it is necessary to reduce the expected wage of the agent.

For this purpose, let the principal justify every evaluation $I \notin \mathcal{I}_C \cup R$ with $c(I) \leq \bar{w} - \kappa$. Denote the set of these I by $\mathcal{I}'$. Now set $\mathcal{I}'_C = \mathcal{I}'_C \cup \mathcal{I}'$, $c'(I, m_A) = c(I)$ and

$$W'(I, m_A) = \begin{cases} c(I) & \text{if } I(t) = 1 \quad \forall t \in m_A \\ \bar{w} + \kappa & \text{if } \exists t \in m_A \colon I(t) \neq 1 \end{cases} \qquad \forall I \in \mathcal{I}', \forall m_A \in T^n.$$

Thus, for any remaining evaluations that are not justified, $I \notin \mathcal{I}'_C \cup R$, the agent's wage is $\bar{w} - \kappa < c(I) \leq \bar{w}$ in contract $\mathcal{W}$. Finally, increase the justification set and make the principal justify a fraction α of the evaluations $M(\mu) \setminus \mathcal{I}'_C$, such that

$$\int u(c(I))\mathrm{d}P\left(I, S \,|e, I \in M(\mu)\right) = \int \beta'(I) u(c(I))\mathrm{d}P\left(I, S \,|e, I \in M(\mu)\right) + \\ + \left(\alpha u(\bar{w} - \kappa) + (1 - \alpha) u(\bar{w})\right) \int 1 - \beta'(I)\mathrm{d}P\left(I, S \,|e, I \in M(\mu)\right)$$

As the agent's wage is $\bar{w} - \kappa < c(I) \leq \bar{w}$ for all $I \notin \mathcal{I}'_C \cup R$, it is possible to find such an $\alpha \in [0, 1]$. Set $\alpha = 1$ if α is not uniquely determined, as $M(\mu) \setminus \mathcal{I}'_C$ has mass 0. Denote the addition evaluations that are justified by $\mathcal{I}''$. To make justification optimal, adjust contract $\mathcal{W}'$ by $\mathcal{I}'_C = \mathcal{I}'_C \cup (\mathcal{I}'' \setminus R)$, $c'(I, m_A) = \bar{w} - \kappa$ and

$$W'(I, m_A) = \begin{cases} \bar{w} - \kappa & \text{if } I(t) = 1 \quad \forall t \in m_A \\ \bar{w} + \kappa & \text{if } \exists t \in m_A \colon I(t) \neq 1 \end{cases} \qquad \forall I \in (\mathcal{I}'' \setminus R), \forall m_A \in T^n.$$

Repeat these steps for every μ. Then the agent's incentives in the new contract $\mathcal{W}'$ are the same as in contract $\mathcal{W}$. In addition, the agent's participation constraint is also satisfied in contract $\mathcal{W}'$, so that contract $\mathcal{W}'$ implements effort e at (weakly) lower costs than contract $\mathcal{W}$. □

Without loss of generality, I can concentrate on contracts in which the agent's wage depends only on the average of the principal's information.

Lemma 4. *For every contract $\mathcal{W}$ there is a contract $\mathcal{W}'$ implementing the same effort e at the same costs as $\mathcal{W}$. Moreover, in contract $\mathcal{W}'$ the justification choice β and the agent's wages just depend on the average of the principal's information, i.e., $\beta(m_P) = \beta(m'_P)$ and $c'(m_P, m_A) = c'(m'_P, m_A)$ for all $m_A \in T^n, m_P, m'_P \in \mathcal{I}_C$ with $\int m_P(t)\mathrm{d}t = \int m'_P(t)\mathrm{d}t$.*

Proof: Lemma 3 shows that the agent's wage does not depend on his message m_A. In addition, according to equation (7), the agent's incentive compatibility (IC_A) depends only on the expected utility of the agent given the average of the principal's information. The same is valid for the agent's participation constraint in Program A. Therefore it is possible without violating these constraints to set $c'(m_P, m_A) = \tilde{c}\left(\int m_P(t)\mathrm{d}t\right)$ for all $m_P \in \mathcal{I}_C$ and $m_A \in T^n$ with $\tilde{c}(\mu)$ defined by

$$u\left(\tilde{c}(\mu)\right) = \int u\big(c(\hat{I}, S)\big)\mathrm{d}P\left(\hat{I}, S \,\middle|\, e, \hat{I} \in M(\mu) \cap \mathcal{I}_C\right)$$

and $M(\mu) = \{I \in \mathcal{I} | \int I(t)\mathrm{d}t = \mu\}$. This modification reduces at least weakly the

expected wage, because the agent is risk-averse,

$$\tilde{c}\left(\mu\right) \leq \int c(\hat{I},S)\mathrm{d}P\left(\hat{I},S\left|e,\hat{I}\in M\left(\mu\right)\cap\mathcal{I}_C\right.\right). \tag{9}$$

Yet, the agent's new wage $c'(m_P,m_A)=\tilde{c}(\mu)$ might be higher than the principal's payment $W(m_P,m_A)$ for some m_A and m_P. To make the contract feasible and satisfy constraint (1) in Program A, set

$$W'(m_P,m_A)=\int W(\hat{I},S)\mathrm{d}P\left(\hat{I},S\left|e,\hat{I}\in M\left(\mu\right)\cap\mathcal{I}_C\right.\right) \qquad \forall m_P\in\mathcal{I}_C,\forall m_A\in T^n$$

with $\mu=\int m_P(t)\mathrm{d}t$. This ensures that $W'(m_P,m_A)\geq\tilde{c}(\mu)$ for all m_P,m_A, because $W(m_P,m_A)\geq c(m_P,m_A)$ in contract $\mathcal{W}$ and the new wage, $\tilde{c}(\mu)$, is lower than the previous expected wage according to equation (9). In addition, the principal's expected payments in the new contract $\mathcal{W}'$ are the same as in contract $\mathcal{W}$, as

$$\int W'_{eq}(I,S)\mathrm{d}P(I,S|e,I\in M\left(\mu\right))=\int W_{eq}(I,S)\mathrm{d}P(I,S|e,I\in M\left(\mu\right)) \quad \forall\mu.$$

Consequently, there exists a contract that gives the principal and the agent the same utility as the old contract $\mathcal{W}$, but makes the agent's utility depend only on the average of the principal's information, i.e., $c(m_P,m_A)=\tilde{c}(\mu)$ for all m_A and all $m_P\in\mathcal{I}_C$. It remains to ensure constraint (IC_P) in contract $\mathcal{W}'$. For $I\in\mathcal{I}_C$ in the justification set, $W'(I,S)$ with $I(t)\neq 1$ for some $t\in S$ matters only on the right-hand side of constraint (IC_P). Therefore increasing $W'(I,S)$ to $\kappa+\sup_{\bar{I},\bar{S}}W(\bar{I},\bar{S})$ does not affect the objective function or other constraints, but gives the principal incentives to communicate truthfully. Additionally, the wage in the justification set is lower than outside this set including the communication costs, $W'(I,S)+\kappa<W'_{eq}(\hat{I},\hat{S})$ for all $I\in\mathcal{I}_C,\hat{I}\notin\mathcal{I}_C$ and all $S,\hat{S}\in T^n$, because contract $\mathcal{W}$ meets this condition according to Lemma 3. This guarantees that there is a $\bar{N}\in\mathbb{N}$ such that for all $n>\bar{N}$ any deviation in the justification choice, $\beta(I)$, and/or the message m_P makes the principal worse off. Therefore also the new contract $\mathcal{W}'$ satisfies the principal's incentive compatibility.

It remains to prove that the justification choice does not change within the set $M(\mu)$ for any μ.

1. Suppose there is a μ such that $\hat{c}\left(\mu\right)\leq\bar{w}-\kappa$ with $\bar{w}$ the wage outside the justification set according to Lemma 3 and $\hat{c}$ defined by

$$u\left(\hat{c}\left(\mu\right)\right)=\int u\big(c(\hat{I},S)\big)\mathrm{d}P\left(\hat{I},S\left|e,\hat{I}\in M\left(\mu\right)\right.\right).$$

Then it is possible to justify all $I\in M(\mu)$ by setting $\mathcal{I}'_C=\mathcal{I}'_C\cup M(\mu)$ and by adapting the wage as in the first part of the proof.

2. If, on the other hand, $\hat{c}(\mu)=\bar{w}$, the principal does not provide justification for

$I \in M(\mu)$ by setting $\mathcal{I}'_C = \mathcal{I}'_C \setminus M(\mu)$ and $W(I, m_A) = c(I, m_A) = \bar{w}$ for all $m_A \in T^n$.

3. Finally, in the last case $\bar{w} - \kappa < \hat{c}(\mu) < \bar{w}$. Denote by A the set of all μ with this property:

$$A = \{\mu \in [0,1] |\, \bar{w} - \kappa < \hat{c}(\mu) < \bar{w}\}.$$

If A has no mass, $\int_A f(\mu|e)\mathrm{d}\mu = 0$, set $\mathcal{I}'_C = \mathcal{I}'_C \cup M(\mu)$ and $\hat{c}(\mu) = \bar{w} - \kappa$ for all $\mu \in A$ and adapt the wage as in the first part of the proof. If $\int_A f(\mu|e)\mathrm{d}\mu > 0$, there exists a unique δ, such that justifying only evaluations below δ does not change the agent's expected utility, i.e.,

$$\int_A u(\hat{c}(\mu))\mathrm{d}F(\mu|e) = u(\bar{w} - \kappa)\int_{A^1} \mathrm{d}F(\mu|e) + u(\bar{w})\int_{A^2} \mathrm{d}F(\mu|e)$$

with the sets $A^1 = \{\mu \in A | \mu \leq \delta\}$ and $A^2 = \{\mu \in A | \mu > \delta\}$. For this purpose, modify the contract to

$$\beta'(m_P) = \begin{cases} 1 & \text{if } \bar{\mu} \in A^1 \\ 0 & \text{if } \bar{\mu} \in A^2 \end{cases} \qquad c'(m_P, m_A) = \begin{cases} \bar{w} - \kappa & \text{if } \bar{\mu} \in A^1 \\ \bar{w} & \text{if } \bar{\mu} \in A^2 \end{cases}$$

$$W'(m_P, m_A) = \begin{cases} \bar{w} & \text{if } \bar{\mu} \in A^2 \\ \bar{w} - \kappa & \text{if } \bar{\mu} \in A^1 \text{ and } m_P(t) = 1 \text{ for all } t \in m_A \\ \bar{w} + \kappa & \text{if } \bar{\mu} \in A^1 \text{ and } m_P(t) \neq 1 \text{ for some } t \in m_A \end{cases}$$

for all $m_P \in M(\bar{\mu})$, all $m_A \in T^n$ and all $\bar{\mu} \in A$. As the agent is risk-averse and $W(m_P, m_A) \geq c(m_P, m_A) \in \big(\{\bar{w}\} \cup (0, \bar{w} - \kappa]\big)$, this reduces the principal's payments:

$$\int_A \left(\int W(\hat{I}, S)\mathrm{d}P\left(\hat{I}, S \,\middle|\, e, \hat{I} \in M(\mu)\right)\right) \mathrm{d}F(\mu|e) \geq (\bar{w} - \kappa)\int_{A^1} \mathrm{d}F(\mu|e) + \bar{w}\int_{A^2} \mathrm{d}F(\mu|e)$$

In addition, the left-hand side of the agent's incentive compatibility (7) increases:

$$\begin{aligned} &\int_{A^1} \big(u(\bar{w} - \kappa) - u(\hat{c}(\mu))\big)\Delta^f(\mu)\mathrm{d}\mu + \int_{A^2} \big(u(\bar{w}) - u(\hat{c}(\mu))\big)\Delta^f(\mu)\mathrm{d}\mu = \\ =&\int_{A^1} \underbrace{\big(u(\bar{w} - \kappa) - u(\hat{c}(\mu))\big)}_{<0} \frac{\Delta^f(\mu)f(\mu|e)}{f(\mu|e)}\mathrm{d}\mu + \int_{A^2} \underbrace{\big(u(\bar{w}) - u(\hat{c}(\mu))\big)}_{>0} \frac{\Delta^f(\mu)f(\mu|e)}{f(\mu|e)}\mathrm{d}\mu > \\ >&\frac{f^H(\delta) - f^L(\delta)}{f(\delta|e)} \left(\int_{A^1} u(\bar{w} - \kappa)\mathrm{d}F(\mu|e) + \int_{A^2} u(\bar{w})\mathrm{d}F(\mu|e) - \int_A u(\hat{c}(\mu))\mathrm{d}F(\mu|e)\right) = 0 \end{aligned}$$

with $\Delta^f(\mu) = f^H(\mu) - f^L(\mu)$. The monotone likelihood ratio property ensures the strict inequality and the constant value of the participation constraint yields the final equality. This shows that the modified contract satisfies the agent's incentive compatibility (IC_A).

Consequently, contract $\mathcal{W}'$ satisfies all the constraints of Program A and the justification decision just depends on the average of the principal's information. □

Proposition 2 characterizes the optimal contract.

Proof of Proposition 2: Lemma 3 determines the basic structure of the optimal contract and this proof uses the notation introduced there. Lemma 4 shows that the justification decision and the agent's wage just depend on the average of the principal's message. It remains to calculate the agent's wage. First, I prove that there are no payments to third parties in equilibrium. For this purpose, simplify Program A according to Lemma 3 and 4 to

$$\inf \int w(I) + \kappa\beta(I)\mathrm{d}P(I,S|e), \tag{B}$$

$$\text{subject to} \int u(c(\mu))\mathrm{d}F(\mu|e) - d(e) \geq \bar{u}, \tag{PC}$$

$$\int u(c(\mu))(f^H(\mu) - f^L(\mu))\mathrm{d}\mu = d'(e), \tag{IC$_A$}$$

$$\bar{w} \geq w(I) + \beta(I)\kappa \geq (1-\beta(I))\bar{w} \qquad \forall I \in \mathcal{I}, \tag{IC$_P$}$$

$$w(I) \geq c\left(\int I(t)\mathrm{d}t\right) \qquad \forall I \in \mathcal{I}. \tag{10}$$

(IC$_P$) ensures that the wage is constant outside the justification set. Furthermore, the condition requires that the principal's costs in the justification set, accounting for the communication costs, is lower than outside this set. According to Lemma 3 this is equivalent to condition (IC$_P$) in Program A. Condition (10) is the equivalent to (1) in the initial program and guarantees that payments to third parties are nonnegative. Assume to the contrary that in the optimal contract there is an $I \in \mathcal{I}$, such that $w(I) > c\left(\int I(t)\mathrm{d}t\right)$. Then $\beta(I) = 1$, because Lemma 3 and 4 show that $c\left(\int I(t)\mathrm{d}t\right) = \bar{w} = w(I)$ for $I \notin \mathcal{I}_C$. In addition, it is possible to (weakly) decrease the objective function without violating any constraint by setting $w(I) = c\left(\int I(t)\mathrm{d}t\right)$. Therefore the principal's equilibrium payment just depends on the average of her message and $w(I) = c\left(\int I(t)\mathrm{d}t\right)$ for all $I \in \mathcal{I}$. This results in payments to a third party of

$$W(m_P, m_A) - c\left(\int m_P(t)\mathrm{d}t\right) = \begin{cases} \bar{w} - w(m_P) + \kappa & \text{if } m_P \in \mathcal{I}_C \text{ and } \exists t \in S\colon I(t) \neq 1 \\ 0 & \text{otherwise.} \end{cases}$$

As the principal's equilibrium payments and her justification decision just depend on the average of the principal's information, writing them as functions of the average μ instead of the information I simplifies Program B to:

$$C^c(e) = \inf \int w(\mu) + \kappa\beta(\mu)\mathrm{d}F(\mu|e), \tag{C}$$

$$\text{subject to} \int u\big(w(\mu)\big)\mathrm{d}F(\mu|e) - d(e) \geq \bar{u}, \tag{PC}$$

$$\int u\big(w(\mu)\big)(f^H(\mu) - f^L(\mu))\mathrm{d}\mu = d'(e), \tag{IC$_A$}$$

$$(1-\beta(\mu))\,(w(\mu) - w) = 0 \qquad \forall \mu, \tag{11}$$

$$w(\mu) + \kappa\beta(\mu) \leq w \qquad \forall \mu. \tag{12}$$

As before, the objective is to minimize the expected costs, here communication costs and the principal's payments. The agent's participation constraint (PC) and incentive compatibility (IC_A) remain unchanged. Constraints (11) and (12) replace condition (IC_P) ensuring a constant wage outside the communication set. □

A.3 The Optimal Communication Pattern

A threshold strategy is optimal, with justification below a threshold δ and no justification above δ.

Proof of Proposition 3: Proposition 1 proves that complete feedback is suboptimal and $\Pr(\mathcal{I}_C) < 1$. In order to show that only bad evaluations will be justified, assume to the contrary that in the optimal contract there is a μ^* with the following properties: With positive probability the principal justifies evaluations μ above μ^* and with positive probability the principal does not justify evaluations μ below μ^*.
Denote by $A^K = \{\mu \in [0,1] | \mu \geq \mu^* \text{ and } \beta(\mu) = 1\}$ justified evaluations above μ^* and by $A^N = \{\mu \in [0,1] | \mu \leq \mu^* \text{ and } \beta(\mu) = 0\}$ evaluations without justification below μ^*. By assumption, $\Pr(A^K), \Pr(A^N) > 0$. Then rewrite program C to simplify the next steps of the proof. Change constraint (11) to

$$f(\mu|e)\,(1-\beta(\mu))\,(w(\mu)-w)^2 \leq 0 \tag{13}$$

for all μ. (13) is equivalent to (11). This condition guarantees that the wage is constant outside the justification set, in particular, in A^N. In the justification set and thus in the subset A^K, condition (13) is trivially fulfilled, as $1-\beta(\mu) = 0$. Furthermore, multiply constraint (12) by $f(\mu|e)$ to obtain

$$f(\mu|e)(w(\mu) + \kappa\beta(\mu) - w) \leq 0 \tag{14}$$

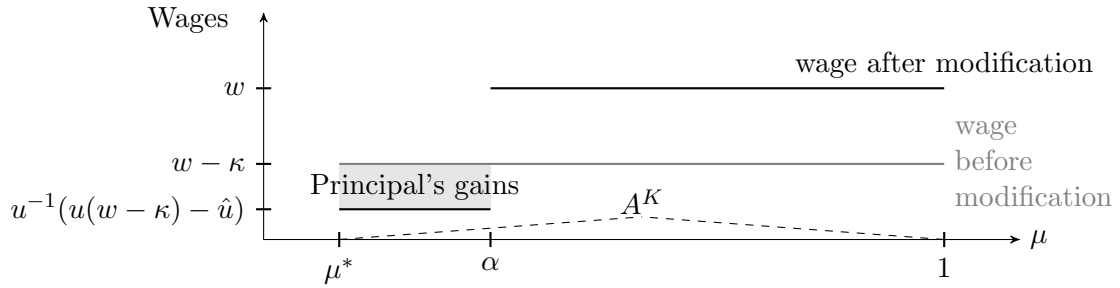
for all μ. (14) is equivalent to (12) and guarantees that it is optimal to justify an evaluation if justification is required by the contract, i.e., $\beta(\mu) = 1$. Together, both constraints make justification optimal, if and only if $\beta(\mu) = 1$. In addition, they ensure that the wage has to be lower in A^K than in A^N, as

$$w(\mu) < w = w(\mu') \qquad \forall \mu \in A^K, \mu' \in A^N. \tag{15}$$

The inequality follows from condition (14), while the equality is given by condition (13). Define λ_1, λ_2, $\nu_1(\mu)$ and $\nu_2(\mu)$ to be the Lagrange multipliers of the constraints (PC), (IC_A), (13) and (14) respectively. Pointwise optimization[15] with respect to $w(\mu)$ yields

$$\begin{aligned} &1 - u'(w(\mu))\left(\lambda_1 + \lambda_2 \frac{f^H(\mu) - f^L(\mu)}{f(\mu|e)}\right) + \nu_1(\mu)(1-\beta(\mu))2(w(\mu)-w) + \nu_2(\mu) = 0 \\ &\Leftrightarrow 1 + \nu_2(\mu) = u'(w(\mu))\left(\lambda_1 + \lambda_2 \frac{f^H(\mu) - f^L(\mu)}{f(\mu|e)}\right), \quad \text{a.s.} \end{aligned} \tag{16}$$

[15] Cf. footnote 13 for the generality of this method.



In the figure the set A^K is an interval to ease exposition. In general, A^K need not be convex. The gray rectangle captures the reduction in expected costs for the principal.

Figure 5: The Modified Wage Contract

because constraint (13) guarantees $(1-\beta(\mu))(w(\mu)-w)=0$ for all μ. The monotone likelihood ratio property ensures that

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

is increasing in μ. Together with (15) this proves that the right-hand side of equation (16) is higher for $\mu\in A^K$ than for $\mu'\in A^N$, because $\mu'<\mu$. To match this increase, $\nu_2(\mu)$ almost surely has to be positive in A^K. Hence, $w(\mu)=w-\kappa$ almost surely for $\mu\in A^K$.

The next step modifies the contract to implement effort e cheaper than before. For this purpose, determine the median α of A^K, such that

$$\int_{A_1^K} f(\mu|e)\mathrm{d}\mu=\int_{A_2^K} f(\mu|e)\mathrm{d}\mu \tag{17}$$

for $A_1^K=A^K\cap[\mu^*,\alpha)$ and $A_2^K=A^K\cap[\alpha,1]$. In the new contract, the principal does not justify elements of A^K above α, $\beta(\mu)=0$, while below α, she provides justification. Then, I add (resp. subtract) the monetary equivalent of $\hat{u}=u(w)-u(w-\kappa)$ to the wage, so that

$$w'(\mu)=\begin{cases}u^{-1}(u(w-\kappa)+\hat{u})=w & \text{for } \mu\in A_2^K\\ u^{-1}(u(w-\kappa)-\hat{u})=u^{-1}(2u(w-\kappa)-u(w)) & \text{for } \mu\in A_1^K.\end{cases}$$

This means that the wage increases by κ for values in A^K above α and decreases by an amount adjusted for the changes in marginal utility below α. Figure 5 illustrates this modification of the wage. The modified wage still satisfies the condition (PC). Additionally, the left-hand side of condition (IC_A) is now bigger than the marginal cost of effort, $d'(e)$, because the left-hand side of condition (IC_A) increases by

$$\begin{aligned}&\hat{u}\left(-\int_{A_1^K} f^H(\mu)-f^L(\mu)\mathrm{d}\mu+\int_{A_2^K} f^H(\mu)-f^L(\mu)\mathrm{d}\mu\right)=\\ &=\hat{u}\left(-\int_{A_1^K}\frac{f^H(\mu)-f^L(\mu)}{f(\mu|e)}f(\mu|e)\mathrm{d}\mu+\int_{A_2^K}\frac{f^H(\mu)-f^L(\mu)}{f(\mu|e)}f(\mu|e)\mathrm{d}\mu\right)>\\ &>\hat{u}\frac{f^H(\alpha)-f^L(\alpha)}{f(\alpha|e)}\left(-\int_{A_1^K} f(\mu|e)\mathrm{d}\mu+\int_{A_2^K} f(\mu|e)\mathrm{d}\mu\right)=0.\end{aligned}$$

The last equality follows from the definition of α in (17). The main inequality follows from the increasing likelihood ratio. By reducing the variance of the wage payments until condition (IC_A) holds with equality, condition (PC) becomes slack. This allows decreasing the wage and shows a contradiction to the existence of μ^*. Therefore the optimal contract does not require the principal to justify good evaluations, as the agent does not suspect a distortion of these evaluations. On the other hand, the principal has to justify bad evaluations. The set A^K can be replaced by any set of justified evaluations and a wage of $w - \kappa$. Thus, in the optimal contract any set with these properties has no mass and, hence, does not influence the optimal contract. Therefore, the communication pattern can be adjusted to a threshold rule. Combining these results with Program C, allows me to prove analogously to Lemma 1 that $w(\mu)$ is strictly increasing in $\mu \in [0, \delta]$. □

Proof of Proposition 4: Consider the relaxed problem:

$$\inf_{w(\mu),w,\delta} \int_0^\delta w(\mu) + \kappa \mathrm{d}F(\mu|e) + (1 - F(\delta|e))w, \tag{D}$$

$$\text{subject to} \int_0^\delta u\big(w(\mu)\big)\mathrm{d}F(\mu|e) + (1 - F(\delta|e))u(w) - d(e) \geq \bar{u}, \tag{PC}$$

$$\int_0^\delta u\big(w(\mu)\big)(f^H(\mu) - f^L(\mu))\mathrm{d}\mu + u(w)\int_\delta^1 f^H(\mu) - f^L(\mu)\mathrm{d}\mu = d'(e) \quad (\text{IC}_A)$$

without constraint (12). The solution to this problem gives the agent the necessary incentives at minimal costs. Yet this solution might for some μ violate the principal's incentive compatibility, namely, $w \geq \kappa + w(\mu)$. Nevertheless, I show that the solution to this problem always satisfies constraint (12). Therefore the solution to the relaxed problem is also a solution to the initial program.

Denote the Lagrange multiplier for the participation constraint by λ_1 and the one for the agent's incentive compatibility by λ_2. Program D yields three first-order conditions:

$$f(\mu|e) - \lambda_1 u'(w(\mu))f(\mu|e) - \lambda_2 u'(w(\mu))(f^H(\mu) - f^L(\mu)) = 0 \quad \text{a.s. for } \mu$$

$$(1 - F(\delta|e)) - \lambda_1 u'(w)(1 - F(\delta|e)) - \lambda_2 u'(w)\int_\delta^1 f^H(\mu) - f^L(\mu)\mathrm{d}\mu = 0$$

$$f(\delta|e)(w(\delta) + \kappa - w) - (u(w(\delta)) - u(w))\Big(\lambda_1 f(\delta|e) + \lambda_2(f^H(\delta) - f^L(\delta))\Big) = 0$$

The last equation ensures $w \neq w(\delta)$. Rearranging yields

$$\lambda_1 + \lambda_2 \frac{f^H(\mu) - f^L(\mu)}{f(\mu|e)} = \frac{1}{u'(w(\mu))} \quad \text{a.s. for } \mu \tag{18}$$

$$\lambda_1 + \lambda_2 \frac{\int_\delta^1 f^H(\mu) - f^L(\mu)\mathrm{d}\mu}{1 - F(\delta|e)} = \frac{1}{u'(w)} \tag{19}$$

$$\lambda_1 + \lambda_2 \frac{f^H(\delta) - f^L(\delta)}{f(\delta|e)} = \frac{w - w(\delta) - \kappa}{u(w) - u(w(\delta))} \tag{20}$$

The monotone likelihood ratio guarantees that

$$\frac{\int_\delta^1 f^H(\mu) - f^L(\mu)\mathrm{d}\mu}{1-F(\delta|e)} = \frac{\int_\delta^1 \frac{f^H(\mu)-f^L(\mu)}{f(\mu|e)} f(\mu|e)\mathrm{d}\mu}{\int_\delta^1 f(\mu|e)\mathrm{d}\mu} > \frac{\frac{f^H(\delta)-f^L(\delta)}{f(\delta|e)} \int_\delta^1 f(\delta|e)\mathrm{d}\mu}{\int_\delta^1 f(\delta|e)\mathrm{d}\mu} = \frac{f^H(\delta)-f^L(\delta)}{f(\delta|e)}.$$

Additionally, as before, λ_1 and λ_2 are positive. Then (18) and (19) imply $w > w(\mu)$ a.s. for $\mu \in [0, \delta]$. Comparing (18) for $\mu = \delta$ and (20) results in

$$\frac{1}{u'(w(\delta))} = \frac{w - w(\delta) - \kappa}{u(w) - u(w(\delta))} \tag{21}$$

Assume to the contrary that $w \leq w(\delta) + \kappa$. Then the left-hand side is non-positive, while the right-hand side is positive. Therefore $w > w(\delta) + \kappa$ and any solution to the relaxed problem also satisfies condition (12). Consequently, (18), (19), and (20) in combination with the participation constraint and the agent's incentive compatibility determine the optimal contract. □

Proof of Proposition 5: The proof proceeds along the following lines. If $e > 0$ and the principal justifies almost surely no evaluation, e.g., due to high communication costs κ, the optimal contract does not exist.[16] Yet, for every ϵ there is a contract whose costs are at most ϵ higher than the infimum costs to implement effort $e > 0$ by the agent. Finally, the proof shows that condition (2) ensures that the optimal contract provides justification with positive probability.

In the absence of justification, the principal's payments have to be constant according to Lemma 3, as it is impossible to verify her message m_P. Denote her payments by $\bar{w}$. The next steps calculate the infimum costs to implement positive effort $e > 0$ of the agent. For this purpose, the set of feasible contracts is reduced step by step, as I show that contracts with specific characteristics are suboptimal. First consider a contract with $c(\mu) < \bar{w}$ almost surely. In this case, it is possible to reduce the principal's payments $\bar{w}$ without violating any constraint. Therefore I only have to take contracts into account with a positive probability for $\{\mu \in [0,1] | c(\mu) = \bar{w}\}$.

Second, consider an optimal contract $\mathcal{W}$ with a μ^*, such that there are payments to a third party for $\mu > \mu^*$ with positive probability, but with positive probability there are no such payments for $\mu \leq \mu^*$. Denote the corresponding sets by $A^S = \{\mu \in [0,1] | \mu > \mu^*$ and $c(\mu) < \bar{w}\}$ with third-party payments above μ^* and the set of evaluations without such payments below μ^* by $A^D = \{\mu \in [0,1] | \mu \leq \mu^*$ and $c(\mu) = \bar{w}\}$. By assumption $\int_{A^S} f(\mu|e)\mathrm{d}\mu,\ \int_{A^D} f(\mu|e)\mathrm{d}\mu > 0$. Now reduce the bigger set, until both sets have the same mass, $\int_{A^S} f(\mu|e)\mathrm{d}\mu = \int_{A^D} f(\mu|e)\mathrm{d}\mu$. In the next step, I modify the wage to

$$c'(\mu) = \begin{cases} u^{-1}\left(\frac{1}{\int_{A^D} f(\mu|e)\mathrm{d}\mu} \int_{A^S} u(c(\mu)) f(\mu|e)\mathrm{d}\mu\right) & \text{for } \mu \in A^D \\ \bar{w} & \text{for } \mu \in A^S \\ c(\mu) & \text{otherwise.} \end{cases} \tag{22}$$

[16] There is an optimal contract if the distributions F^H and F^L have atoms at 0.

On A^D the agent's wage is reduced to the average wage on A^S in contract $\mathcal{W}$, while on A^D the wage increases to $\bar{w}$. Otherwise, the contract remains unchanged. In order to check whether this contract is feasible, I analyze the remaining constraints (PC) and (IC$_A$). By the definition of the wage modification $\mathcal{W}'$ in (22), the agent's participation constraint (PC) is still satisfied. On the other hand, the left-hand side of the agent's incentive compatibility (IC$_A$) increases by

$$\int_{A^D} \underbrace{\big(u(c'(\mu)) - u(\bar{w})\big)}_{<0}(f^H(\mu) - f^L(\mu))\mathrm{d}\mu + \int_{A^S} \underbrace{\big(u(\bar{w}) - u(c(\mu))\big)}_{>0}(f^H(\mu) - f^L(\mu))\mathrm{d}\mu >$$
$$> \frac{f^H(\mu^*) - f^L(\mu^*)}{f(\mu^*|e)}\Big(\int_{A^D} u(c'(\mu))f(\mu|e)\mathrm{d}\mu - \int_{A^S} u(c(\mu))f(\mu|e)\mathrm{d}\mu\Big) = 0.$$

The monotone likelihood ratio property ensures the strict inequality and the constant value of the participation constraint yields the equality. Therefore the modified wage $\mathcal{W}'$ in (22) makes the left-hand side of condition (IC$_A$) bigger than the marginal cost of effort, $d'(e)$. In addition, the contract satisfies condition (PC). By reducing the variance of the wage payments, until condition (IC$_A$) holds with equality, condition (PC) becomes slack. This allows decreasing the wage. Consequently, restrict attention to contracts with a $\delta < 1$, such that $c(\mu) = \bar{w}$ for all $\mu > \delta$.

Third, consider such a contract $\mathcal{W}$ with $\hat{\delta}$. To satisfy the agent's incentive compatibility, there has to be a payment to a third party, i.e., $c(\mu) < \bar{w}$, with positive probability. Denote the corresponding set by $A^S = \{\mu \in [0,1] | c(\mu) < \bar{w}\}$ and its median by α. By the previous remarks, these values are below $\hat{\delta}$, i.e., $\mu \leq \hat{\delta}$ for all $\mu \in A^S$. Now, modify the wage to

$$c''(\mu) = \begin{cases} u^{-1}(\hat{u}(\mu)) & \text{for } \mu \leq \alpha \text{ and } \mu \in A^S \\ \bar{w} & \text{otherwise} \end{cases}$$
$$\hat{u}(\mu) = u(c(\mu)) - u(\bar{w}) + \frac{2}{\int_{A^S} f(\bar{\mu}|e)\mathrm{d}\bar{\mu}} \int_{\bar{\mu}\in A^S \text{ and } \bar{\mu}>\alpha} u(c(\bar{\mu}))f(\bar{\mu}|e)\mathrm{d}\bar{\mu}.$$

For evaluations above the median α, the agent's wage increases to $\bar{w}$, while below the median the wage is reduced to balance the utility gain above the median. Otherwise, the contract remains unchanged. The proof is now analogous to the last case. Therefore it is possible to improve the contract, as long as $\hat{\delta} > 0$. Yet the contract with a constant wage for the agent and $\hat{\delta} = 0$ does not satisfy the agent's incentive compatibility. Therefore an optimal contract does not exist and I have to consider a sequence of contracts.

For this purpose, construct a sequence of feasible contracts that satisfy the properties derived in this proof. For those contracts, I derive upper and lower bounds for their costs.

Consider the following contracts for a small $\delta > 0$: $W_\delta(m_P, m_A) = \bar{w}(\delta)$ and

$$\bar{c}_\delta(m_P, m_A) = c_\delta\left(\int m_P(t)\mathrm{d}t\right) = \begin{cases} \bar{w}(\delta) & \text{if } \int m_P(t)\mathrm{d}t > \delta \\ \bar{w}(\delta) - \Lambda_\delta(\int m_P(t)\mathrm{d}t) & \text{if } \int m_P(t)\mathrm{d}t \leq \delta, \end{cases}$$

with $\bar{w}(\delta)$ and third-party payments $\Lambda_\delta(\mu)$ such that the agent's incentive compatibility and his participation constraint are satisfied. This requires

$$\int_0^\delta u(c_\delta(\mu))f(\mu|e)\mathrm{d}\mu + (1 - F(\delta|e))u(\bar{w}(\delta)) = \bar{u} + d(e) \tag{PC}$$

$$\int_0^\delta u(c_\delta(\mu))(f^H(\mu) - f^L(\mu))\mathrm{d}\mu - u(\bar{w}(\delta))\int_0^\delta f^H(\mu) - f^L(\mu)\mathrm{d}\mu = d'(e) \tag{IC$_A$}$$

as $\int_\delta^1 f^H(\mu) - f^L(\mu)\mathrm{d}\mu = \int_0^\delta f^L(\mu) - f^H(\mu)\mathrm{d}\mu$. The constraint (IC$_A$) implies $u(c_\delta(\mu)) < 0$ for all $\mu \leq \delta$ with δ sufficiently small, because $\int_0^\delta f^H(\mu) - f^L(\mu)\mathrm{d}\mu \to 0$ for $\delta \to 0$ and $f^H(\mu) - f^L(\mu) < 0$ for all $\mu \leq \delta$. Rearranging the agent's incentive compatibility and approximating it from above results in

$$d'(e) \leq -u(\bar{w}(\delta))\int_0^\delta f^H(\mu) - f^L(\mu)\mathrm{d}\mu + \frac{f^H(0) - f^L(0)}{f(0|e)}\int_0^\delta u(c_\delta(\mu))f(\mu|e)\mathrm{d}\mu.$$

In the next step, insert the participation constraint for $\int_0^\delta u(c_\delta(\mu))f(\mu|e)\mathrm{d}\mu$ to get

$$u(\bar{w}(\delta)) \geq \frac{d'(e) + \frac{f^L(0) - f^H(0)}{f(0|e)}(\bar{u} + d(e))}{\int_0^\delta f^L(\mu) - f^H(\mu)\mathrm{d}\mu + \frac{f^L(0) - f^H(0)}{f(0|e)}(1 - F(\delta|e))}$$

as a lower bound or

$$u(\bar{w}(\delta)) \leq \frac{d'(e) + \frac{f^L(\delta) - f^H(\delta)}{f(\delta|e)}(\bar{u} + d(e))}{\int_0^\delta f^L(\mu) - f^H(\mu)\mathrm{d}\mu + \frac{f^L(\delta) - f^H(\delta)}{f(\delta|e)}(1 - F(\delta|e))}$$

as an upper bound. For $\delta \to 0$, both bounds converge to

$$\bar{u} + d(e) + \frac{f(0|e)}{f^L(0) - f^H(0)}d'(e).$$

This also coincides with the optimal contract in the case of atoms at 0, which guarantee existence of an optimal contract. To ensure that the set of these contracts $\mathcal{W}_\delta$ is nonempty, consider such a contract with $\bar{w}(\delta)$ and $\Lambda_\delta(\bar{\mu})$ determined by:

$$u(\bar{w}(\delta)) = \bar{u} + d(e) + \frac{F(\delta|e)}{\int_0^\delta f^L(\mu) - f^H(\mu)\mathrm{d}\mu}d'(e) \qquad \text{and}$$

$$u(\bar{w}(\delta) - \Lambda_\delta(\bar{\mu})) = \bar{u} + d(e) - \frac{1 - F(\delta|e)}{\int_0^\delta f^L(\mu) - f^H(\mu)\mathrm{d}\mu}d'(e), \qquad \forall \bar{\mu} \leq \delta.$$

This contract satisfies all the properties derived in this proof as well as constraints (PC) and (IC$_A$) with equality. The costs of this contract are $\bar{w}(\delta)$ which by l'Hôpital's rule

converges to $C^0(e) = u^{-1}\left(\bar{u} + d(e) + \frac{f(0|e)}{f^L(0)-f^H(0)}d'(e)\right)$ for $\delta \searrow 0$. The convergence is also monotone in the costs of the contract, because $\bar{w}(\delta)$ is strictly increasing in δ. As the densities are continuous, for every $\epsilon > 0$ there exists a $\delta' > 0$, such that the costs of the contract $\mathcal{W}_\delta$ for all $0 < \delta \leq \delta'$ are lower than $C^0(e) + \epsilon$. Yet it is impossible to approximate first-best, because $C^0(e) > u^{-1}(\bar{u} + d(e))$.

For the final step of the proof, compare the principal's costs with and without justification. If she provides justification with probability 0, $\Pr(\mathcal{I}_C) = 0$ and the principal's costs are $C^0(e)$. Notice that $\Lambda_\delta(\bar{\mu}) \to C^0(e)$ for $\delta \searrow 0$. If $\kappa < C^0(e)$, there is a $\bar{\delta}$, such that $\Lambda_\delta(\bar{\mu}) < \kappa$ for $\delta \in (0, \bar{\delta})$ and the following contract implies lower costs for the principal than $C^0(e)$ for $\delta \in (0, \bar{\delta})$:

$$\bar{c}_\delta(m_P, m_A) = \begin{cases} \bar{w}(\delta) & \text{if } \int m_P(t)\mathrm{d}t > \delta \\ \bar{w}(\delta) - \Lambda_\delta(\bar{\mu}) & \text{if } \int m_P(t)\mathrm{d}t \leq \delta, \end{cases}$$

$$W_\delta(m_P, m_A) = \begin{cases} \bar{w}(\delta) & \text{if } \int m_P(t)\mathrm{d}t > \delta \\ \bar{w}(\delta) - \Lambda_\delta(\bar{\mu}) & \text{if } \int m_P(t)\mathrm{d}t \leq \delta \text{ and } m_P(t) = 1 \text{ for all } t \in m_A \\ \bar{w}(\delta) + \kappa & \text{otherwise } \int m_P(t)\mathrm{d}t \leq \delta. \end{cases}$$

In this contract, the principal justifies evaluations below δ. Intuitively, communication costs are lower than the third-party payments $\Lambda_\delta(\bar{\mu})$. Therefore, justifying some low evaluations allows reducing the principal's costs. Hence, as long as the communication costs are lower than this bound, the principal will justify evaluations with positive probability.

If $\kappa \geq C^0(e)$, there is no justification in the optimal contract. □

Proof of Proposition 6: Return to condition (21) which characterizes the optimal justification threshold δ. Rearranging yields

$$u(w^{**}) - u(w^{**}(\delta)) - u'(w^{**}(\delta))(w^{**} - w^{**}(\delta)) = -\kappa u'(w^{**}(\delta))$$

Taylor's theorem for u at $w^{**}(\delta)$ ensures that there is a $\hat{w} \in [w^{**}(\delta), w^{**}]$ such that the left-hand side of the last equation equals $u''(\hat{w})(w^{**} - w^{**}(\delta))^2/2 < 0$. Therefore

$$\kappa = \frac{-u''(\hat{w})}{u'(w^{**}(\delta))}\frac{1}{2}(w^{**} - w^{**}(\delta))^2.$$

For $\kappa \searrow 0$, $w^{**}(\delta) \nearrow w^{**}$ satisfies this equation. Then (18) and (19) imply $\delta \nearrow 1$. □

A.4 Budget-Balanced Contracts

It is possible to make the optimal contract ex-post budget-balanced.

Proof of Proposition 7: In order to capture stochastic payments, change the interpretation of the notation. Now $w(m_P, m_A)$ denotes the expected wage after principal and agent sent messages m_P and m_A. On the other hand, $c(m_P, m_A)$ is the agent's certainty equivalent of the wage payment. Finally, lotteries with the corresponding mean and cer-

tainty equivalent will be specified. This formalization captures any stochastic payment without loss of generality. The principal could pay the agent a lottery or could discard certain messages with some probability by 'turning a blind eye'.[17]

Program A still describes the problem. Yet constraint (1) now captures the agent's risk aversion. Proposition 2 states the solution to Program A. The solution $\mathcal{W}^{**}$ gives the principal and the agent the same utilities as the contract $\bar{\mathcal{W}}^{**}$ in Proposition 7 if the lotteries are chosen accordingly. For the principal this is obvious, as the lotteries have mean κ. The agent's expected utility also remains unchanged, because the certainty equivalent equals his former wage.

Finally, specify a lottery with the desired properties. Let Λ_z denote a lottery that pays $\kappa + z$ and $\kappa - z$ with probability 1/2, respectively.[18] $z(\mu)$ is determined, such that

$$\mathbb{E}u(w^{**} + \Lambda_{z(\mu)}) = \frac{1}{2}u(w^{**} + \kappa + z(\mu)) + \frac{1}{2}u(w^{**} + \kappa - z(\mu)) = u(w^{**}(\mu)).$$

It is possible to find such a $z(\mu)$. Additionally, $z(\mu)$ is unique for every μ due to the strict concavity of u. The lottery does not change the principal's expected payments, but reduces the agent's certainty equivalent. The mean preserving spread introduced by the lottery is the reason for this loss of utility. Once the lottery is realized, the party who gains in the lottery has an incentive to avoid renegotiations. Therefore the lottery ought to be realized as soon as the messages are available in order to make the contract renegotiation-proof.[19] □

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.

Aghion, P., Fudenberg, D., Holden, R., Kunimoto, T., and Tercieux, O. (2012). Subgame-Perfect Implementation Under Value Perturbations and the Hold-up Problem. *Quarterly Journal of Economics*, 127(4): 1843–1881.

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.

[17] This is optimal for example in Herweg et al. (2010).

[18] If the agent has limited liability, the lottery Λ_z can be redefined to limit the negative realization, but to assign it a higher probability.

[19] Rasmusen (1987) and Maskin (2002) provide a good discussion of this problem.

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.

Bull, C. (1987). The existence of self-enforcing implicit contracts. *Quarterly Journal of Economics*, 102(1):147–159.

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, Upper Saddle River, N.J., 11th edition.

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

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. (1991). An Economic Approach to Process in Pay and Performance Appraisals. *mimeo*.

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

Green, J. R. and Laffont, J.-J. (1986). Partially Verifiable Information and Mechanism Design. *Review of Economic Studies*, 53(3):447–456.

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

Grossman, S. J. and Hart, O. D. (1986). The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. *Journal of Political Economy*, 94(4):691–719.

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

Harris, M. M., Smith, D. E., and Champagne, D. (1995). A field study of performance appraisal purpose: Research versus administrative based ratings. *Personnel Psychology*, 48(1):151–160.

Hart, O. and Moore, J. (1988). Incomplete Contracts and Renegotiation. *Econometrica*, 56(4):755–785.

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

Herweg, F., Müller, D., and Weinschenk, P. (2010). Binary payment schemes: Moral Hazard and Loss Aversion. *American Economic Review*, 100(5):2451–2477.

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.

Jawahar, I. M. and Williams, C. R. (1997). Where all the children are above average: the performance appraisal purpose effect. *Personnel Psychology*, 50(4):905–925.

Judd, K. L. (1985). The law of large numbers with a continuum of IID random variables. *Journal of Economic Theory*, 35(1):19–25.

Kamien, M. I. and Schwartz, N. L. (1991). *Dynamic optimization – the calculus of variations and optimal control in economics and management*. Advanced textbooks in economics. Elsevier, Amsterdam, second edition.

Kandori, M. and Matsushima, H. (1998). Private Observation, Communication and Collusion. *Econometrica*, 66(3):627–652.

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

Landy, F. J. and Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, 87(1):72–107.

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

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

MacLeod, W. B. 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.

Maskin, E. (2002). On indescribable contingencies and incomplete contracts. *European Economic Review*, 46(4):725–733.

Maskin, E. and Tirole, J. (1999). Unforeseen contingencies and incomplete contracts. *Review of Economic Studies*, 66(1):83–114.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., and Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79(4):599–616.

Medoff, J. L. and Abraham, K. G. (1980). Experience, performance, and earnings. *Quarterly Journal of Economics*, 95(4):703–736.

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

Milkovich, G. T. and Wigdor, A. K. (1991). *Pay for Performance: Evaluating Performance Appraisal and Merit Pay*. National Academy Press, Washington, DC.

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

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

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

Rasmusen, E. (1987). Moral Hazard in Risk-Averse Teams. *RAND Journal of Economics*, 18(3):428–435.

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.

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.

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.