Binary Payment Schemes: Moral Hazard and Loss Aversion*

Fabian Herweg, Daniel Müller, and Philipp Weinschenk

November 18, 2009

We modify the principal-agent model with moral hazard by assuming that the agent is expectation-based loss averse according to Kőszegi and Rabin (2006, 2007). The optimal contract is a binary payment scheme even for a rich performance measure, where standard preferences predict a fully contingent contract. The logic is that, due to the stochastic reference point, increasing the number of different wages reduces the agent's expected utility without providing strong additional incentives. Moreover, for diminutive occurrence probabilities for all signals the agent is rewarded with the fixed bonus if his performance exceeds a certain threshold. (*JEL* D82, M12, M52)

*Herweg: University of Bonn, Department of Economics, Chair of Economic Theory II, Lenné Str. 43, D-53113 Bonn, Germany, fherweg@uni-bonn.de. Müller: Bonn Graduate School of Economics, University of Bonn, Adenauerallee 24-42, D-53113 Bonn, Germany, daniel.mueller@uni-bonn.de. Weinschenk: Bonn Graduate School of Economics and Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn, Germany, weinschenk@coll.mpg.de. In preparing this paper we have greatly benefited from comments made by Felix Bierbrauer, Patrick Bolton, Jörg Budde, Paul Heidhues, Martin Hellwig, Botond Kőszegi, Sebastian Kranz, Matthias Lang, Kristóf Madarász, Patrick Schmitz, Urs Schweizer, and four anonymous referees. We also thank seminar participants at University of Bonn and the MPI, Bonn, as well as participants at the IMEBE at Alicante (2008), EEA/ESEM at Milan (2008), the annual congress of the Verein für Socialpolitik at Graz (2008), Nordic Conference on Behavioral and Experimental Economics at Copenhagen (2008), the Workshop on Industrial Organization and Antitrust Policy at DIW Berlin (2008), Behavioral Models of Market Competition at Bad Homburg (2009), and EARIE at Ljubljana (2009). An earlier draft of this paper was circulated under the title "The Optimality of Simple Contracts: Moral Hazard and Loss Aversion".

The recent literature provides very strong evidence that contractual forms have large effects on behavior. As the notion that "incentive matters" is one of the central tenets of economists of every persuasion, this should be comforting to the community. On the other hand, it raises an old puzzle: if contractual form matters so much, why do we observe such a prevalence of fairly simple contracts?

—Bernard Salanié (2003, 474)

A lump-sum bonus contract, with the bonus being a payment for achieving a certain level of performance, is probably one of the most simple incentive schemes for employees one can think of. According to Thomas J. Steenburgh (2008), salesforce compensation plans provide incentives mainly via a lump-sum bonus for meeting or exceeding the annual sales quota.[1] The observed plainness of contractual arrangements, however, is at odds with predictions made by economic theory, as nicely stated in the above quote by Bernard Salanié (2003). While Canice Prendergast (1999) already referred to the discrepancy between theoretically predicted and actually observed contractual form, over time this question was raised again and again, recently by Edward P. Lazear and Oyer (2007), and the answer still is not fully understood.

Beside this gap between theoretical prediction and observed practice, both theoretical and empirical studies demonstrate that these simple contractual arrangements create incentives for misbehavior of the agent that is outside the scope of most standard models. As Oyer (1998) points out, facing an annual sales quota provides incentives for salespeople to manipulate prices and timing of business to maximize their own income rather than firms' profits. This observation raises "the interesting question of why these (...) contracts are so prevalent. (...) It appears that there must be some benefit of these contracts that outweighs these apparent costs" (Lazear and Oyer, 2007, 16).

To give one possible explanation for the widespread use of binary payment schemes, we modify the principal-agent model with moral hazard by assuming that the agent is expectation-based loss averse according to Botond Kőszegi and Matthew Rabin (2006, 2007).[2] With the tradeoff between incentive provision and risk sharing being at the heart of moral

[1] Incentives for salespeople in the food manufacturing industry often are solely created by a lump-sum bonus, see Paul Oyer (2000). Moreover, in his book about designing effective sales compensation plans, John K. Moynahan (1980) argues that for a wide range of industries lump-sum bonus contracts are optimal. For a survey on salesforce compensation plans, see Kissan Joseph and Manohar U. Kalwani (1998). Simple binary contracts are commonly found not only in labor contexts, but also in insurance markets, where straight-deductible contracts are prevalent.

[2] We will use the terms bonus contract, bonus scheme, and binary payment scheme interchangeably to refer to a contract that specifies exactly two distinct wage payments, a base wage and a lump-sum bonus.

hazard, allowing for a richer description of the agent's risk preferences that goes beyond standard risk aversion seems a natural starting point to gain deeper insights into contract design. Following Kőszegi and Rabin, we posit that the agent—next to standard consumption utility—derives gain-loss utility from comparing the actual outcome with his lagged expectations. Specifically, the agent compares his actual wage pairwise which each other wage that he could have received instead, where each comparison is weighted by the occurrence probability of the alternative outcome. Our main finding is that a simple (lump-sum) bonus scheme is optimal when loss aversion is the driving force of the agent's risk preferences. We derive this finding in a model where the principal can make use of a rich performance measure and where the standard notion of risk aversion would predict fully contingent contracts. Intuitively, specifying many different payments induces uncertainty for the agent as to what he will receive. If he earns a relatively low wage, he compares this to higher wages he could have received, and experiences the sensation of a loss from this comparison. The anticipation of these losses reduces the agent's expected utility and thus he demands a higher average payment. While the principal has a classic rationale for rewarding signals strictly higher if they are stronger indicators of good performance, the negative "comparison effect" dominates this consideration if standard risk aversion plays a minor role.[3] In this sense, reference-dependent preferences according to Kőszegi and Rabin introduce an endogenous complexity cost into contracting based on psychological foundations.

We establish several properties displayed by the optimal contract. Let a signal that is the more likely to be observed the higher the agent's effort be referred to as a "good" signal. We find that the subset of signals that is rewarded with the bonus payment contains either only good signals, or all good signals and possibly a few bad signals as well.[4] By paying the bonus very often or very rarely, the principal can minimize the weight the agent puts ex ante on the disappointing event of feeling a loss when not obtaining the bonus. When abstracting from integer-programming problems, it is optimal for the principal to order the signals according to their relative informativeness (likelihood ratio). Put differently, the agent receives the bonus for all signals that are more indicative of high effort than a cutoff signal, e.g., a salesperson receiving a bonus for meeting or exceeding the annual sales quota.

In addition, we show that an increase in the agent's degree of loss aversion may allow the

[3]The term "comparison effect" was first introduced by Kőszegi and Rabin (2006).

[4]The theoretical prediction that inferior performance may also well be rewarded with a bonus is in line with both Joseph and Kalwani's (1998) suggestion that organizations tend to view the payment of a bonus as a reward for good or even acceptable performance rather than an award for exceptional performance, and Gilbert A. Churchill, Neil M. Ford and Orville C. Walker's (1993) prescription that bonuses should be based on objectives that can be achieved with reasonable rather than Herculean efforts.

principal to use a lower-powered incentive scheme. The reason is that a higher degree of loss aversion may be associated with a stronger incentive for the agent to choose a high effort in order to reduce the scope for incurring a loss. The overall cost of implementation, however, increases in the agent's degree of loss aversion.

While assuming for most of the paper that the agent is not too loss averse, which guarantees that the first-order approach is valid, we also briefly investigate the principal's problem for higher degrees of loss aversion. Here, to keep the analysis tractable, we focus on binary measures of performance. We show that if the agent's degree of loss aversion is sufficiently high and if the performance measure is sufficiently informative, then only extreme actions—work as hard as possible or do not work at all—are incentive compatible. Put differently, the principal may face severe problems in fine-tuning the agent's incentives. These implementation problems, however, can be remedied if the principal can commit herself to stochastically ignoring the performance measure. Moreover, for high degrees of loss aversion, stochastic ignorance of the performance measure also lowers the cost of implementing the desired level of effort. The logic of this result is that stochastic ignorance allows the principal to pay the bonus to the agent even if she observes the signal that indicates low effort. By doing this, the agent considers it ex ante less likely that he will be disappointed at the end of the day, and thus he demands a lower average payment. In this case, with the optimal contract including randomization which would not be optimal under the standard notion of risk aversion, loss aversion leads to more complex contracts than predicted by orthodox theory.

Before launching out into the model description, we briefly review the existing evidence documenting that expectations matter in the determination of the reference point, which is a key feature of the Kőszegi-Rabin concept.[5] While mainly based on findings in the psychological literature,[6] evidence for this assumption is provided also by some recent contributions to the economic literature. Investigating decision making in a large-stake game show, Thierry Post et al. (2008, 62) come to the conclusion that observed behavior is "consistent with the idea that the reference point is based on expectations." Alike, analyzing field data, Vincent P. Crawford and Juanjuan Meng (2009) propose a model of cabdrivers' labor supply that builds on the Kőszegi-Rabin theory of reference-dependent preferences. Their estimates suggest that a reference-dependent model of drivers' labor supply where targets are carefully

[5]The feature that the reference point is determined by the decision maker's forward-looking expectations is shared with the disappointment aversion models of David E. Bell (1985), Graham Loomes and Robert Sugden (1986), and Faruk Gul (1991).

[6]For instance, Barbara Mellers, Alan Schwartz and Ilana Ritov (1999) and Hans C. Breiter et al. (2001) document that both the actual outcome and unattained possible outcomes affect subjects' satisfaction with their payoff.

modeled significantly improves on the neoclassic model. In a real-effort experiment, Johannes Abeler et al. (forthcoming) manipulate the rational expectations of subjects. They find that effort provision is significantly different between treatments in the way predicted by models of expectation-based loss aversion.

In the following Section I, we formulate the principal-agent relationship. Section II specifies the principal's problem and derives the set of feasible contracts. In Section III, the principal's problem is solved and properties of the optimal contract are discussed. Section IV investigates the implications of high degrees of loss aversion. In Section V, next to the related literature, alternative notions of loss aversion are discussed. Section VI concludes. All proofs of Sections II and III deferred to the Appendix. The proofs corresponding to Section IV as well as additional technical discussions are relegated to the Web Appendix.

I. The Model

A principal offers a one-period employment contract to an agent, who has an outside employment opportunity yielding expected utility $\bar{u}$.[7] If the agent accepts the contract, then he chooses an effort level $a \in \mathcal{A} \equiv [0,1]$. The agent's action a equals the probability that the principal receives a benefit $B > 0$. The principal's expected net benefit is

$$\pi = aB - E[W] \,,$$

where W is the compensation payment the principal pays to the agent.[8] The principal is assumed to be risk and loss neutral, thus she maximizes π. We wish to inquire into the form that contracts take under moral hazard and loss aversion. Therefore, we focus on the cost minimization problem to implement a certain action $\hat{a} \in (0,1)$.

The action choice $a \in \mathcal{A}$ is private information of the agent and unobservable for the principal. Furthermore, the realization of B is not directly observable. A possible interpretation is that B corresponds to a complex good whose quality cannot be determined by a court, thus a contract cannot depend on the realization of B. Instead the principal observes a contractible measure of performance, $\hat{\boldsymbol{\gamma}}$, with $s \in \mathcal{S} \equiv \{1, \ldots, S\}$ being the realization of the performance measure. Let $S \geq 2$. The probability of observing signal s conditional on B being realized is denoted by γ_s^H. Accordingly, γ_s^L is the probability of observing signal s conditional on B not being realized. Hence, the unconditional probability of observing signal

[7] The framework is based on W. Bentley MacLeod (2003), who analyzes subjective performance measures without considering loss-averse agents.

[8] The particular functional form of the principal's profit function is not crucial for our analysis. We assume this specific structure since it allows for a straight-forward interpretation of the performance measure.

s for a given action a is $\gamma_s(a) \equiv a\gamma_s^H + (1-a)\gamma_s^L$.[9] For technical convenience, we make the following assumption.

ASSUMPTION (A1): *For all $s, \tau \in \mathcal{S}$ with $s \neq \tau$,*

(i) $\gamma_s^H/\gamma_s^L \neq 1$ *(informative signals),*

(ii) $\gamma_s^H, \gamma_s^L \in (0,1)$ *(full support),*

(iii) $\gamma_s^H/\gamma_s^L \neq \gamma_\tau^H/\gamma_\tau^L$ *(different signals).*

Part (i) guarantees that any signal s is either a good or a bad signal, i.e., the overall probability of observing that signal unambiguously increases or decreases in a. Part (ii) ensures that for all $a \in \mathcal{A}$, all signals occur with positive probability. Last, with part (iii) signals can unambiguously be ranked according to the relative impact of an increase in effort on the probability of observing a particular signal.

The contract which the principal offers to the agent consists of a payment for each realization of the performance measure, $(w_s)_{s=1}^S \in \mathbb{R}^S$.[10]

The agent is assumed to have reference-dependent preferences in the sense of Kőszegi and Rabin (2006): overall utility from consuming $\boldsymbol{x} = (x_1, \ldots, x_K) \in \mathbb{R}^K$—when having reference level $\boldsymbol{r} = (r_1, \ldots, r_K) \in \mathbb{R}^K$ for each dimension of consumption—is given by

$$v(\boldsymbol{x}|\boldsymbol{r}) \equiv \sum_{k=1}^{K} m_k(x_k) + \sum_{k=1}^{K} \mu(m_k(x_k) - m_k(r_k)).$$

Put verbally, overall utility is assumed to have two components: consumption utility and gain-loss utility. Consumption utility, also called intrinsic utility, from consuming in dimension k is denoted by $m_k(x_k)$. How a person feels about gaining or losing in a dimension is assumed to depend in a universal way on the changes in consumption utility associated with such gains and losses. The universal gain-loss function $\mu(\cdot)$ satisfies the assumptions imposed by Amos Tversky and Daniel Kahneman (1991) on their "value function". In our model, the agent's consumption space comprises of two dimensions, money income ($x_1 = W$) and effort ($x_2 = a$).[11] The agent's intrinsic utility for money is assumed to be a strictly increasing, (weakly) concave, and unbounded function. Formally, $m_1(W) = u(W)$ with $u'(\cdot) > \varepsilon > 0$,

[9] The results of Section III do not rely on the linear structure of the performance measure. The linearity is needed to show validity of the first-order approach and in Section IV.

[10] Restricting the principal to offer nonstochastic wage payments is standard in the principal-agent literature and also in accordance with observed practice. In Section IV, we comment on this assumption.

[11] We implicitly assume that the agent is a "narrow bracketer", in the sense that he ignores that the risk from the current employment relationship is incorporated with substantial other risk.

$u''(\cdot) \leq 0$. The intrinsic disutility from exerting effort $a \in [0,1]$ is a strictly increasing, strictly convex function of effort, $m_2(a) = -c(a)$ with $c'(0) = 0$, $c'(a) > 0$ for $a > 0$, $c''(\cdot) > 0$, and $\lim_{a \to 1} c(a) = \infty$. We assume that the gain-loss function is piece-wise linear,[12]

$$\mu(m) = \begin{cases} m \ , & \text{for } m \geq 0 \\ \lambda m, & \text{for } m < 0 \end{cases} .$$

The parameter $\lambda \geq 1$ characterizes the weight put on losses relative to gains.[13] The weight on gains is normalized to one. When $\lambda > 1$, the agent is loss averse in the sense that losses loom larger than equally-sized gains.

Following Kőszegi and Rabin (2006, 2007), the agent's reference point is determined by his rational expectations about outcomes. A given outcome is then evaluated by comparing it to all possible outcomes, where each comparison is weighted with the ex ante probability with which the alternative outcome occurs. With the actual outcome being itself uncertain, the agent's expected utility is obtained by averaging over all these comparisons. We apply the concept of choice-acclimating personal equilibrium (CPE) as defined in Kőszegi and Rabin (2007), which assumes that a person correctly predicts his choice set, the environment he faces, in particular the set of possible outcomes and how the distribution of these outcomes depends on his decisions, and his own reaction to this environment. The eponymous feature of CPE is that the agent's reference point is affected by his choice of action. As pointed out by Kőszegi and Rabin, CPE refers to the analysis of risk preferences regarding outcomes that are resolved long after all decisions are made. This environment seems well-suited for many principal-agent relationships: often the outcome of a project becomes observable, and thus performance-based wage compensation feasible, long after the agent finished working on that project. Under CPE, the expectations relative to which a decision's outcome is evaluated are formed at the moment the decision is made and, therefore, incorporate the implications of the decision. More precisely, suppose the agent chooses action a and that signal s is observed. The agent receives wage w_s and incurs effort cost $c(a)$. While the agent expected signal s to come up with probability $\gamma_s(a)$, with probability $\gamma_\tau(a)$ he expected signal $\tau \neq s$ to be observed. If $w_\tau > w_s$, the agent experiences a loss of $\lambda(u(w_s) - u(w_\tau))$, whereas if $w_\tau < w_s$, the agent experiences a gain of $u(w_s) - u(w_\tau)$. If $w_s = w_\tau$, there is no sensation of

[12] In their work on asset pricing, Nicholas Barberis, Ming Huang, and Tano Santos (2001) argue that for prospects involving both gains and losses, loss aversion at the kink is more relevant than the degree of curvature away from the kink. Implications of a more general gain-loss function are discussed in Section VI.

[13] Alternatively, one could introduce a weight attached to gain-loss utility relative to intrinsic utility, $\eta \geq 0$. We implicitly normalized $\eta = 1$ which can be done without much loss, since this normalization does not qualitatively affect any of our results.

gaining or losing involved. The agent's utility from this particular outcome is given by

$$u(w_s) + \sum_{\{\tau|w_\tau<w_s\}} \gamma_\tau(a)(u(w_s) - u(w_\tau)) + \sum_{\{\tau|w_\tau\geq w_s\}} \gamma_\tau(a)\lambda(u(w_s) - u(w_\tau)) - c(a).$$

Note that since the agent's expected and actual effort choice coincide, there is neither a gain nor a loss in the effort dimension. We conclude this section by briefly summarizing the underlying timing.

1) The principal makes a take-it-or-leave-it offer to the agent.

2) The agent either accepts or rejects the contract. If the agent rejects, the game ends and each party receives her/his reservation payoff. If the agent accepts, the game moves to the next stage.

3) The agent chooses his action and forms rational expectations about the monetary outcomes. The contract and the agent's rational expectations about the realization of the performance measure determine his reference point.

4) Both parties observe the realization of the performance measure and payments are made according to the contract.

II. Preliminary Analysis

Let $h(\cdot) := u^{-1}(\cdot)$, i.e., the monetary cost for the principal to offer the agent utility u_s is $h(u_s) = w_s$. Due to the assumptions imposed on $u(\cdot)$, $h(\cdot)$ is a strictly increasing and weakly convex function. Following Sanford J. Grossman and Oliver D. Hart (1983), we regard $\boldsymbol{u} = (u_1, \ldots, u_S)$ as the principal's control variables in her cost minimization problem to implement action $\hat{a} \in (0, 1)$. The principal offers the agent a contract that specifies for each signal a monetary payment or, equivalently, an intrinsic utility level. With this notation, the agent's expected utility from exerting effort a is

$$EU(a) = \sum_{s\in\mathcal{S}} \gamma_s(a)u_s - (\lambda - 1)\sum_{s\in\mathcal{S}} \sum_{\{\tau|u_\tau>u_s\}} \gamma_\tau(a)\gamma_s(a)(u_\tau - u_s) - c(a). \tag{1}$$

For $\lambda = 1$ the agent's expected utility equals expected net intrinsic utility. Thus, for $\lambda = 1$ we are in the standard case without loss aversion. Moreover, from the above formulation of the agent's utility it becomes clear that λ captures not only the weight put on losses relative to gains, but that $(\lambda - 1)$ can also be interpreted as the weight put on gain-loss utility relative to intrinsic utility. Thus, for $\lambda \leq 2$, the weight attached to gain-loss utility is

below the weight attached to intrinsic utility. For a given contract $\boldsymbol{u}$, the agent's marginal utility of effort is

$$(2)\quad EU'(a) = \sum_{s\in\mathcal{S}}(\gamma_s^H - \gamma_s^L)u_s \\ - (\lambda-1)\sum_{s\in\mathcal{S}}\sum_{\{\tau|u_\tau>u_s\}}[\gamma_\tau(a)(\gamma_s^H-\gamma_s^L)+\gamma_s(a)(\gamma_\tau^H-\gamma_\tau^L)](u_\tau-u_s) - c'(a).$$

A principal who wants to implement action $\hat{a}\in(0,1)$ minimizes her expected wage payment subject to the usual individual rationality and incentive compatibility constraints:

$$\min_{u_1,\ldots,u_S}\sum_{s\in\mathcal{S}}\gamma_s(\hat{a})h(u_s)$$

$$\text{(IR)}\qquad \text{subject to}\quad EU(\hat{a})\geq\bar{u}\ ,$$

$$\text{(IC)}\qquad \hat{a}\in\arg\max_{a\in\mathcal{A}}EU(a)\ .$$

Suppose the agent's action choice is contractible, i.e., the incentive constraint (IC) is absent. In this first-best situation, the principal pays a risk- or loss-averse agent a fixed wage $u^{FB}=\bar{u}+c(\hat{a})$. In the presence of moral hazard, on the other hand, the principal faces the classic tradeoff between risk sharing and providing incentives: when the agent is anything but risk and loss neutral, it is neither optimal to have the agent bear the complete risk, nor fully to insure the agent.

At this point we simplify the analysis by imposing two assumptions. These assumptions are sufficient to guarantee that the principal's cost minimization problem exhibits the following two properties: first, there are incentive-compatible wage contracts, i.e., contracts under which it is optimal for the agent to choose the desired action $\hat{a}$. Second, the first-order approach is valid, i.e., the incentive constraint to implement action $\hat{a}$ can equivalently be represented as $EU'(\hat{a})=0$. The first assumption that we introduce requires that the "weight" attached to gain-loss utility does not exceed the weight put on intrinsic utility.

ASSUMPTION (A2): *No dominance of gain-loss utility,* $\lambda\leq 2$.

As carefully laid out in Kőszegi and Rabin (2007), CPE implies a strong notion of risk aversion, in the sense that a decision maker may choose stochastically dominated options when $\lambda>2$. The reason is that, with losses looming larger than gains of equal size, the person ex ante expects to experience a net loss. In consequence, if reducing the scope of possibly incurring a loss is the decision maker's primary concern, the person would rather give up the slim hope of experiencing a gain at all in order to avoid the disappointment in case of not experiencing this gain. In our model, if the agent is sufficiently loss averse, the

principal may be unable to implement any action $\hat{a} \in (0,1)$. The reason is that the agent minimizes his expected net loss by choosing one of the two extreme actions. The values of λ for which this behavior is optimal for the agent depend on the precise structure of the performance measure. Assumption (A2) is sufficient, but not necessary, to ensure that there is a contract such that $\hat{a} \in (0,1)$ satisfies the necessary condition for incentive compatibility. In Section IV, we relax Assumption (A2) and discuss in detail the implications of higher degrees of loss aversion.

To keep the analysis tractable, we impose the following additional assumption on the agent's cost function.

ASSUMPTION (A3): *Convex marginal cost function,* $\forall\, a \in [0,1]:\ c'''(a) \geq 0$.

Given (A2), Assumption (A3) is a sufficient but not a necessary condition for the first-order approach to be applicable.[14] In fact, our results only require the validity of the first-order approach, not that Assumption (A3) holds. In Section IV, we consider the case in which the first-order approach is invalid.

LEMMA 1: *Suppose (A1)-(A3) hold, then the constraint set of the principal's cost minimization problem is nonempty for all* $\hat{a} \in (0,1)$.

The above lemma states that there are wage contracts such that the agent is willing to accept the contract and then chooses the desired action. Existence of a second-best optimal contract is shown separately for the three cases analyzed: pure risk aversion, pure loss aversion, and the intermediate case.

Sometimes it will be convenient to state the constraints in terms of increases in intrinsic utilities instead of absolute utilities. Note that whatever contract $(\hat{u}_s)_{s\in\mathcal{S}}$ the principal offers, we can relabel the signals such that this contract is equivalent to a contract $(u_s)_{s=1}^S$ with $u_{s-1} \leq u_s$ for all $s \in \{2,\ldots,S\}$. This, in turn, allows us to write the contract as $u_s = u_1 + \sum_{\tau=2}^{s} b_\tau$, where $b_\tau = u_\tau - u_{\tau-1} \geq 0$. Let $\boldsymbol{b} = (b_2,\ldots,b_S)$. With this notation the individual rationality constraint can be stated as follows:

$$\text{(IR}') \qquad u_1 + \sum_{s=2}^{S} b_s \left[\sum_{\tau=s}^{S} \gamma_\tau(\hat{a}) - \rho_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})\right] \geq \bar{u} + c(\hat{a})\ ,$$

where

$$\rho_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) := (\lambda - 1)\left[\sum_{\tau=s}^{S} \gamma_\tau(\hat{a})\right]\left[\sum_{t=1}^{s-1} \gamma_t(\hat{a})\right]\ .$$

[14] The validity of the first-order approach under assumptions (A1)-(A3) is proven in the Web Appendix. The reader should be aware that the proof requires notation introduced later in this section.

Let $\boldsymbol{\rho}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) = (\rho_2(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}), \ldots, \rho_S(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}))$. The first part of the agent's utility, $u_1 + \sum_{s=2}^{S} b_s(\sum_{\tau=s}^{S} \gamma_\tau(\hat{a}))$, is the expected intrinsic utility for money. Due to loss aversion, however, the agent's utility has a second negative component, the term $\boldsymbol{b}'\boldsymbol{\rho}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})$. With bonus b_s being paid to the agent whenever a signal higher or equal to s is observed, the agent expects to receive b_s with probability $\sum_{\tau=s}^{S} \gamma_\tau(\hat{a})$. With probability $\sum_{t=1}^{s-1} \gamma_t(\hat{a})$, however, a signal below s will be observed, and the agent will not be paid bonus b_s. Thus, with "probability" $[\sum_{\tau=s}^{S} \gamma_\tau(\hat{a})][\sum_{t=1}^{s-1} \gamma_t(\hat{a})]$ the agent experiences a loss of λb_s. Analogous reasoning implies that the agent will experience a gain of b_s with the same probability. With losses looming larger than gains of equal size, in expectation the agent suffers from deviations from his reference point. This expected net loss is captured by the term, $\boldsymbol{b}'\boldsymbol{\rho}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})$, which we will refer to as the agent's "loss premium".[15] A crucial point is that the loss premium increases in the contract's degree of wage differentiation. When there is no wage differentiation at all, i.e., $\boldsymbol{b} = \boldsymbol{0}$, then the loss premium vanishes. If, in contrast, the contract specifies many different wage payments, then the agent ex ante considers a deviation from his reference point very likely. Put differently, for each additional wage payment an extra negative term enters the agent's loss premium and therefore reduces his expected utility.[16]

Given that the first-order approach is valid, the incentive constraint can be rewritten as

$$\text{(IC}') \qquad \sum_{s=2}^{S} b_s \beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) = c'(\hat{a}),$$

where

$$\beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) := \left(\sum_{\tau=s}^{S} (\gamma_\tau^H - \gamma_\tau^L)\right) - (\lambda - 1)\left[\left(\sum_{t=1}^{s-1} \gamma_t(\hat{a})\right)\left(\sum_{\tau=s}^{S} (\gamma_\tau^H - \gamma_\tau^L)\right) + \left(\sum_{\tau=s}^{S} \gamma_\tau(\hat{a})\right)\left(\sum_{t=1}^{s-1} (\gamma_t^H - \gamma_t^L)\right)\right].$$

Here, $\beta_s(\cdot)$ is the marginal effect on incentives of an increase in the wage payments for signals above $s-1$. Without loss aversion, i.e., $\lambda = 1$, this expression equals the marginal probability of observing at least signal s. If the agent is loss averse, on the other hand, an

[15] Our notion of the agent's loss premium is highly related to the average self-distance of a lottery defined by Kőszegi and Rabin (2007). Let $D(\boldsymbol{u})$ be the average self-distance of incentive scheme $\boldsymbol{u}$, then $[(\lambda - 1)/2]D(\boldsymbol{u}) = \boldsymbol{b}'\boldsymbol{\rho}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})$.

[16] While the exact change of the loss premium from adding more and more wage payments is hard to grasp, this point can heuristically be illustrated by considering the upper bound of the loss premium. Suppose the principal sets $n \leq S$ different wages. It is readily verified that the loss premium is bounded from above by $(\lambda - 1)[(u_S - u_1)/2] \times [(n-1)/n]$, and that this upper bound increases as n increases. Note, however, that even for $n \to \infty$ the upper bound of the loss premium is finite.

increase in the action also affects the agent's loss premium. The agent's action balances the tradeoff between maximizing intrinsic utility and minimizing the expected net loss. Overall, loss aversion may facilitate as well as hamper the creation of incentives. Let $\boldsymbol{\beta}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) = (\beta_2(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}), \ldots, \beta_S(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}))$.

As in the standard case, incentives are created solely by increases in intrinsic utilities, $\boldsymbol{b}$. In consequence, (IR′) is binding in the optimum. It is obvious that (IC′) can only be satisfied if there exists at least one $\beta_s(\cdot) > 0$. If, for example, signals are ordered according to their likelihood ratios, then $\beta_s(\cdot) > 0$ for all $s = 2, \ldots, S$. More precisely, for a given ordering of signals, under (A2) the following equivalence follows:

$$\beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) > 0 \iff \sum_{\tau=s}^{S} (\gamma_\tau^H - \gamma_\tau^L) > 0 \,. \tag{3}$$

III. The Optimal Contract

In this part of the paper, we first review the standard case where the agent is only risk averse but not loss averse. Thereafter, the case of a loss-averse agent with a risk-neutral intrinsic utility function is analyzed. Last, we discuss the intermediate case of a risk- and loss-averse agent.

A. Pure Risk Aversion

Consider an agent who is risk averse in the usual sense, $h''(\cdot) > 0$, but does not exhibit loss aversion, $\lambda = 1$.

PROPOSITION 1 (Holmström, 1979): *Suppose (A1) holds, $h''(\cdot) > 0$, and $\lambda = 1$. Then there exists a second-best optimal contract to implement $\hat{a} \in (0, 1)$. The second-best contract has the property that $u_s^* \neq u_\tau^* \; \forall s, \tau \in \mathcal{S}$ and $s \neq \tau$. Moreover, $u_s^* > u_\tau^*$ if and only if $\gamma_s^H/\gamma_s^L > \gamma_\tau^H/\gamma_\tau^L$.*

Proposition 1 restates the well-known finding by Bengt Holmström (1979) for discrete signals: signals that are more indicative of higher effort, i.e., signals with a higher likelihood ratio γ_s^H/γ_s^L, are rewarded strictly higher. Thus, the optimal wage scheme is complex in the sense that it is fully contingent, with each signal being rewarded differently.

B. Pure Loss Aversion

We now turn to the other extreme, a purely loss-averse agent. Formally, intrinsic utility of money is a linear function, $h''(\cdot) = 0$, and the agent is loss averse, $\lambda > 1$. Whatever contract

the principal offers, relabeling the signals always allows us to represent this contract as an (at least weakly) increasing intrinsic utility profile. Therefore we can decompose the principal's problem into two steps: first, for a given ordering of signals, choose a nondecreasing profile of intrinsic utility levels that implements the desired action $\hat{a}$ at minimum cost; second, choose the signal ordering with the lowest cost of implementation. As we know from the discussion at the end of the previous section, a necessary condition for an upward-sloping incentive scheme to achieve incentive compatibility is that for the underlying signal ordering at least one $\beta_s(\cdot) > 0$. In what follows, we restrict attention to the set of signal orderings that are incentive feasible in the aforementioned sense. Nonemptiness of this set follows from Lemma 1.

The Optimality of Bonus Contracts.—Consider the first step of the principal's problem, i.e., taking the ordering of signals as given, find the nondecreasing payment scheme with the lowest cost of implementation. With $h(\cdot)$ being linear, the principal's objective function is $C(u_1, \boldsymbol{b}) = u_1 + \sum_{s=2}^{S} b_s(\sum_{\tau=2}^{S} \gamma_\tau(\hat{a}))$. Remember that at the optimum, (IR′) holds with equality. Inserting (IR′) into the principal's objective allows us to write the cost minimization problem for a given order of signals in the following simple way:

Program ML:

$$\min_{\boldsymbol{b} \in \mathbb{R}_+^{S-1}} \ \boldsymbol{b}'\boldsymbol{\rho}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})$$

$$\text{(IC}') \qquad \text{subject to} \quad \boldsymbol{b}'\boldsymbol{\beta}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) = c'(\hat{a}).$$

Intuitively, the principal seeks to minimize the agent's expected net loss. Due to the incentive constraint, however, this loss premium has to be strictly positive.

We want to emphasize that solving Program ML also yields insights for the case with a concave intrinsic utility function. Even though the principal's objective will not reduce to minimizing the agent's loss premium alone, this nevertheless remains an important aspect of her problem. Since the solution to Program ML tells us how to minimize the loss premium irrespective of the functional form of intrinsic utility, one should expect its properties to carry over to some extent to the solution of the more general problem.

The principal's cost minimization problem for a given order of signals is a simple linear programming problem: minimize a linear objective function subject to one linear equality constraint. Since we restricted attention to orderings of signals with $\beta_s(\cdot) > 0$ for at least one signal s, a solution to ML exists. Due to the linear nature of problem ML, (generically) this solution sets exactly one $b_s > 0$ and all other $b_s = 0$. Put differently, the problem is to find that b_s which creates incentives at the lowest cost. What remains to do for the principal, in

a second step, is to find the signal ordering that leads to the lowest cost of implementation; this problem clearly has a solution.

PROPOSITION 2: *Suppose (A1)-(A3) hold,* $h''(\cdot) = 0$ *and* $\lambda > 1$. *Then there exists a second-best optimal contract to implement action* $\hat{a} \in (0,1)$. *Generically, the second-best optimal incentive scheme* $(u_s^*)_{s=1}^S$ *is a bonus contract, i.e.,* $u_s^* = u_H^*$ *for* $s \in \mathcal{B}^* \subset \mathcal{S}$ *and* $u_s^* = u_L^*$ *for* $s \in \mathcal{S} \setminus \mathcal{B}^*$, *where* $u_H^* > u_L^*$.

According to Proposition 2, the principal considers it optimal to offer the agent a bonus contract which entails only a minimum degree of wage differentiation in the sense that the contract specifies only two different wage payments no matter how rich the signal space. This endeavor to reduce the complexity of the contract is plausible since a high degree of wage differentiation increases the loss premium. A loss-averse agent considers a wage schedule as riskier if the average margin between any two wages is higher. The principal can reduce the riskiness of the contract by setting the spread of as many wage pairs as possible equal to zero.

More intuitively, what are the effects of the principal specifying many different wage payments? With a contract specifying many different wages, receiving a relatively low wage feels like a loss when comparing it to possible higher ones, which in turn decreases the agent's utility. Likewise, in the case of obtaining a high wage most comparisons are drawn to lower wages, with the associated gains increasing the agent's utility. Since losses loom larger than gains, anticipating these comparisons ex ante reduces the agent's expected utility and thus a higher average payment is needed to make him accept the contract. In order to avoid these unfavorable comparisons, the principal has an incentive to lump together wages for different signals.

With effort being unobservable but costly for the agent, however, any incentive-compatible contract has to display at least some degree of wage differentiation. Under the standard notion of risk aversion, creating incentives via increasing the utility margin between two signal realizations becomes more and more costly due to the agent's marginal utility of money being decreasing. Loosely speaking, instead of creating incentives via one big bonus payment, provision of incentives is achieved at lower cost by setting many small wage spreads. When facing a purely loss-averse agent, whose marginal intrinsic utility of money is constant, the principal cannot capitalize on differentiating payments according to performance. In this case, pooling together as many wages as possible is beneficial to the principal and thus the optimal contract is a binary payment scheme.

Features of the Optimal Contract.—Up to now we have not specified which signals are generally included in the set $\mathcal{B}^*$. In light of the above observation, the principal's problem

boils down to choosing a binary partition of the set of signals, $\mathcal{B} \subset \mathcal{S}$, which characterizes for which signals the agent receives the high wage and for which signals he receives the low wage. The wages u_L and u_H are then uniquely determined by the corresponding individual rationality and incentive constraints. The problem of choosing the optimal partition of signals, $\mathcal{B}^*$, is an integer programming problem. As is typical for this class of problems, and as is nicely illustrated by the well-known "Knapsack Problem", it is impossible to provide a general characterization of the solution.[17]

Next to these standard intricacies of integer programming, there is an additional difficulty in our model: the principal's objective behaves nonmonotonically when including an additional signal into the "bonus set" $\mathcal{B}$. From Program ML it follows that, for a given bonus set $\mathcal{B}$, the minimum cost of implementing action $\hat{a}$ is

$$C_{\mathcal{B}} = \bar{u} + c(\hat{a}) + \frac{c'(\hat{a})(\lambda - 1)P_{\mathcal{B}}(1 - P_{\mathcal{B}})}{[\sum_{s \in \mathcal{B}}(\gamma_s^H - \gamma_s^L)][1 - (\lambda - 1)(1 - 2P_{\mathcal{B}})]}\ , \tag{4}$$

where $P_{\mathcal{B}} := \sum_{s \in \mathcal{B}} \gamma_s(\hat{a})$. The above costs can be rewritten such that the principal's problem amounts to

$$\max_{\mathcal{B} \subset \mathcal{S}} \left[\sum_{s \in \mathcal{B}}(\gamma_s^H - \gamma_s^L)\right] \left\{\frac{1}{(\lambda - 1)P_{\mathcal{B}}(1 - P_{\mathcal{B}})} - \frac{1}{P_{\mathcal{B}}} + \frac{1}{1 - P_{\mathcal{B}}}\right\}\ . \tag{5}$$

This objective function illustrates the tradeoff that the principal faces. The first term, $\sum_{s \in \mathcal{B}}(\gamma_s^H - \gamma_s^L)$, is the aggregate marginal impact of effort on the probability of the bonus $b := u_H - u_L$ being paid out. In order to create incentives for the agent, the principal would like to make this term as large as possible, which in turn allows her to lower the bonus payment. This can be achieved by including only good signals in $\mathcal{B}$. The second term, on the other hand, is maximized by making the probability of paying the agent the high wage either as large as possible or as small as possible, depending on the exact signal structure and the action to be implemented. Intuitively, by making the event of paying the high wage very likely or unlikely, the principal minimizes the scope for the agent to experience a loss that he demands to be compensated for. These two goals may conflict with each other. Nevertheless, it can be shown that the optimal contract displays the following plausible property.

PROPOSITION 3: *Let $\mathcal{S}^+ \equiv \{s \in \mathcal{S} | \gamma_s^H - \gamma_s^L > 0\}$. The optimal partition of the signals for which the high wage is paid, $\mathcal{B}^*$, has the following property: either $\mathcal{B}^* \subseteq \mathcal{S}^+$ or $\mathcal{S}^+ \subseteq \mathcal{B}^*$.*

Put verbally, the optimal partition of the signal set takes one of the two possible forms: the high wage is paid out to the agent (i) either only for good signals though possibly not for all good signals, or (ii) for all good signals and possibly a few bad signals as well.

[17]The Knapsack Problem refers to a hiker who has to select from a group of items, all of which may be suitable for her trip, a subset that has greatest value while not exceeding the capacity of her knapsack.

Back to the Knapsack Problem, here it is well-established for the continuous version of the problem that the solution can easily be found by ordering the items according to their value-to-weight ratio. Defining $\kappa := \max_{\{s,t\}\subseteq\mathcal{S}} |\gamma_s(\hat{a}) - \gamma_t(\hat{a})|$, we can obtain a similar result. Assuming that κ is sufficiently small, which is likely to hold if the performance measure is, for instance, sales revenues measured in cents, makes the principal's problem of choosing $\mathcal{B}^*$ similar to a continuous problem.[18] With this assumption, we can show that it is optimal to order the signals according to their likelihood ratios.

PROPOSITION 4: *Suppose κ is sufficiently small, then there exists a constant K such that $\mathcal{B}^* = \{s \in \mathcal{S} \mid \gamma_s^H/\gamma_s^L \geq K\}$.*

If one is prepared to assume that higher sales revenues are associated with higher likelihood ratios, then Proposition 4 states that the sales agent receives the bonus only if his sales exceed a previously specified sales quota.

Comparative Statics.—Last, we want to point out the following comparative static results.

PROPOSITION 5: *(i) The minimum cost of implementing action $\hat{a}$ strictly increases in λ. (ii) For a given feasible bonus set $\mathcal{B}$, the wage spread necessary to implement action $\hat{a}$ decreases in λ if and only if $P_{\mathcal{B}} > 1/2$.*

Part (ii) of Proposition 5 relates to the reasoning by Kőszegi and Rabin (2006, 1156) that if the agent is expectation-based loss averse, then "in principal-agent models, performance-contingent pay may not only directly motivate the agent to work harder in pursuit of higher income, but also indirectly motivate [him] by changing [his] expected income and effort." The agent's expected utility comprises of two components, the first of which is expected net intrinsic utility from choosing effort level $\hat{a}$, $u_L^* + b^* \sum_{s\in\mathcal{B}^*} \gamma_s(\hat{a}) - c(\hat{a})$. Due to loss aversion there is a second component since in expectation the agent suffers from deviations from his reference point. A deviation from the agent's reference point occurs with probability $P_{\mathcal{B}^*}(1 - P_{\mathcal{B}^*})$, which we refer to as loss probability. Therefore, when choosing his action, the agent has to balance off two possibly conflicting targets, maximizing expected net intrinsic utility and minimizing the loss probability. The loss probability is locally decreasing at $\hat{a}$ if and only if $P_{\mathcal{B}^*} > 1/2$. In this case, an increase in λ, which makes reducing the loss probability more important, leads to the agent choosing a higher effort level, which in turn allows the principal to use lower-powered incentives. The principal, however, cannot capitalize on this since, according to part (i) of Proposition 5, the overall cost of implementation strictly increases in the agent's degree of loss aversion.

[18]Here, the probability of observing a specific signal, say, sales revenues of exactly \$13,825.32 is rather small.

C. The General Case: Loss Aversion and Risk Aversion

While binary wage schemes based on a rich signal space are hard to reconcile with the orthodox notion of risk aversion, it is well-known that bonus contracts may be optimal if both contracting parties are risk (and loss) neutral. This finding, however, immediately collapses when the agent is somewhat risk averse.[19] As we argue in this section, our finding that under loss aversion the optimal contractual arrangement takes the form of a bonus scheme is robust towards introducing a slightly concave intrinsic utility function. The intuition is as follows: with the intrinsic utility function for money being concave the principal has a classic rationale for rewarding signals that are stronger indicators of good performance strictly higher. Due to the agent being loss averse, however, the principal still has an incentive to lump together wages in order to eliminate the negative comparison effect. If risk aversion is relatively unimportant compared to loss aversion, then this motive outweighs the principal's benefit from differentiating payments according to performance, and the optimal contract is a binary payment scheme. More formal, when the agent's intrinsic utility function becomes close to linearity the risk premium goes to zero, whereas due to loss aversion there are still first-order costs of wage differentiation. While we provide a more thorough discussion as well as a formal proof of this intuition in the Web Appendix, at this point we content ourselves by illustrating this conjecture by means of an example.

Suppose $h(u) = u^r$, with $r > 1$. More precisely, $R = 1 - \frac{1}{r}$ denotes the Arrow-Pratt measure for relative risk aversion of the intrinsic utility function. The agent's effort cost is $c(a) = (1/2)a^2$, the effort level to be implemented is $\hat{a} = 1/2$, and the reservation utility $\bar{u} = 10$. Assume that the agent's performance can take only three values, excellent (E), satisfactory (S) or inadequate (I). Let

$$\begin{array}{lll} \gamma_E^H = 5/10 & \gamma_S^H = 4/10 & \gamma_I^H = 1/10 \\ \gamma_E^L = 1/10 & \gamma_S^L = 3/10 & \gamma_I^L = 6/10. \end{array}$$

It turns out that it is always (weakly) optimal to order signals according to their likelihood ratio, i.e., $u_1 = u_I$, $u_2 = u_S$ and $u_3 = u_E$. The structure of the optimal contract for this specification and various values of r and λ is presented in Table 1. Table 1 suggests that the optimal contract typically involves pooling of the two good signals, in particular when the agent's intrinsic utility is not too concave. Table 1 nicely illustrates the tradeoff the

[19] With both contracting parties being risk neutral, a broad range of contracts—including bonus schemes—is optimal. If the agent is protected by limited liability, Eun-Soo Park (1995), Son Ku Kim (1997), Oyer (2000), and Dominique Demougin and Claude Fluet (1998) show that the unique optimal contract is a bonus scheme. As demonstrated by Ian Jewitt, Ohad Kadan, and Jeroen M. Swinkels (2008), these findings break down if risk aversion is introduced even to the slightest degree.

r \ λ	1.0	1.1	1.3	1.5
1.5	$u_1 < u_2 < u_3$	$u_1 < u_2 = u_3$	$u_1 < u_2 = u_3$	$u_1 < u_2 = u_3$
2	$u_1 < u_2 < u_3$	$u_1 < u_2 < u_3$	$u_1 < u_2 = u_3$	$u_1 < u_2 = u_3$
3	$u_1 < u_2 < u_3$	$u_1 < u_2 < u_3$	$u_1 < u_2 = u_3$	$u_1 < u_2 = u_3$

Table 1: Structure of the optimal contract with two "good" signals.

principal faces when the agent is both risk and loss averse: if the agent becomes more risk averse, pooling is less likely to be optimal. If, on the other hand, he becomes more loss averse, pooling is more likely to be optimal.[20]

IV. Implementation Problems and Stochastic Contracts

In order to explore the implications of a higher degree of loss aversion, we relax assumption (A2), which implies that the first-order approach is not necessarily valid. To ease the exposition, we consider a purely loss-averse agent and restrict attention to binary measures of performance, i.e., $\mathcal{S} = \{1, 2\}$. For notational convenience , let γ^H and γ^L denote the probabilities of observing signal $s = 2$ conditional on B being realized and not being realized, respectively.[21] Thus, the unconditional probability of observing signal $s = 2$ for a given action a is $\gamma(a) \equiv a\gamma^H + (1-a)\gamma^L$. Let $\hat{\boldsymbol{\gamma}} = (\gamma^H, \gamma^L)$. We assume that $s = 2$ is the good signal.

ASSUMPTION (A4): $\quad 1 > \gamma^H > \gamma^L > 0.$

With only two possible signals to be observed, the contract takes the form of a bonus contract: the agent is paid a base wage u if the bad signal is observed, and he is paid the base wage plus a bonus b if the good signal is observed. For now assume that $b \geq 0$.[22] We assume that the agent's intrinsic disutility of effort is a quadratic function, $c(a) = (k/2)a^2$.[23] The first derivative of the agent's expected utility with respect to effort is given by

$$EU'(a) = \underbrace{(\gamma^H - \gamma^L)b\,[2 - \lambda + 2\gamma(a)(\lambda - 1)]}_{MB(a)} - \underbrace{ka}_{MC(a)}. \tag{6}$$

[20]For a given r, the degree of pooling actually may decrease in λ. This can happen, however, only locally: at some point, the degree of pooling increases in λ.

[21]In the notation introduced above, we have $\gamma_1^H = 1 - \gamma^H$, $\gamma_2^H = \gamma^H$, $\gamma_1^L = 1 - \gamma^L$ and $\gamma_2^L = \gamma^L$.

[22]The assumption $b \geq 0$ is made only for expositional purposes, the results hold true for $b \in \mathbb{R}$.

[23]Allowing for more general effort cost functions does not qualitatively change the insights that are to be obtained.

While the marginal cost, $MC(a)$, obviously is a straight line through the origin with slope k, the marginal benefit, $MB(a)$, also is a positively sloped, linear function of effort a. An increase in b unambiguously makes $MB(a)$ steeper. Letting a_0 denote the intercept of $MB(a)$ with the horizontal axis, we have

$$a_0 = \frac{\lambda - 2 - 2\gamma^L(\lambda - 1)}{2(\gamma^H - \gamma^L)(\lambda - 1)}.$$

Implementation problems in our sense refer to a situation where there are actions $a \in (0, 1)$ that are not incentive compatible for any bonus payment.

PROPOSITION 6: *Suppose (A4) holds, then effort level $\hat{a} \in (0, 1)$ is implementable if and only if $a_0 \leq 0$.*

Obviously, implementation problems do not arise when (A2) is satisfied. Implementation problems do occur, however, when $a_0 > 0$, or equivalently, when $\gamma^L < 1/2$ and $\lambda > 2(1 - \gamma^L)/(1 - 2\gamma^L) > 2$. Somewhat surprisingly, this includes performance measures with $\gamma^L < 1/2 < \gamma^H$, which are highly informative. These implementation problems arise because the agent has two possibly conflicting targets: on the one hand, he seeks to maximize net intrinsic utility, $u + b\gamma(a) - (k/2)a^2$, while on the other hand, he wants to minimize the expected loss by choosing an action such that the loss probability, $\gamma(a)(1 - \gamma(a))$, becomes small. For $\gamma^L \geq 1/2$ these targets are perfectly aligned: the loss probability is strictly decreasing in the agent's action, which implies that an increase in the bonus unambiguously increases effort and thus each action $a \in (0, 1)$ is implementable. For $\gamma^L < 1/2$, however, implementation problems do arise when λ is sufficiently large. Roughly speaking, being very loss averse, the agent primarily cares about reducing the loss probability. With the loss probability being inverted U-shaped in this case, the agent achieves this by choosing one of the two extreme actions $a \in \{0, 1\}$, i.e., the principal faces severe implementation problems.

Turning a Blind Eye.—One might wonder if there is a remedy for these implementation problems. The answer is "yes". The principal can manipulate the signal in her favor by not paying attention to the signal from time to time, but nevertheless paying the bonus in these cases. Formally, suppose the principal commits herself to stochastically ignoring the signal with probability $p \in [0, 1)$. Thus, the overall probability of receiving the bonus is given by $\gamma(p, a) \equiv p + (1 - p)\gamma(a)$. This strategic ignorance of information gives rise to a transformed performance measure with $\gamma^H(p) = p + (1 - p)\gamma^H$ and $\gamma^L(p) = p + (1 - p)\gamma^L$ denoting the probabilities that the bonus is paid to the agent conditional on benefit B being realized and not being realized, respectively. We refer to the principal not paying attention to the performance measure as turning a blind eye. It is readily verified that under the transformed

performance measure $\hat{\boldsymbol{\gamma}}(p)$ the intercept of the $MB(a)$ function with the horizontal axis,

$$a_0(p) \equiv \frac{\lambda - 2 - 2\left[p + (1-p)\gamma^L\right](\lambda - 1)}{2(1-p)(\gamma^H - \gamma^L)(\lambda - 1)},$$

not only is decreasing in p but also can be made arbitrarily small, in particular, arbitrarily negative. In the light of Proposition 6, this immediately implies that the principal can eliminate any implementation problems by choosing p sufficiently high.

Besides alleviating possible implementation problems, turning a blind eye can also benefit the principal from a cost perspective. Differentiating the minimum cost of implementing action $\hat{a}$ under the transformed performance measure,

$$(7)\ C(p;\hat{a}) = \overline{u} + \frac{k}{2}\hat{a}^2 + \frac{k\hat{a}(\lambda-1)(1-\gamma(\hat{a}))}{(\gamma^H - \gamma^L)} \frac{\gamma(\hat{a}) + p(1-\gamma(\hat{a}))}{1 - (\lambda-1)\left[1 - 2\gamma(\hat{a}) - 2p(1-\gamma(\hat{a}))\right]},$$

with respect to p reveals that $\text{sign}\{dC(p;\hat{a})/dp\} = \text{sign}\{2-\lambda\}$. Hence, an increase in the probability of ignoring the performance measure decreases the cost of implementing a certain action if and only if $\lambda > 2$. Hence, whenever the principal turns a blind eye in order to remedy implementation problems, she will do so to the largest possible extent.[24]

PROPOSITION 7: *Suppose the principal can commit herself to stochastic ignorance of the signal. Then each action $\hat{a} \in [0,1]$ can be implemented. Moreover, the implementation costs are strictly decreasing in p if and only if $\lambda > 2$.*

To grasp this finding intuitively, remember the intuition underlying Proposition 2: by implementation of a bonus contract, the principal reduces the ex ante probability of the agent incurring a loss by making it more likely that the agent receives what he expects to receive. By the same token, turning a blind eye allows the principal to reduce the agent's loss premium even beyond what is achieved by a deterministic bonus contract. While this reduction comes at the cost of making the performance measure less informative, according to Proposition 7, the positive effect on the agent's loss premium outweighs the negative effect on incentives if the agent is sufficiently loss averse.

We restricted the principal to offer nonstochastic payments conditional on which signal is observed. If the principal was able to do just that, then she could remedy implementation problems by paying the base wage plus a lottery in the case of the bad signal. For instance, when the lottery yields b with probability p and zero otherwise, this is just the same as turning a blind eye. This observation suggests that the principal may benefit from offering a contract

[24]Formally, for $\lambda > 2$, the solution to the principal's problem of choosing the optimal probability to turn a blind eye, p^*, is not well defined because $p^* \to 1$. If the agent is subject to limited liability or if there is a cost of ignorance, however, the optimal probability of turning a blind eye is well defined.

that includes randomization, which is in contrast to the finding under conventional risk aversion, see Holmström (1979).[25] In this sense, while the optimal contract under standard risk aversion would specify only two distinct wages, loss aversion increases the complexity of the optimal contract.

We conclude this section by pointing out an interesting implication of the above analysis. Suppose the principal has no access to a randomization device. Then the above considerations allow a straight-forward comparison of performance measures $\hat{\boldsymbol{\zeta}} = (\zeta^H, \zeta^L)$ and $\hat{\boldsymbol{\gamma}} = (\gamma^H, \gamma^L)$ if $\hat{\boldsymbol{\zeta}}$ is a convex combination of $\hat{\boldsymbol{\gamma}}$ and $\mathbf{1} \equiv (1, 1)$.

COROLLARY 1: *Let $\hat{\boldsymbol{\zeta}} = p\mathbf{1} + (1-p)\hat{\boldsymbol{\gamma}}$ with $p \in (0, 1)$. Then the principal at least weakly prefers performance measure $\hat{\boldsymbol{\zeta}}$ to $\hat{\boldsymbol{\gamma}}$ if and only if $\lambda \geq 2$.*

The finding that the principal prefers the "garbled" performance measure $\hat{\boldsymbol{\zeta}}$ over performance measure $\hat{\boldsymbol{\gamma}}$ is at odds with Blackwell's theorem. While Kim (1995) has already shown that the necessary part of Blackwell's theorem does not hold in the agency model, the sufficiency part was proven to be applicable to the agency framework by Frøystein Gjesdal (1982).[26] Our findings, however, show that the latter is not the case anymore when the agent is loss averse.

V. Alternative Notions of Loss Aversion and Related Literature

With only little being known about how exactly expectations enter into the formation of a person's reference point, a discussion seems warranted to what extent our results depend on the notion of loss aversion according to Kőszegi and Rabin (2007). The agent in our model compares an obtained outcome with all other possible outcomes. This pairwise comparison, which may lead to one and the same outcome being perceived as both a gain and a loss at the same time, is in fact responsible for our main findings.[27] An increase in the margins of payments always increases the agent's expected loss. Even though they are closely related to the CPE concept, this latter effect does not arise under the forward-looking notions of loss aversion according to Bell (1985), Loomes and Sugden (1986), or Gul (1991), which do not allow for mixed feelings. For the sake of argument, consider an agent with linear intrinsic

[25]The finding that stochastic contracts may be optimal is not novel to the principal-agent literature. Hans Haller (1985) shows that in the case of a satisficing agent, who wants to achieve certain aspiration levels of income with certain probabilities, randomization may pay for the principal. Moreover, Roland Strausz (2006) finds that deterministic contracts may be suboptimal in a screening context.

[26]The sufficiency part of Blackwell's theorem states that making use of more informative performance measure implies that the principal is not worse off. See David Blackwell (1951, 1953).

[27]For at least suggestive evidence on mixed feelings, see Jeff T. Larsen et al. (2004).

utility for money who is loss averse in the sense of Bell (1985), i.e., his reference point is the arithmetic mean of the wage distribution. Suppose there are only three signals, $s = 1, 2, 3$, which are equiprobable for the action which is to be implemented. The associated wages are $w_1 < w_2 = w_3 =: w_{23}$. If the principal increases the wage for signal 3 by $\varepsilon > 0$ and reduces the wage for signal 2 by the same amount, then the average payment, and in consequence both the principal's cost and the agent's reference point remain unaffected. Moreover, given that ε is not too large, also the loss premium the principal has to pay turns out to be independent of ε,[28]

$$LP^{Bell}(\varepsilon) = (2/9)(\lambda - 1)(w_{23} - w_1). \tag{8}$$

Thus, the individual rationality constraint also holds under this new contract. Since an increase in the degree of wage differentiation often is accompanied by an improvement of incentives, it is easily imagined that the principal benefits from specifying more than two wages. With loss aversion à la Kőszegi and Rabin, in contrast, the loss premium is strictly increasing in ε:

$$LP^{KR}(\varepsilon) = (2/9)(\lambda - 1)(w_{23} - w_1) + (2/9)(\lambda - 1)\varepsilon. \tag{9}$$

In order to illustrate the differences between these two concepts more vividly, we discuss in more detail the sensations of losses and gains under both concepts. Under Bell's notion of loss aversion, if $s = 1$ occurs, the loss felt is the same under both contracts. For $s = 2$, under the new contract the agent feels a lower gain than under the original contract. This lower gain, however, is exactly offset in expectations by an increased gain for $s = 3$. With loss aversion à la Kőszegi and Rabin, under the new contract, if $s = 1$ is realized then the agent feels a lower loss compared to the outcome for $s = 2$ and a larger loss compared to the outcome for $s = 3$. In expectations, these changes exactly cancel out. For $s = 2$, in contrast, under the new contract the agent now feels a loss in comparison to the outcome for $s = 3$, while under the original contract this comparison did not lead to the sensation of a loss. Thus, under the more differentiated wage scheme, the ex ante probability of incurring a loss is higher, which in turn increases the agent's gain-loss disutility.

The above observations suggest that increasing the degree of wage differentiation always increases the principal's cost if the agent is loss averse à la Kőszegi and Rabin. If, on the other hand, the agent is loss-averse according to Bell, then paying slightly different wages for different signals is costless, except when differentiating wages that originally were equal

[28]The independence of the loss premium of ε does neither rely on the wages being equiprobable nor on using Bell's concept instead of Loomes and Sugden's or Gul's.

to the reference point. With wage differentiation being less costly in the absence of mixed feelings, one would expect the optimal contract to be more differentiated under Bell's notion of loss aversion. Nevertheless, with losses still being painful for the agent, a fully contingent contract, which maximizes the scope for the agent to incur a loss, hardly seems optimal for a rich performance measure even if the agent's reference point has no stochastic component.

This conjecture is highly in line with the extant literature on incentive design under loss aversion.[29] With no unifying approach provided how to determine a decision maker's reference point, it is little surprising that all contributions differ in this aspect. Nevertheless, none of the earlier contributions applies a notion of loss aversion that allows for mixed feelings. David de Meza and David C. Webb (2007) apply the concept of Gul (1991), which posits that the reference point is the certainty equivalent of the prospect and thus is closely related to Bell (1985). The optimal contract consists of three regions: first compensation increases with performance up to the reference point, thereafter for a range of signals the wage equals the reference point, and for high performance the wage is strictly increasing in performance. As an alternative to Gul's concept, de Meza and Webb also consider the median as reference wage, which captures the idea that a loss is incurred at all incomes for which it is odds-on that a higher income would be drawn. Now, the optimal contract is discontinuous after the flat-part, but otherwise qualitatively similar. Thus, the optimal contract derived by de Meza and Webb provides a theoretical underpinning for the usage of option-like incentive schemes in CEO compensation.

Focusing only on gain-loss utility, Emil P. Iantchev (2009) applies the concept of Luis Rayo and Gary S. Becker (2007) to a multi-principal/multi-agent environment in which an agent's reference point is determined by the equilibrium conditions in the market.[30] Next to a dismissal region for very low performance, the optimal contract is found to display a performance-independent flat part for intermediate performance, which is followed by a region where rewards are increasing in performance. Evidence for this theoretically predicted contractual form is shown to be found in panel data from Safelite Glass Corporation.

Also abstracting from intrinsic utility but assuming that the reference point equals previous year's income, Ingolf Dittmann, Ernst Maug, and Oliver G. Spalt (forthcoming) find that a loss aversion model dominates an equivalent risk aversion model in explaining observed CEO compensation contracts. The resulting contract under loss aversion qualitatively resembles

[29]Nonstandard risk preferences different from loss aversion are analyzed in a moral hazard framework by Ulrich Schmidt (1999), who applies Menahem E. Yaari's (1987) concept of dual expected utility theory, and by Ján Zábojník (2002), who incorporates Friedman-Savage utility.

[30]The assumption that only changes in wealth matter is based on Kahneman and Tversky's (1979) original formulation of prospect theory.

the optimal contract identified by Iantchev (2009).

The commonality of all loss aversion concepts, irrespective of mixed-feelings possibly arising or not, is that there typically is a range of signals where payment does not vary with performance.[31] Without mixed feelings, however, the optimal wage schedule displays high sensitivity of pay to performance at least for signals that are very indicative for high effort. Thus, none of the aforementioned papers provides a rationale for the prevalence of binary payment schemes.[32]

To the best of our knowledge, Kohei Daido and Hideshi Itoh (2007) is the only paper that also applies reference dependence à la Kőszegi and Rabin to a principal-agent setting. The focus of Daido and Itoh greatly differs from ours. Assuming that the performance measure comprises of only two signals, two types of self-fulfilling prophecy regarding the impact of expectations on performance are explained. While sufficient to capture these two effects, the assumption of a binary measure of performance does not allow to inquire into the form that contracts take under moral hazard.

Though not placed in the literature on incentive design, the findings in Paul Heidhues and Kőszegi (2005) in spirit are closely related to our results. Here it is shown that consumer loss aversion à la Kőszegi and Rabin can explain why monopoly prices react less sensitively to cost shocks than predicted by orthodox theory. The driving force underlying this price stickiness is the aforementioned comparison effect: the probability of the consumer buying the good at some price is negatively affected by the comparison of this price to lower prices in the distribution. Therefore, just like our principal lumps together wages despite possibly negative incentive effects in order to avoid the unfavorable comparison of some relatively low wage with higher wages, the monopolist has an incentive to lump together prices even though this means foregoing the benefit from differentiating production according to cost. In a similar vein, Heidhues and Kőszegi (2008) provide an answer to the question why nonidentical competitors charge identical prices for differentiated products.

VI. Closing Discussion

In this paper, we explore the implications of loss aversion à la Kőszegi and Rabin (2007) on contract design in the presence of moral hazard. With a stochastic reference point component, increasing the number of different wages increases the agent's gain-loss disutility

[31] Put differently, due to first-order risk aversion Holmström's informativeness principle is violated.

[32] De Meza and Webb (2007) find conditions under which a bonus contract is optimal. For this to be the case, however, they assume that the reference point is exogenously given and that all wage payments are in the loss region, where the agent is assumed to be risk loving.

significantly without necessarily supplying strong additional incentives. The most drastic implication is the use of binary payment schemes even in situations where the principal has access to an arbitrarily rich performance measure and the optimal contract thus would be fully contingent under the standard notion of risk aversion. Moreover, we find that under reasonable conditions the optimal contract needs to specify only a cut-off performance, e.g., a sales quota. Thus, loss aversion provides a theoretical rationale for bonus contracts, the wide application of which is hard to reconcile with obvious drawbacks—such as seasonality effects—that come along with this particular contractual form. In the aforementioned sense loss aversion leads to simpler contracts than predicted by orthodox theory. Reduced complexity of the contract, however, is not a general prediction of loss aversion. We derived circumstances under which the optimal contract consists of stochastic payments if the agent is loss averse but does not include randomization if the agent is risk averse in the usual sense.

We adopted the concept of choice-acclimating personal equilibrium (CPE). Kőszegi and Rabin (2006, 2007) provide another concept, called unacclimating personal equilibrium (UPE). The major difference is the timing of expectation formation and actual decision making.[33] Under UPE a decision maker first forms his expectations, which determine his reference point, and thereafter, given these expectations, chooses his preferred action. To guarantee internal consistency, UPE requires that individuals can only make plans that they will follow through. With expectations being met on the equilibrium path under UPE, the expected utility takes the same form under both concepts. Since the optimality of bonus schemes is rooted in the agent's dislike of being exposed ex ante to numerous outcomes, we would expect bonus contracts to be optimal also under UPE.

Throughout the analysis we ignored diminishing sensitivity of the gain-loss function. A more general gain-loss function complicates the analysis because neither the incentive constraint nor the participation constraint are linear functions in the intrinsic utility levels any longer. Nevertheless, we expect that a reduction of the pay-performance sensitivity will benefit the principal in this case as well. Diminishing sensitivity implies that the sum of two net losses of two monetary outcomes exceeds the net loss of the sum of these two monetary outcomes. Therefore, in addition to the effects discussed in the paper, under diminishing sensitivity there is another channel through which melting two bonus payments into one "big" bonus affects, and in tendency reduces, the agent's expected net loss. There is, however, an argument running counter to this intuition. As we have shown, loss aversion may help the principal to create incentives. Therefore, setting many different wage payments, and thereby—in a sense—creating many kinks, may have favorable incentive effects.

[33] A dynamic model of reference-dependent preferences which allows for changes in beliefs about outcomes is developed in Kőszegi and Rabin (2009).

Last, while several notions of loss aversion proposed in the literature are forward looking in the sense that the reference point is determined by the decision maker's rational expectations, our findings depend on the mixed-feelings approach embodied in the concept of Kőszegi and Rabin. With the exact way of how expectations enter the process of reference point formation being an understudied question, this issue clearly warrants further investigation.

Mathematical Appendix

A. Proofs of Propositions and Lemmas

PROOF OF LEMMA 1:

Suppose that signals are ordered according to their likelihood ratio, that is, $s > s'$ if and only if $\gamma_s^H/\gamma_s^L > \gamma_{s'}^H/\gamma_{s'}^L$. Consider a contract of the form

$$u_s = \begin{cases} \underline{u} & \text{if} \quad s < \hat{s} \\ \underline{u} + b & \text{if} \quad s \geq \hat{s} \end{cases},$$

where $b > 0$ and $1 < \hat{s} \leq S$. Under this contractual form and given that the first-order approach is valid, (IC) can be rewritten as

$$b\Bigg\{ \left[\sum_{s=\hat{s}}^{S}(\gamma_s^H - \gamma_s^L)\right]\left(1 - (\lambda-1)\sum_{s=1}^{\hat{s}-1}\gamma_s(\hat{a})\right) - (\lambda-1)\left(\sum_{s=1}^{\hat{s}-1}(\gamma_s^H - \gamma_s^L)\right)\left(\sum_{s=\hat{s}}^{S}\gamma_s(\hat{a})\right)\Bigg\} = c'(\hat{a}).$$

Since signals are ordered according to their likelihood ratio, we have $\sum_{s=\hat{s}}^{S}(\gamma_s^H - \gamma_s^L) > 0$ and $\sum_{s=1}^{\hat{s}-1}(\gamma_s^H - \gamma_s^L) < 0$ for all $1 < \hat{s} \leq S$. This implies that the term in curly brackets is strictly positive for $\lambda \leq 2$. Hence, with $c'(\hat{a}) > 0$, b can alway be chosen such that (IC) is met. Rearranging the participation constraint,

$$\underline{u} \geq \bar{u} + c(\hat{a}) - b\left(\sum_{s=\hat{s}}^{S}\gamma_s(\hat{a})\right)\left[1 - (\lambda-1)\left(\sum_{s=1}^{\hat{s}-1}\gamma_s(\hat{a})\right)\right],$$

reveals that (IR) can be satisfied for any b by choosing $\underline{u}$ appropriately. This concludes the proof.

PROOF OF PROPOSITION 1:

It is readily verified that Assumptions 1-3 from Grossman and Hart (1983) are satisfied. Thus, the cost-minimization problem is well defined, in the sense that for each action $a \in (0,1)$ there exists a second-best incentive scheme. Suppose the principal wants to implement

action $\hat{a} \in (0,1)$ at minimum cost. Since the agent's action is not observable, the principal's problem is given by

$$\text{(MR)} \qquad \min_{\{u_s\}_{s=1}^S} \sum_{s=1}^S \gamma_s(\hat{a}) h(u_s)$$

subject to

$$\text{(IR}_R\text{)} \qquad \sum_{s=1}^S \gamma_s(\hat{a}) u_s - c(\hat{a}) \geq \bar{u}\,,$$

$$\text{(IC}_R\text{)} \qquad \sum_{s=1}^S (\gamma_s^H - \gamma_s^L) u_s - c'(\hat{a}) = 0\,.$$

where the first constraint is the individual rationality constraint and the second is the incentive compatibility constraint. Note that the first-order approach is valid, since the agent's expected utility is a strictly concave function of his effort. The Lagrangian to the resulting problem is

$$\mathcal{L} \;=\; \sum_{s=1}^S \gamma_s(a) h(u_s) \;-\; \mu_0 \left\{ \sum_{s=1}^S \gamma_s(a) u_s - c(a) - \bar{u} \right\} \;-\; \mu_1 \left\{ \sum_{s=1}^S (\gamma_s^H - \gamma_s^L) u_s - c'(a) \right\},$$

where μ_0 and μ_1 denote the Lagrange multipliers of the individual rationality constraint and the incentive compatibility constraint, respectively. Setting the partial derivative of $\mathcal{L}$ with respect to u_s equal to zero yields

$$\text{(A.1)} \qquad \frac{\partial \mathcal{L}}{\partial u_s} = 0 \iff h'(u_s) = \mu_0 + \mu_1 \frac{\gamma_s^H - \gamma_s^L}{\gamma_s(\hat{a})}, \;\; \forall s \in \mathcal{S}.$$

Irrespective of the value of μ_0, if $\mu_1 > 0$, convexity of $h(\cdot)$ implies that $u_s > u_{s'}$ if and only if $(\gamma_s^H - \gamma_s^L)/\gamma_s(\hat{a}) > (\gamma_{s'}^H - \gamma_{s'}^L)/\gamma_{s'}(\hat{a})$, which in turn is equivalent to $\gamma_s^H/\gamma_s^L > \gamma_{s'}^H/\gamma_{s'}^L$. Thus it remains to show that μ_1 is strictly positive. Suppose, in contradiction, that $\mu_1 \leq 0$. Consider the case $\mu_1 = 0$ first. From (A.1) it follows that $u_s = u^f$ for all $s \in \mathcal{S}$, where u^f satisfies $h'(u^f) = \mu_0$. This, however, violates (IC$_R$), a contradiction. Next, consider $\mu_1 < 0$. From (A.1) it follows that $u_s < u_{s'}$ if and only if $(\gamma_s^H - \gamma_s^L)/\gamma_s(\hat{a}) > (\gamma_{s'}^H - \gamma_{s'}^L)/\gamma_t(\hat{a})$. Let $\mathcal{S}^+ \equiv \{s|\gamma_s^H - \gamma_s^L > 0\}$, $\mathcal{S}^- \equiv \{s|\gamma_s^H - \gamma_s^L < 0\}$, and $\hat{u} \equiv \min\{u_s | s \in \mathcal{S}^-\}$. Since $\hat{u} > u_s$ for

all $s \in \mathcal{S}^+$, we have

$$\begin{aligned}\sum_{s=1}^{S}(\gamma_s^H - \gamma_s^L)u_s &= \sum_{\mathcal{S}^-}(\gamma_s^H - \gamma_s^L)u_s + \sum_{\mathcal{S}^+}(\gamma_s^H - \gamma_s^L)u_s \\ &< \sum_{\mathcal{S}^-}(\gamma_s^H - \gamma_s^L)\hat{u} + \sum_{\mathcal{S}^+}(\gamma_s^H - \gamma_s^L)\hat{u} \\ &= \hat{u}\sum_{s=1}^{S}(\gamma_s^H - \gamma_s^L) \\ &= 0,\end{aligned}$$

again a contradiction to (IC_R). Hence, $\mu_1 > 0$ and the desired result follows.

PROOF OF PROPOSITION 2:

The problem of finding the optimal contract $\boldsymbol{u}^*$ to implement action $\hat{a} \in (0,1)$ is decomposed into two subproblems. First, for a given incentive feasible ordering of signals, we derive the optimal nondecreasing incentive scheme that implements action $\hat{a} \in (0,1)$. Then, in a second step, we choose the ordering of signals for which the ordering specific cost of implementation is lowest.

Step 1: Remember that the ordering of signals is incentive feasible if $\beta_s(\cdot) > 0$ for at least one signal s. For a given incentive feasible ordering of signals, in this first step we solve Program ML. First, note that it is optimal to set $b_s = 0$ if $\beta_s(\cdot) < 0$. To see this, suppose, in contradiction, that in the optimum (IC′) holds and $b_s > 0$ for some signal s with $\beta_s(\cdot) \leq 0$. If $\beta_s(\cdot) = 0$, then setting $b_s = 0$ leaves (IC′) unchanged, but leads to a lower value of the objective function of Program ML, contradicting that the original contract is optimal. If $\beta_s(\cdot) < 0$, then setting $b_s = 0$ not only reduces the value of the objective function, but also relaxes (IC′), which in turn allows to lower other bonus payments, thereby lowering the value of the objective function even further. Again, a contradiction to the original contract being optimal. Let $\mathcal{S}_\beta \equiv \{s \in \mathcal{S} | \beta_s(\cdot) > 0\}$ denote the set of signals for which $\beta_s(\cdot)$ is strictly positive under the considered ordering of signals, and let S_β denote the number of elements in this set. Thus, Program ML can be rewritten as

PROGRAM ML^+:

$$\begin{aligned}&\min_{(b_s)_{s\in\mathcal{S}_\beta}} \sum_{s\in\mathcal{S}_\beta} b_s\rho_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) \\ (\text{IC}^+) \qquad \text{subject to} \quad &(i) \sum_{s\in\mathcal{S}_\beta} b_s\beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) = c'(\hat{a}) \\ &(ii)\ b_s \geq 0, \ \ \forall s \in \mathcal{S}_\beta\ .\end{aligned}$$

Program ML$^+$ is a linear programming problem. It is well-known that if a linear programming problem has a solution, it must have a solution at an extreme point of the constraint set. Generically, there is a unique solution and this solution is an extreme point. Since the constraint set of Program ML$^+$, $\mathcal{M} \equiv \{(b_s)_{s\in\mathcal{S}_\beta} \in \mathbb{R}_+^{S_\beta} | \sum_{s\in\mathcal{S}_\beta} b_s\beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) = c'(\hat{a})\}$, is closed and bounded, Program ML$^+$ has a solution. Hence, generically $\sum_{s\in\mathcal{S}_\beta} b_s\rho_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})$ achieves its greatest lower bound at one of the extreme points of $\mathcal{M}$. (We comment on genericity below.) With $\mathcal{M}$ describing a hyperplane in $\mathbb{R}_+^{S_\beta}$, all extreme points of $\mathcal{M}$ are characterized by the following property: $b_s > 0$ for exactly one signal $s \in S_\beta$ and $b_t = 0$ for all $t \in S_\beta$, $t \neq s$. It remains to determine for which signal the bonus is set strictly positive. The size of the bonus payment, which is set strictly positive, is uniquely determined by (IC$^+$):

$$b_s\beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) = c'(\hat{a}) \iff b_s = \frac{c'(\hat{a})}{\beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})}. \tag{A.2}$$

Therefore, from the objective function of Program ML$^+$ it follows that, for the signal ordering under consideration, the optimal signal for which the bonus is set strictly positive, $\hat{s}$, is characterized by

$$\hat{s} \in \arg\min_{s\in\mathcal{S}_\beta} \frac{c'(\hat{a})}{\beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})}\rho_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}).$$

Step 2: From all incentive feasible signal orders, the principal chooses the one which minimizes her cost of implementation. With the number of incentive feasible signal orders being finite, this problem clearly has a solution. Let s^* denote the resulting cutoff, i.e.,

$$u_s^* = \begin{cases} u^* & \text{if} \quad s < s^* \\ u^* + b^* & \text{if} \quad s \geq s^* \end{cases},$$

where $b^* = c'(\hat{a})/\beta_{s^*}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})$ and $u^* = \bar{u} + c(\hat{a}) - b^*\left[\sum_{\tau=s^*}^{S} \gamma_\tau(\hat{a}) - \rho_{s^*}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})\right]$. Letting $u_L^* = u^*$, $u_H^* = u^* + b^*$, and $\mathcal{B}^* = \{s \in \mathcal{S} | s \geq s^*\}$ establishes the desired result.

On genericity: We claimed that, for any given feasible ordering of signals, generically Program ML$^+$ has a unique solution at one of the extreme points of the constraint set. To see this, note that a necessary condition for the existence of multiple solutions is $\beta_s/\beta_{s'} = \rho_s/\rho_{s'}$ for some $s, s' \in \mathcal{S}_\beta$, $s \neq s'$. This condition is characterized by the action to be implemented, $\hat{a}$, the structure of the performance measure, $\left\{(\gamma_s^H, \gamma_s^L)\right\}_{s=1}^S$, and the agent's degree of loss aversion, λ. Now, fix $\hat{a}$ and $\left\{(\gamma_s^H, \gamma_s^L)\right\}_{s=1}^S$. With both $\beta_s > 0$ and $\rho_s > 0$ for all $s \in \mathcal{S}_\beta$, it is readily verified, that exactly one value of λ equates $\beta_s/\beta_{s'}$ with $\rho_s/\rho_{s'}$. Since λ is drawn from the interval $(1, 2]$, and with the number of signals being finite, this necessary condition for Program ML$^+$ having multiple solutions for a given feasible ordering of signals generically will not hold. With the number of feasible orderings being finite, generic optimality of a corner solution carries over to the overall problem.

PROOF OF PROPOSITION 3:

$\mathcal{B}^*$ maximizes $X(\mathcal{B}) := \left[\sum_{s\in\mathcal{B}}(\gamma_s^H - \gamma_s^L)\right] \times Y(P_\mathcal{B})$, where

$$Y(P_\mathcal{B}) := \frac{1}{(\lambda-1)P_\mathcal{B}(1-P_\mathcal{B})} - \frac{1}{P_\mathcal{B}} + \frac{1}{1-P_\mathcal{B}} .$$

Suppose for the moment that $P_\mathcal{B}$ is a continuous decision variable. Accordingly,

$$\frac{dY(P_\mathcal{B})}{dP_\mathcal{B}} = \frac{1}{P_\mathcal{B}^2(1-P_\mathcal{B})^2}\left[2P_\mathcal{B}^2 + \frac{2-\lambda}{\lambda-1}(2P_\mathcal{B}-1)\right] . \tag{A.3}$$

It is readily verified that $dY(P_\mathcal{B})/dP_\mathcal{B} < 0$ for $0 < P_\mathcal{B} < \bar{P}(\lambda)$ and $dY(P_\mathcal{B})/dP_\mathcal{B} > 0$ for $\bar{P}(\lambda) < P_\mathcal{B} < 1$, where

$$\bar{P}(\lambda) \equiv \frac{\lambda - 2 + \sqrt{\lambda(2-\lambda)}}{2(\lambda-1)} .$$

Note that for $\lambda \leq 2$ the critical value $\bar{P}(\lambda) \in [0, 1/2)$. Hence, excluding a signal of $\mathcal{B}$ increases $Y(P_\mathcal{B})$ if $P_\mathcal{B} < \bar{P}(\lambda)$, whereas including a signal to $\mathcal{B}$ increases $Y(P_\mathcal{B})$ if $P_\mathcal{B} \geq \bar{P}(\lambda)$. With these insights the next two implications follow immediately.

(i) $P_{\mathcal{B}^*} < \bar{P}(\lambda) \implies \mathcal{B}^* \subseteq \mathcal{S}^+$

(ii) $P_{\mathcal{B}^*} \geq \bar{P}(\lambda) \implies \mathcal{S}^+ \subseteq \mathcal{B}^*$

We prove both statements in turn by contradiction. (i) Suppose $P_{\mathcal{B}^*} < \bar{P}(\lambda)$ and that there exists a signal $\hat{s} \in \mathcal{S}^-$ which is also contained in $\mathcal{B}^*$, i.e., $\hat{s} \in \mathcal{B}^*$. Clearly, $\sum_{s\in\mathcal{B}^*}(\gamma_s^H - \gamma_s^L) < \sum_{s\in\mathcal{B}^*\setminus\{\hat{s}\}}(\gamma_s^H - \gamma_s^L)$ because $\hat{s}$ is a bad signal. Moreover, $Y(\mathcal{B}^*) < Y(\mathcal{B}^*\setminus\{\hat{s}\})$ because $Y(\cdot)$ increases when signals are excluded of $\mathcal{B}^*$. Thus $X(\mathcal{B}^*) < X(\mathcal{B}^*\setminus\{\hat{s}\})$, a contradiction to the assumption that $\mathcal{B}^*$ is the optimal partition. (ii) Suppose $P_{\mathcal{B}^*} \geq \bar{P}(\lambda)$ and that there exists a signal $\tilde{s} \in \mathcal{S}^+$ that is not contained in $\mathcal{B}^*$, i.e., $\mathcal{B}^* \cap \{\tilde{s}\} = \emptyset$. Since $\hat{s}$ is a good signal $\sum_{s\in\mathcal{B}^*}(\gamma_s^H - \gamma_s^L) < \sum_{s\in\mathcal{B}^*\cup\{\hat{s}\}}(\gamma_s^H - \gamma_s^L)$. $P_{\mathcal{B}^*} \geq \bar{P}(\lambda)$ implies that $Y(\mathcal{B}^* \cup \{\tilde{s}\}) > Y(\mathcal{B}^*)$. Thus, $X(\mathcal{B}^*) < X(\mathcal{B}^* \cup \{\tilde{s}\})$ a contradiction to the assumption that $\mathcal{B}^*$ maximizes $X(\mathcal{B}^*)$. Finally, since for any $\mathcal{B}^*$ we are either in case (i) or in case (ii), the desired result follows.

PROOF OF PROPOSITION 4:

Suppose, in contradiction, that in the optimum there are signals $s, t \in \mathcal{S}$ such that $s \in \mathcal{B}^*$, $t \notin \mathcal{B}^*$ and $(\gamma_s^H - \gamma_s^L)/\gamma_s(\hat{a}) < (\gamma_t^H - \gamma_t^L)/\gamma_t(\hat{a})$. We derive a contradiction by showing that exchanging signal s for signal t reduces the principal's cost, which implies that the original contract cannot be optimal. Let $\bar{\mathcal{B}} \equiv (\mathcal{B}^* \setminus \{s\}) \cup \{t\}$.

$$\left(\sum_{j\in\mathcal{B}^*}(\gamma_j^H - \gamma_j^L) + (\gamma_t^H - \gamma_t^L) - (\gamma_s^H - \gamma_s^L)\right)\left[\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right] > \left(\sum_{j\in\mathcal{B}^*}(\gamma_j^H - \gamma_j^L)\right)\left[\frac{1-(\lambda-1)(1-2P_{\mathcal{B}^*})}{(\lambda-1)P_{\mathcal{B}^*}(1-P_{\mathcal{B}^*})}\right] .$$

Rearranging yields

$$\text{(A.4)}\quad \left[(\gamma_t^H-\gamma_t^L)-(\gamma_s^H-\gamma_s^L)\right]\left[\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right] > \left(\sum_{j\in\mathcal{B}^*}(\gamma_j^H-\gamma_j^L)\right)\left[\frac{1-(\lambda-1)(1-2P_{\mathcal{B}^*})}{(\lambda-1)P_{\mathcal{B}^*}(1-P_{\mathcal{B}^*})}-\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right].$$

With $Y(P_{\mathcal{B}})$ being defined as in the proof of Proposition 3, we have to consider two cases, (i) $dY(P_{\mathcal{B}^*})/P_{\mathcal{B}}\geq 0$, and (ii) $dY(P_{\mathcal{B}^*})/P_{\mathcal{B}}<0$.

Case (i): Since $\gamma_s(\hat{a})-\gamma_t(\hat{a})\leq\kappa$, we have $P_{\mathcal{B}^*}\leq P_{\bar{\mathcal{B}}}+\kappa$. With $Y(P_{\mathcal{B}})$ being (weakly) increasing at $P_{\mathcal{B}^*}$, inequality (A.4) is least likely to hold for $P_{\mathcal{B}^*}=P_{\bar{\mathcal{B}}}+\kappa$. Inserting $P_{\mathcal{B}^*}=P_{\bar{\mathcal{B}}}+\kappa$ into (A.4) yields

$$\text{(A.5)}\quad \left[(\gamma_t^H-\gamma_t^L)-(\gamma_s^H-\gamma_s^L)\right]\left[\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right] > \left(\sum_{j\in\mathcal{B}^*}(\gamma_j^H-\gamma_j^L)\right)\left[\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}}-2\kappa)}{(\lambda-1)[P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})+\kappa(1-2P_{\bar{\mathcal{B}}})-\kappa^2]}-\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right].$$

The right-hand side of (A.5) becomes arbitrarily close to zero for $\kappa\to 0$, thus it remains to show that

$$\text{(A.6)}\qquad \left[(\gamma_t^H-\gamma_t^L)-(\gamma_s^H-\gamma_s^L)\right]\left[\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right]>0\,.$$

For (A.6) to hold, we must have $(\gamma_t^H-\gamma_t^L)-(\gamma_s^H-\gamma_s^L)>0$. From the proof of Proposition 3 we know that $\mathcal{S}^+\subseteq\mathcal{B}^*$ if $Y(P_{\mathcal{B}})$ is increasing at $\mathcal{B}^*$. Since the principal will end up including all good signals in the set $\mathcal{B}^*$ anyway, the question of interest is whether she can benefit from swapping two bad signals. Therefore, we consider case $s,t\in\mathcal{S}^-$, where $\mathcal{S}^-\equiv\{s\in\mathcal{S}|\gamma_s^H-\gamma_s^L<0\}$. With $s,t\in\mathcal{S}^-$, we have

$$\text{(A.7)}\left[(\gamma_t^H-\gamma_t^L)-(\gamma_s^H-\gamma_s^L)\right]\geq\gamma_t(\hat{a})\gamma_s(\hat{a})\left[\frac{1}{\gamma_s(\hat{a})}\frac{\gamma_t^H-\gamma_t^L}{\gamma_t(\hat{a})}-\frac{1}{\gamma_s(\hat{a})+\kappa}\frac{\gamma_s^H-\gamma_s^L}{\gamma_s(\hat{a})}\right],$$

where the inequality holds because $\gamma_t(\hat{a})-\gamma_s(\hat{a})\leq\kappa$. Note that for $\kappa\to 0$ the right-hand side of (A.7) becomes strictly positive, thus $(\gamma_t^H-\gamma_t^L)-(\gamma_s^H-\gamma_s^L)>0$ for $\kappa\to 0$. Hence, for κ sufficiently small, $X(\mathcal{B}^*)<X(\bar{\mathcal{B}})$, a contradiction to $\mathcal{B}^*$ being optimal.

Case (ii): Since $\gamma_t(\hat{a})-\gamma_s(\hat{a})\leq\kappa$, we have $P_{\mathcal{B}^*}\geq P_{\bar{\mathcal{B}}}-\kappa$. With $Y(P_{\mathcal{B}})$ being decreasing at $P_{\mathcal{B}^*}$, inequality (A.4) is least likely to hold for $P_{\mathcal{B}^*}=P_{\bar{\mathcal{B}}}-\kappa$. Inserting $P_{\mathcal{B}^*}=P_{\bar{\mathcal{B}}}-\kappa$ into (A.4), and running along the lines of case (i) allows us to establish that, for κ sufficiently small, $X(\mathcal{B}^*)<X(\bar{\mathcal{B}})$, a contradiction to $\mathcal{B}^*$ being optimal.

To sum up, for κ sufficiently small we have

$$\max_{s\in\mathcal{S}\setminus\mathcal{B}^*}\{(\gamma_s^H-\gamma_s^L)/\gamma_s(\hat{a})\}<\min_{s\in\mathcal{B}^*}\{(\gamma_s^H-\gamma_s^L)/\gamma_s(\hat{a})\}\ ,$$

or equivalently, $\max_{s\in\mathcal{S}\setminus\mathcal{B}^*}\{\gamma_s^H/\gamma_s^L\}<\min_{s\in\mathcal{B}^*}\{\gamma_s^H/\gamma_s^L\}=:K$, which establishes the result.

PROOF OF PROPOSITION 5:

We first prove part (ii). Consider a feasible partition $\mathcal{B}$. The corresponding bonus to implement $\hat{a}$ is given by

$$b=\frac{c'(\hat{a})}{\sum_{s\in\mathcal{B}}(\gamma_s^H-\gamma_s^L)-(\lambda-1)\left[\sum_{s\in\mathcal{B}}(\gamma_s^H-\gamma_s^L)\right][1-2P_{\mathcal{B}}]}. \tag{A.8}$$

Straight-forward differentiation reveals that

$$\frac{db}{d\lambda}=\frac{c'(\hat{a})\left[\sum_{s\in\mathcal{B}}(\gamma_s^H-\gamma_s^L)\right][1-2P_{\mathcal{B}}]}{\left\{\sum_{s\in\mathcal{B}}(\gamma_s^H-\gamma_s^L)-(\lambda-1)\left[\sum_{s\in\mathcal{B}}(\gamma_s^H-\gamma_s^L)\right][1-2P_{\mathcal{B}}]\right\}^2}.$$

Since for a feasible partition $\sum_{s\in\mathcal{B}}(\gamma_s^H-\gamma_s^L)>0$, the desired result follows.

To prove part (i), let $\mathcal{B}^+\equiv\left\{\mathcal{B}\subset\mathcal{S}|\sum_{s\in\mathcal{B}}(\gamma_s^H-\gamma_s^L)>0\right\}$. For any $\tilde{\mathcal{B}}\in\mathcal{B}^+$, let

$$b_{\tilde{\mathcal{B}}}=\frac{c'(\hat{a})}{\sum_{s\in\tilde{\mathcal{B}}}(\gamma_s^H-\gamma_s^L)-(\lambda-1)\left[\sum_{s\in\tilde{\mathcal{B}}}(\gamma_s^H-\gamma_s^L)\right][1-2P_{\tilde{\mathcal{B}}}]}$$

and

$$\underline{u}_{\tilde{\mathcal{B}}}=\bar{u}+c(\hat{a})-b_{\tilde{\mathcal{B}}}P_{\tilde{\mathcal{B}}}+(\lambda-1)P_{\tilde{\mathcal{B}}}(1-P_{\tilde{\mathcal{B}}})b_{\tilde{\mathcal{B}}}.$$

The cost of implementing action $\hat{a}$ when paying $\underline{u}_{\tilde{\mathcal{B}}}$ for signals in $\mathcal{S}\setminus\tilde{\mathcal{B}}$ and $\underline{u}_{\tilde{\mathcal{B}}}+b_{\tilde{\mathcal{B}}}$ for signals in $\tilde{\mathcal{B}}$ is given by

$$C_{\tilde{\mathcal{B}}}\ =\ \underline{u}_{\tilde{\mathcal{B}}}+b_{\tilde{\mathcal{B}}}P_{\tilde{\mathcal{B}}}\ =\ \bar{u}+c(\hat{a})+\frac{c'(\hat{a})(\lambda-1)P_{\tilde{\mathcal{B}}}(1-P_{\tilde{\mathcal{B}}})}{\left[\sum_{s\in\tilde{\mathcal{B}}}(\gamma_s^H-\gamma_s^L)\right][1-(\lambda-1)(1-2P_{\tilde{\mathcal{B}}})]}. \tag{A.9}$$

Differentiation of $C_{\tilde{\mathcal{B}}}$ with respect to λ yields

$$\frac{dC_{\tilde{\mathcal{B}}}}{d\lambda}=\frac{c'(\hat{a})P_{\tilde{\mathcal{B}}}(1-P_{\tilde{\mathcal{B}}})}{\left[\sum_{s\in\tilde{\mathcal{B}}}(\gamma_s^H-\gamma_s^L)\right][1-(\lambda-1)(1-2P_{\tilde{\mathcal{B}}})]^2}.$$

Obviously, $dC_{\mathcal{B}}/d\lambda>0$ for all $\mathcal{B}\in\mathcal{B}^+$. Since the optimal partition of $\mathcal{S}$ may change as λ changes, the minimum cost of implementing action $\hat{a}$ is given by

$$C(\hat{a})=\min_{\mathcal{B}\in\mathcal{B}^+}C_{\mathcal{B}}.$$

Put differently, $C(\hat{a})$ is the lower envelope of all $C_{\mathcal{B}}$ for $\mathcal{B}\in\mathcal{B}^+$. With $C_{\mathcal{B}}$ being continuous and strictly increasing in λ for all $\mathcal{B}\in\mathcal{B}^+$, it follows that also $C(\hat{a})$ is continuous and strictly increasing in λ. This completes the proof.

PROOF OF PROPOSITION 6:

First consider $b \geq 0$. We divide the analysis for $b \geq 0$ into three subcases.

Case 1 ($a_0 < 0$): For the effort level $\hat{a}$ to be chosen by the agent, this effort level has to satisfy the following incentive compatibility constraint:

$$\text{(IC)} \qquad \hat{a} \in \arg\max_{a \in [0,1]} \; u + \gamma(a)b - \gamma(a)(1-\gamma(a))b(\lambda - 1) - \frac{k}{2}a^2.$$

For $\hat{a}$ to be a zero of $dEU(a)/da$, the bonus has to be chosen according to

$$b^*(\hat{a}) = \frac{k\hat{a}}{(\gamma^H - \gamma^L)\left[2 - \lambda + 2\gamma(\hat{a})(\lambda - 1)\right]}.$$

Since $a_0 < 0$, $b^*(a)$ is a strictly increasing and strictly concave function with $b^*(0) = 0$. Hence, each $\hat{a} \in [0,1]$ can be made a zero of $dEU(a)/da$ with a nonnegative bonus. By choosing the bonus according to $b^*(\hat{a})$, $\hat{a}$ satisfies, by construction, the first-order condition. Inserting $b^*(\hat{a})$ into $d^2EU(a)/da^2$ shows that expected utility is strictly concave function if $a_0 < 0$. Hence, with the bonus set equal to $b^*(\hat{a})$, effort level $\hat{a}$ satisfies the second-order condition for optimality and therefore is incentive compatible.

Case 2 ($a_0 = 0$): Just like in the case where $a_0 < 0$, each effort level $a \in [0,1]$ turns out to be implementable with a nonnegative bonus. To see this, consider bonus

$$b_0 = \frac{k}{2(\gamma^H - \gamma^L)^2(\lambda - 1)}.$$

For $b < b_0$, $dEU(a)/da < 0$ for each $a > 0$, that is, lowering effort increases expected utility. Hence, the agent wants to choose an effort level as low as possible and therefore exerts no effort at all. If, on the other hand, $b > b_0$, then $dEU(a)/da > 0$. Now, increasing effort increases expected utility, and the agent wants to choose effort as high as possible. For $b = b_0$, expected utility is constant over all $a \in [0,1]$, that is, as long as his participation constraint is satisfied, the agent is indifferent which effort level to choose. As a tie-breaking rule we assume that, if indifferent between several effort levels, the agent chooses the effort level that the principal prefers.

Case 3 ($a_0 > 0$): If $a_0 > 0$, the agent either chooses $a = 0$ or $a = 1$. To see this, again consider bonus b_0. For $b \leq b_0$, $dEU(a)/da < 0$ for each $a > 0$. Hence, the agent wants to exert as little effort as possible and chooses $a = 0$. If, on the other hand, $b > b_0$, then $d^2EU(a)/da^2 > 0$, that is, expected utility is a strictly convex function of effort. In order to maximize expected utility, the agent will choose either $a = 0$ or $a = 1$ depending on whether $EU(0)$ exceeds $EU(1)$ or not.

Negative Bonus: $b < 0$

Let $b^- < 0$ denote the monetary punishment that the agent receives if the good signal is observed. With a negative bonus, the agent's expected utility is

$$EU(a) = u + \gamma(a)b^- + \gamma(a)(1-\gamma(a))\lambda b^- + (1-\gamma(a))\gamma(a)(-b^-) - \frac{k}{2}a^2. \tag{A.10}$$

The first derivative with respect to effort,

$$\frac{dEU(a)}{da} = \underbrace{(\gamma^H - \gamma^L)b^- \left[\lambda - 2\gamma(a)(\lambda - 1)\right]}_{MB^-(a)} - \underbrace{ka}_{MC(a)},$$

reveals that $MB^-(a)$ is a positively sloped function, which is steeper the harsher the punishment is, that is, the more negative b^- is. It is worthwhile to point out that if bonus and punishment are equal in absolute value, $|b^-| = b$, then also the slopes of $MB^-(a)$ and $MB(a)$ are identical. The intercept of $MB^-(a)$ with the horizontal axis, a_0^- again is completely determined by the model parameters:

$$a_0^- = \frac{\lambda - 2\gamma^L(\lambda - 1)}{2(\gamma^H - \gamma^L)(\lambda - 1)}.$$

Note that $a_0^- > 0$ for $\gamma^L \leq 1/2$. For $\gamma^L > 1/2$ we have $a_0^- < 0$ if and only if $\lambda > 2\gamma^L/(2\gamma^L - 1)$. Proceeding in exactly the same way as in the case of a nonnegative bonus yields a familiar results: effort level $\hat{a} \in [0,1]$ is implementable with a strictly negative bonus if and only if $a_0^- \leq 0$. Finally, note that $a_0 < a_0^-$. Hence a negative bonus does not improve the scope for implementation.

PROOF OF PROPOSITION 7:

Throughout the analysis we restricted attention to nonnegative bonus payment. It remains to be shown that the principal cannot benefit from offering a negative bonus payment: implementing action $\hat{a}$ with a negative bonus is at least as costly as implementing action $\hat{a}$ with a positive bonus. In what follows, we make use of notation introduced in the paper as well as in the proof of Proposition 6. Let $a_0(p)$, $a_0^-(p)$, $b^*(p;\hat{a})$, and $u^*(p;\hat{a})$ denote the expressions obtained from a_0, a_0^-, $b^*(\hat{a})$, and $u^*(\hat{a})$, respectively, by replacing $\gamma(\hat{a})$, γ^L, and γ^H with $\gamma(p,\hat{a})$, $\gamma^L(p)$, and $\gamma^H(p)$. From the proof of Proposition 6 we know that (i) action $\hat{a}$ is implementable with a nonnegative bonus (negative bonus) if and only if $a_0(p) \leq 0$ ($a_0^-(p) \leq 0$), and (ii) $a_0^-(p) \leq 0$ implies $a_0(p) < 0$. We will show that, for a given value of p, if $\hat{a}$ is implementable with a negative bonus then it is less costly to implement $\hat{a}$ with a nonnegative bonus.

Consider first the case where $a_0^-(p) < 0$. The negative bonus payment satisfying incentive compatibility is given by

$$b^-(p;\hat{a}) = \frac{k\hat{a}}{(\gamma^H(p) - \gamma^L(p))\left[\lambda - 2\gamma(p,\hat{a})(\lambda - 1)\right]}.$$

It is easy to verify that the required punishment to implement $\hat{a}$ is larger in absolute value than than the respective nonnegative bonus which is needed to implement $\hat{a}$, that is, $b^*(p;\hat{a}) < |b^-(p;\hat{a})|$ for all $\hat{a} \in (0,1)$ and all $p \in [0,1)$. When punishing the agent with a negative bonus $b^-(p;\hat{a})$, $u^-(p;\hat{a})$ will be chosen to satisfy the corresponding participation constraint with equality, that is,

$$u^-(p;\hat{a}) = \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(p,\hat{a})b^-(p;\hat{a})\left[\lambda - \gamma(p,\hat{a})(\lambda - 1)\right].$$

Remember that, if $\hat{a}$ is implemented with a nonnegative bonus, we have

$$u^*(p;\hat{a}) = \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(p,\hat{a})b^*(p;\hat{a})\left[2 - \lambda + \gamma(p,\hat{a})(\lambda - 1)\right].$$

It follows immediately that the minimum cost of implementing $\hat{a}$ with a nonnegative bonus is lower than the minimum implementation cost with a strictly negative bonus:

$$\begin{aligned}
C^-(p;\hat{a}) &= u^-(p;\hat{a}) + \gamma(p,\hat{a})b^-(p;\hat{a}) \\
&= \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(p,\hat{a})b^-(p;\hat{a})\left[\lambda - \gamma(p,\hat{a})(\lambda - 1) - 1\right] \\
&> \bar{u} + \frac{k}{2}\hat{a}^2 + \gamma(p,\hat{a})b^*(p;\hat{a})\left[\lambda - \gamma(p,\hat{a})(\lambda - 1) - 1\right] \\
&= \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(p,\hat{a})b^*(p;\hat{a})\left[1 - \lambda + \gamma(p,\hat{a})(\lambda - 1)\right] \\
&= \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(p,\hat{a})b^*(p;\hat{a})\left[2 - \lambda + \gamma(p,\hat{a})(\lambda - 1)\right] + \gamma(p,\hat{a})b^*(p;\hat{a}) \\
&= u^*(p;\hat{a}) + \gamma(p,\hat{a})b^*(p;\hat{a}) \\
&= C(p;\hat{a}).
\end{aligned}$$

The same line of argument holds when $a_0^- = 0$: the bonus which satisfies the (IC) is

$$b_0^-(p;\hat{a}) = -\frac{k}{2(\gamma^H(p) - \gamma^L(p))^2(\lambda - 1)},$$

and so $b^*(p;\hat{a}) < |b_0^-(p;\hat{a})|$ for all $\hat{a} \in (0,1)$ and all $p \in [0,1)$.

PROOF OF COROLLARY 1:

Let $p \in (0,1)$. With $\hat{\boldsymbol{\zeta}}$ being a convex combination of $\hat{\boldsymbol{\gamma}}$ and $\mathbf{1}$ we have $(\zeta^H, \zeta^L) = p(1,1) + (1-p)(\gamma^H, \gamma^L) = (\gamma^H + p(1-\gamma^H), \gamma^L + p(1-\gamma^L))$. The desired result follows

immediately from Proposition 7. Consider $\lambda > 2$. Implementation problems are less likely to be encountered under $\hat{\boldsymbol{\zeta}}$ than under $\hat{\boldsymbol{\gamma}}$. Moreover, if implementation problems are not an issue under both performance measures, then implementation of a certain action is less costly under $\hat{\boldsymbol{\zeta}}$ than under $\hat{\boldsymbol{\gamma}}$. For $\lambda = 2$ implementation problems do not arise and implementation costs are identical under both performance measures. Last, if $\lambda < 2$, implementation problems are not an issue under either performance measure, but the cost of implementation is strictly lower under $\hat{\boldsymbol{\gamma}}$ than under $\hat{\boldsymbol{\zeta}}$.

B. Validity of the First-Order Approach

LEMMA 2: *Suppose (A1)-(A3) hold, then the incentive constraint in the principal's cost minimization problem can be represented as* $EU'(\hat{a}) = 0$.

PROOF:

Consider a contract $(u_1, (b_s)_{s=2}^S)$ with $b_s \geq 0$ for $s = 2, \ldots, S$. In what follows, we write β_s instead of $\beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})$ to cut back on notation. The proof proceeds in two steps. First, for a given contract with the property $b_s > 0$ only if $\beta_s > 0$, we show that all actions that satisfy the first-order condition of the agent's utility maximization problem characterize a local maximum of his utility function. Since the utility function is twice continuously differentiable and all extreme points are local maxima, if there exists some action that fulfills the first-order condition, this action corresponds to the unique maximum. In the second step we show that under the optimal contract we cannot have $b_s > 0$ if $\beta_s \leq 0$.

Step 1: The second derivative of the agent's utility with respect to a is

$$EU''(a) = -2(\lambda - 1)\sum_{s=2}^{S} b_s \sigma_s - c''(a) \;, \tag{B.1}$$

where $\sigma_s := (\sum_{i=1}^{s-1}(\gamma_i^H - \gamma_i^L))(\sum_{i=s}^{S}(\gamma_i^H - \gamma_i^L)) < 0$. Suppose action $\hat{a}$ satisfies the first-order condition. Formally

$$\sum_{s=2}^{S} b_s \beta_s = c'(\hat{a}) \iff \sum_{s=2}^{S} b_s \frac{\beta_s}{\hat{a}} = \frac{c'(\hat{a})}{\hat{a}} \;. \tag{B.2}$$

Action $\hat{a}$ locally maximizes the agent's utility if

$$-2(\lambda - 1)\sum_{s=2}^{S} b_s \sigma_s < c''(\hat{a}) \;. \tag{B.3}$$

Under Assumption (A3), we have $c''(\hat{a}) > c'(\hat{a})/\hat{a}$. Therefore, if

$$\sum_{s=2}^{S} b_s \big[-2(\lambda - 1)\sigma_s - \beta_s/\hat{a}\big] < 0 \;, \tag{B.4}$$

then (B.2) implies (B.3), and each action $\hat{a}$ satisfying the first-order condition of the agent's maximization problem is a local maximum of his expected utility. Inequality (B.4) obviously is satisfied if each element of the sum is negative. Summand s is negative if and only if

$$
\begin{aligned}
&-2(\lambda-1)\left(\sum_{i=1}^{s-1}(\gamma_i^H-\gamma_i^L)\right)\left(\sum_{i=s}^{S}(\gamma_i^H-\gamma_i^L)\right)\hat{a} \\
&-\left(\sum_{\tau=s}^{S}(\gamma_\tau^H-\gamma_\tau^L)\right)\left[1-(\lambda-1)\left(\sum_{t=1}^{s-1}\gamma_t(\hat{a})\right)\right]+(\lambda-1)\left[\sum_{\tau=s}^{S}\gamma_\tau(\hat{a})\right]\left(\sum_{t=1}^{s-1}(\gamma_t^H-\gamma_t^L)\right)<0\,.
\end{aligned}
$$

Rearranging the above inequality yields

$$
\begin{aligned}
&\left(\sum_{i=s}^{S}(\gamma_i^H-\gamma_i^L)\right)\left\{\lambda+2(\lambda-1)\left[\hat{a}\sum_{i=1}^{s-1}(\gamma_i^H-\gamma_i^L)-\sum_{i=1}^{s-1}\gamma_i(\hat{a})\right]\right\} \;>\; 0 \\
\iff\quad &\left(\sum_{i=s}^{S}(\gamma_i^H-\gamma_i^L)\right)\left\{\lambda\left(1-\sum_{i=1}^{s-1}\gamma_i^L\right)+(2-\lambda)\sum_{i=1}^{s-1}\gamma_i^L\right\} \;>\; 0. \qquad \text{(B.5)}
\end{aligned}
$$

The term in curly brackets is positive, since $\lambda \leq 2$ and $\sum_{i=1}^{s-1}\gamma_i^L < 1$. Furthermore, note that $\sum_{i=s}^{S}(\gamma_i^H-\gamma_i^L)>0$ since $\beta_s>0$ for all $b_s>0$. This completes the first step of the proof.

Step 2: Consider a contract with $b_s>0$ and $\beta_s\leq 0$ for at least one signal $s\in\{2,\ldots,S\}$ that implements $\hat{a}\in(0,1)$. Then, under this contract, (IC$'$) is satisfied and there exists at least one signal t with $\beta_t>0$ and $b_t>0$. Obviously, the principal can reduce both b_s and b_t without violating (IC$'$). This reasoning goes through up to the point where (IC$'$) is satisfied and $b_s=0$ for all signals s with $\beta_s\leq 0$. From the first step of the proof we know that the resulting contract implements $\hat{a}$ incentive compatibly. Next, we show that reducing any spread, say b_k, always reduces the principal's cost of implementation.

$$
\begin{aligned}
C(\boldsymbol{b}) \;&=\sum_{s=1}^{S}\gamma_s(\hat{a})h\left(u_1(\boldsymbol{b})+\sum_{t=2}^{s}b_t\right)\,, \qquad \text{(B.6)}\\
\text{where}\qquad u_1(\boldsymbol{b})&=\bar{u}+c(\hat{a})-\sum_{s=2}^{S}b_s\left[\sum_{\tau=s}^{S}\gamma_\tau(\hat{a})-(\lambda-1)\left(\sum_{\tau=s}^{S}\gamma_\tau(\hat{a})\right)\left(\sum_{t=1}^{s-1}\gamma_t(\hat{a})\right)\right]\,.
\end{aligned}
$$

The partial derivative of the cost function with respect to an arbitrary b_k is

$$
\frac{\partial C(\boldsymbol{b})}{\partial b_k}=\sum_{s=1}^{k-1}\gamma_s(\hat{a})h'\left(u_1(\boldsymbol{b})+\sum_{t=2}^{s}b_t\right)\left[\frac{\partial u_1}{\partial b_k}\right]+\sum_{s=k}^{S}\gamma_s(\hat{a})h'\left(u_1(\boldsymbol{b})+\sum_{t=2}^{s}b_t\right)\left[\frac{\partial u_1}{\partial b_k}+1\right]\,.
$$

Rearranging yields

$$\text{(B.7)}\quad \frac{\partial C(\boldsymbol{b})}{\partial b_k} = \sum_{s=1}^{k-1} \gamma_s(\hat{a}) h'(u_s) \underbrace{\left[(\lambda - 1) \left(\sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) \right) \left(\sum_{t=1}^{k-1} \gamma_t(\hat{a}) \right) - \sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) \right]}_{<0} + \sum_{s=k}^{S} \gamma_s(\hat{a}) h'(u_s) \underbrace{\left[(\lambda - 1) \left(\sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) \right) \left(\sum_{t=1}^{k-1} \gamma_t(\hat{a}) \right) - \sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) + 1 \right]}_{>0} .$$

Note $u_s \leq u_{s+1}$ which implies that $h'(u_s) \leq h'(u_{s+1})$. Thus, the following inequality holds

$$\begin{aligned}\text{(B.8)}\quad \frac{\partial C(\boldsymbol{b})}{\partial b_k} \geq & \sum_{s=1}^{k-1} \gamma_s(\hat{a}) h'(u_k) \left[(\lambda - 1) \left(\sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) \right) \left(\sum_{t=1}^{k-1} \gamma_t(\hat{a}) \right) - \sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) \right] \\ & + \sum_{s=k}^{S} \gamma_s(\hat{a}) h'(u_k) \left[(\lambda - 1) \left(\sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) \right) \left(\sum_{t=1}^{k-1} \gamma_t(\hat{a}) \right) - \sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) + 1 \right] .\end{aligned}$$

The above inequality can be rewritten as follows

$$\frac{\partial C(\boldsymbol{b})}{\partial b_k} \geq h'(u_k) \left[(\lambda - 1) \left(\sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) \right) \left(\sum_{t=1}^{k-1} \gamma_t(\hat{a}) \right) \right] > 0 .$$

Since reducing any bonus lowers the principal's cost of implementation, it cannot be optimal to set $b_s > 0$ for $\beta_s \leq 0$. This completes the second step of the proof. In combination with Step 1, this establishes the desired result.

C. The General Case: Loss Aversion and Risk Aversion

In this part of the Web Appendix we provide a thorough discussion of the intermediate case where the agent is both risk and loss averse. The agent's intrinsic utility for money is a strictly increasing and strictly concave function, which implies that $h(\cdot)$ is strictly increasing and strictly convex. Moreover, the agent is loss averse, i.e., $\lambda > 1$. From Lemma 1, we know that the constraint set of the principal's problem is nonempty. By relabeling signals, each contract can be interpreted as a contract that offers the agent a (weakly) increasing intrinsic utility profile. This allows us to assess whether the agent perceives receiving u_s instead of u_t as a gain or a loss. As in the case of pure loss aversion, we analyze the optimal contract for a given feasible ordering of signals.

The principal's problem for a given arrangement of the signals is given by

PROGRAM MG:

$$\min_{u_1,\ldots,u_S} \sum_{s=1}^{S} \gamma_s(\hat{a})h(u_s)$$

subject to

$$(\mathrm{IR}_G) \qquad \sum_{s=1}^{S} \gamma_s(\hat{a})u_s - (\lambda-1)\sum_{s=1}^{S-1}\sum_{t=s+1}^{S} \gamma_s(\hat{a})\gamma_t(\hat{a})[u_t - u_s] - c(\hat{a}) = \bar{u} \ ,$$

$$\sum_{s=1}^{S}(\gamma_s^H - \gamma_s^L)u_s -$$

$$(\mathrm{IC}_G) \qquad (\lambda-1)\sum_{s=1}^{S-1}\sum_{t=s+1}^{S} \left[\gamma_s(\hat{a})(\gamma_t^H - \gamma_t^L) + \gamma_t(\hat{a})(\gamma_s^H - \gamma_s^L)\right][u_t - u_s] = c'(\hat{a}) \ ,$$

$$(\mathrm{OC}_G) \qquad u_S \geq u_{S-1} \geq \ldots \geq u_1 \ .$$

Since the objective function is strictly convex and the constraints are all linear in $\boldsymbol{u} = \{u_1, \ldots, u_S\}$, the Kuhn-Tucker theorem yields necessary and sufficient conditions for optimality. Put differently, if there exists a solution to the problem (MG) the solution is characterized by the partial derivatives of the Lagrangian associated with (MG) set equal to zero.

LEMMA 3: *Suppose (A1)-(A3) hold and $h''(\cdot) > 0$, then there exists a second-best optimal incentive scheme for implementing action $\hat{a} \in (0,1)$, denoted $\boldsymbol{u}^* = (u_1^*, \ldots, u_S^*)$.*

PROOF:

We show that program (MG) has a solution, i.e., $\sum_{s=1}^{S} \gamma_s(\hat{a})h(u_s)$ achieves its greatest lower bound. First, from Lemma 1 we know that the constraint set of program (MG) is not empty for action $\hat{a} \in (0,1)$. Next, note that from (IR_G) it follows that $\sum_{s=1}^{S} \gamma_s(\hat{a})u_s$ is bounded below. Following the reasoning in the proof of Proposition 1 of Grossman and Hart (1983), we can artificially bound the constraint set—roughly spoken because unbounded sequences in the constraint set make $\sum_{s=1}^{S} \gamma_s(\hat{a})h(u_s)$ tend to infinity by a result from Dimitri Bertsekas (1974). Since the constraint set is closed, the existence of a minimum follows from Weierstrass' theorem.

In order to interpret the first-order conditions of the Lagrangian to problem (MG) it is necessary to know whether the Lagrangian multipliers are positive or negative.

LEMMA 4: *The Lagrangian multipliers of program (MG) associated with the incentive compatibility constraint and the individual rationality constraint are both strictly positive, i.e., $\mu_{IC} > 0$ and $\mu_{IR} > 0$.*

PROOF:

Since (IR_G) will always be satisfied with equality due to an appropriate adjustment of the lowest intrinsic utility level offered, relaxing (IR_G) will always lead to strictly lower costs for the principal. Therefore, the shadow value of relaxing (IR_G) is strictly positive, so $\mu_{IR} > 0$.

Next, we show that relaxing (IC_G) has a positive shadow value, $\mu_{IC} > 0$. We do this by showing that a decrease in $c'(\hat{a})$ leads to a reduction in the principal's minimum cost of implementation. Let $(u_s^*)_{s\in\mathcal{S}}$ be the optimal contract under (the original) Program MG, and suppose that $c'(\hat{a})$ decreases. Now the principal can offer a new contract $(u_s^N)_{s\in\mathcal{S}}$ of the form

$$u_s^N = \alpha u_s^* + (1-\alpha)\sum_{t=1}^{S}\gamma_t(\hat{a})u_t^* \,, \tag{C.1}$$

where $\alpha \in (0,1)$, which also satisfies (IR_G), the relaxed (IC_G), and (OC_G), but yields strictly lower costs of implementation than the original contract $(u_s^*)_{s\in\mathcal{S}}$.

Clearly, for $\hat{\alpha} \in (0,1)$, $u_s^N < u_{s'}^N$ if and only if $u_s^* < u_{s'}^*$, so (OC_G) is also satisfied under contract $(u_s^N)_{s\in\mathcal{S}}$.

Next, we check that the relaxed (IC_G) holds under $(u_s^N)_{s\in\mathcal{S}}$. To see this, note that for $\alpha = 1$ we have $(u_s^N)_{s\in\mathcal{S}} \equiv (u_s^*)_{s\in\mathcal{S}}$. Thus, for $\alpha = 1$, the relaxed (IC_G) is oversatisfied under $(u_s^N)_{s\in\mathcal{S}}$. For $\alpha = 0$, on the other hand, the left-hand side of (IC_G) is equal to zero, and the relaxed (IC_G) in consequence is not satisfied. Since the left-hand side of (IC_G) is continuous in α under contract $(u_s^N)_{s\in\mathcal{S}}$, by the intermediate-value theorem there exists $\hat{\alpha} \in (0,1)$ such that the relaxed (IC_G) is satisfied with equality.

Last, consider (IR_G). The left-hand side of (IR_G) under contract $(u_s^N)_{s\in\mathcal{S}}$ with $\alpha = \hat{\alpha}$ amounts to

$$\begin{aligned}
&\sum_{s=1}^{S}\gamma_s(\hat{a})u_s^N - (\lambda-1)\sum_{s=1}^{S-1}\sum_{t=s+1}^{S}\gamma_s(\hat{a})\gamma_t(\hat{a})\left[u_t^N - u_s^N\right] \\
&= \sum_{s=1}^{S}\gamma_s(\hat{a})u_s^* - \hat{\alpha}(\lambda-1)\sum_{s=1}^{S-1}\sum_{t=s+1}^{S}\gamma_s(\hat{a})\gamma_t(\hat{a})\left[u_t^* - u_s^*\right] \\
&> \sum_{s=1}^{S}\gamma_s(\hat{a})u_s^* - (\lambda-1)\sum_{s=1}^{S-1}\sum_{t=s+1}^{S}\gamma_s(\hat{a})\gamma_t(\hat{a})\left[u_t^* - u_s^*\right] \\
&= \bar{u} + c(\hat{a}) \,,
\end{aligned} \tag{C.2}$$

where the last equality follows from the fact that $(u_s^*)_{s\in\mathcal{S}}$ fulfills the (IR_G) with equality. Thus, contract $(u_s^N)_{s\in\mathcal{S}}$ is feasible in the sense that all constraints of program (MG) are met. It remains to show that the principal's costs are reduced. Since $h(\cdot)$ is strictly convex, the principal's objective function is strictly convex in α, with a minimum at $\alpha = 0$. Hence, the

principal's objective function is strictly increasing in α for $\alpha \in (0,1]$. Since $(u_s^N)_{s\in\mathcal{S}} \equiv (u_s^*)_{s\in\mathcal{S}}$ for $\alpha = 1$, for $\alpha = \hat{\alpha}$ we have

$$\sum_{s=1}^{S} \gamma_s(\hat{a}) h(u_s^*) > \sum_{s=1}^{S} \gamma_s(\hat{a}) h(u_s^N),$$

which establishes the desired result.

We now give a heuristic reasoning why pooling of information may well be optimal in this more general case. For the sake of argument, suppose there is no pooling of information in the sense that it is optimal to set distinct wages for distinct signals. In this case all order constraints are slack; formally, if $u_s \neq u_{s'}$ for all $s, s' \in \mathcal{S}$ and $s \neq s'$, then $\mu_{OC,s} = 0$ for all $s \in \{2, \ldots, S\}$. In this case, the first-order condition of optimality with respect to u_s, $\partial \mathcal{L}(\boldsymbol{u})/\partial u_s = 0$, can be written as follows:

$$\text{(C.3)} \quad h'(u_s) = \underbrace{\left(\mu_{IR} + \mu_{IC}\frac{\gamma_s^H - \gamma_s^L}{\gamma_s(\hat{a})}\right)}_{=:H_s} \underbrace{\left[1 - (\lambda - 1)\left(2\sum_{t=1}^{s-1}\gamma_t(\hat{a}) + \gamma_s(\hat{a}) - 1\right)\right]}_{=:\Gamma_s} - \underbrace{\mu_{IC}(\lambda - 1)\left[2\sum_{t=1}^{s-1}(\gamma_t^H - \gamma_t^L) + (\gamma_s^H - \gamma_s^L)\right]}_{=:\Lambda_s}.$$

For $\lambda = 1$ we have $h'(u_s) = H_s$, the standard "Holmström-formula".[34] Note that $\Gamma_s > 0$ for $\lambda \leq 2$. More importantly, irrespective of the signal ordering, we have $\Gamma_s > \Gamma_{s+1}$. The third term, Λ_s, can be either positive or negative. If the compound signal of all signals below s and the signal s itself are bad signals, then $\Lambda_s < 0$.

Since the incentive scheme is nondecreasing, when the order constraints are not binding it has to hold that $h'(u_s) \geq h'(u_{s-1})$. Thus, if $\mu_{OC,s-1} = \mu_{OC,s} = \mu_{OC,s+1} = 0$ the following inequality is satisfied:

$$\text{(C.4)} \qquad H_s \times \Gamma_s - \Lambda_s \geq H_{s-1} \times \Gamma_{s-1} - \Lambda_{s-1}.$$

Note that for the given ordering of signals, if there exists any pair of signals $s, s-1$ such that (C.4) is violated, then the optimal contract for this ordering involves pooling of wages. Even when $H_s > H_{s-1}$, as it is the case when signals are ordered according to their likelihood ratio, it is not clear that inequality (C.4) is satisfied. In particular, when s and $s-1$ are similarly informative it seems to be optimal to pay the same wage for these two signals as can easily be illustrated for the case of two good signals: If s and $s-1$ are similarly informative good

[34] See Holmström (1979).

signals then $H_s \approx H_{s-1} > 0$ but $\Gamma_s < \Gamma_{s-1}$ and $\Lambda_s > \Lambda_{s-1}$, thus condition (C.4) is violated. In summary, it may well be that for a given incentive-feasible ordering of signals, and thus overall as well, the order constraints are binding, i.e., it may be optimal to offer a contract which is less complex than the signal space allows for.

Application with Constant Relative Risk Aversion.—Suppose $h(u) = u^r$, with $r \geq 0$ being a measure for the agent's risk aversion. More precisely, the Arrow-Pratt measure for relative risk aversion of the agent's intrinsic utility function is $R = 1 - \frac{1}{r}$ and therefore constant. The following result states that the optimal contract is still a bonus contract when the agent is not only loss averse, but also slightly risk averse.

PROPOSITION 8: *Suppose (A1)-(A3) hold,* $h(u) = u^r$ *with* $r > 1$*, and* $\lambda > 1$*. Generically, for* r *sufficiently small the optimal incentive scheme* $(u_s^*)_{s=1}^S$ *is a bonus scheme, i.e.,* $u_s^* = u_H^*$ *for* $s \in \mathcal{B}^* \subset \mathcal{S}$ *and* $u_s^* = u_L^*$ *for* $s \in \mathcal{S} \backslash \mathcal{B}^*$ *where* $u_L^* < u_H^*$*.*

PROOF:

For the agent's intrinsic utility function being sufficiently linear, the principal's costs are approximately given by a second-order Taylor polynomial about $r = 1$, thus

$$C(\boldsymbol{u}|r) \approx \sum_{s \in \mathcal{S}} \gamma_s(\hat{a}) u_s + \Omega(\boldsymbol{u}|r) \ , \tag{C.5}$$

where

$$\Omega(\boldsymbol{u}|r) \equiv \sum_{s \in \mathcal{S}} \gamma_s(\hat{a}) \Big[(u_s \ln u_s)(r-1) + (1/2) u_s (\ln u_s)^2 (r-1)^2 \Big] \ . \tag{C.6}$$

Relabeling signals such that the wage profile is increasing allows us to express the incentive scheme in terms of increases in intrinsic utility. The agent's binding participation constraint implies that

$$u_1 = \bar{u} + c(\hat{a}) - \sum_{s=2}^{S} b_s \left\{ \sum_{\tau=s}^{S} \gamma_\tau(\hat{a}) - (\lambda - 1) \Big[\sum_{\tau=s}^{S} \gamma_\tau(\hat{a}) \Big] \Big[\sum_{t=1}^{s-1} \gamma_t(\hat{a}) \Big] \right\} \equiv u_1(\boldsymbol{b}) \tag{C.7}$$

and $u_s = u_1(\boldsymbol{b}) + \sum_{t=2}^{s} b_t \equiv u_s(\boldsymbol{b})$ for all $s = 2, \ldots, S$. Inserting the binding participation constraint into the above cost function and replacing $\Omega(\boldsymbol{u}|r)$ equivalently by $\tilde{\Omega}(\boldsymbol{b}|r) \equiv \Omega(u_1(\boldsymbol{b}), \ldots, u_S(\boldsymbol{b})|r)$ yields

$$C(\boldsymbol{b}|r) \approx \bar{u} + c(\hat{a}) + (\lambda - 1) \sum_{s=2}^{S} b_s \left[\sum_{\tau=s}^{S} \gamma_\tau(\hat{a}) \right] \left[\sum_{t=1}^{s-1} \gamma_t(\hat{a}) \right] + \tilde{\Omega}(\boldsymbol{b}|r) \ . \tag{C.8}$$

Hence, for a given increasing wage profile the principal's cost minimization problem is:

PROGRAM ME:

$$\min_{\boldsymbol{b}\in\mathbb{R}_+^{S-1}} \ \boldsymbol{b}'\boldsymbol{\rho}(\hat{\boldsymbol{\gamma}},\lambda,\hat{a}) + \tilde{\Omega}(\boldsymbol{b}|r)$$

$$\text{(IC}') \qquad \text{subject to} \quad \boldsymbol{b}'\boldsymbol{\beta}(\hat{\boldsymbol{\gamma}},\lambda,\hat{a}) = c'(\hat{a})$$

If r is sufficiently close to 1, then the incentive scheme that solves Program ML also solves Program ME. Note that generically Program ME is solved only by bonus schemes. Put differently, even if there are multiple optimal contracts for Program ML, all these contracts are generically simple bonus contracts. Thus, from Proposition 2 it follows that generically for r close to 1 the optimal incentive scheme entails a minimum of wage differentiation. Note that for $\lambda = 1$ the principal's problem is to minimize $\tilde{\Omega}(\boldsymbol{b}|r)$ even for r sufficiently close to 1.

REFERENCES

Abeler, Johannes, Armin Falk, Lorenz Götte, and David Huffman. 2009. "Reference Points and Effort Provision." IZA Discussion Paper 3939.

Barberis, Nicholas, Ming Huang, and Tano Santos. 2001. "Prospect Theory and Asset Prices." *Quarterly Journal of Economics*, 116(1): 1-53.

Bell, David E. 1985. "Disappointment in Decision Making under Uncertainty." *Operations Research*, 33(1): 1-27.

Bertsekas, Dimitri. 1974. "Necessary and Sufficient Conditions for Existence of an Optimal Portfolio." *Journal of Economic Theory*, 8(2): 235-47.

Blackwell, David. 1951. "Comparison of Experiments." *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, ed. Jerzy Neyman, 93-102. Berkeley: University of California Press.

Blackwell, David. 1953. "Equivalent Comparison of Experiments." *Annals of Mathematics and Statistics*, 24(2): 265-72.

Breiter, Hans C., Itzhak Aharon, Daniel Kahneman, Anders Dale, and Peter Shizgal. 2001. "Functional Imaging of Neural Responses to Expectancy and Experience of Monetary Gains and Losses." *Neuron*, 30(2): 619-39.

Churchill, Gilbert A., Neil M. Ford, and Orville C. Walker. 1993. *Sales Force Management*. Homewood: Irwin.

Crawford, Vincent P., and Juanjuan Meng. 2009. "New York City Cabdrivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income." http://dss.ucsd.edu/~vcrawfor/cabdriver11.3.pdf.

Daido, Kohei, and Hideshi Itoh. 2007. "The Pygmalion and Galatea Effects: An Agency Model with Reference-Dependent Preferences and Applications to Self-Fulfilling Prophecy." School of Economics, Kwansei Gakuin University Discussion Paper 35.

de Meza, David, and David C. Webb. 2007. "Incentive Design Under Loss Aversion." *Journal of the European Economic Association*, 5(1): 66-92.

Demougin, Dominique, and Claude Fluet. 1998. "Mechanism Sufficient Statistic in the Risk-Neutral Agency Problem." *Journal of Institutional and Theoretical Economics*, 154(4): 622-639.

Dittmann, Ingolf, Ernst Maug, and Oliver G. Spalt. Forthcoming. "Sticks or Carrots? Optimal CEO Compensation when Managers are Loss-Averse." *Journal of Finance*.

Gjesdal, Frøystein. 1982. "Information, and Incentives: The Agency Information Problem." *Review of Economic Studies*, 49(3): 373-90.

Grossman, Sanford J., and Oliver D. Hart. 1983. "An Analysis of the Principal-Agent Problem." *Econometrica*, 51(1): 7-45.

Gul, Faruk. 1991. "A Theory of Disappointment Aversion." *Econometrica*, 59(3): 667-86.

Haller, Hans. 1985. "The Principal-Agent Problem with a Satisficing Agent." *Journal of Economic Behavior and Organization*, 6(4): 359-79.

Heidhues, Paul, and Botond Kőszegi. 2005. "The Impact of Consumer Loss Aversion on Pricing." CEPR Discussion Paper 4849.

Heidhues, Paul, and Botond Kőszegi. 2008. "Competition and Price Variation when Consumers are Loss Averse." *American Economic Review*, 98(4): 1245-68.

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

Iantchev, Emil P. 2009. "Risk or Loss Aversion? Evidence from Personnel Records." http://faculty.maxwell.syr.edu/iantchev/Site/Welcome_files/Risk_or_LossAversion4-1.pdf.

Jewitt, Ian, Ohad Kadan, and Jeroen M. Swinkels. 2008. "Moral Hazard with Bounded Payments." *Journal of Economic Theory*, 143(1): 59-82.

Joseph, Kissan, and Manohar.U. Kalwani. 1998. "The Role of Bonus Pay in Salesforce Compensation Plans." *Industrial Marketing Management*, 27(2): 147-59.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47(2): 263-91.

Kim, Son Ku. 1995. "Efficiency of an Information System in an Agency Model." *Econometrica*, 63(1): 89-102.

Kim, Son Ku. 1997. "Limited Liability and Bonus Contracts." *Journal of Economics & Management Strategy*, 6(4): 899-913.

Kőszegi, Botond, and Matthew Rabin. 2006. "A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics*, 121(4): 1133-65.

Kőszegi, Botond, and Matthew Rabin. 2007. "Reference-Dependent Risk Preferences." *American Economic Review*, 97(4): 1047-73.

Kőszegi, Botond, and Matthew Rabin. 2009. "Reference-Dependent Consumption Plans." *American Economic Review*, 99(3): 909-36.

Larsen, Jeff T., A. Peter Mc Graw, Barbara A. Mellers, and John T. Cacioppo. 2004. "The Agony of Victory and Thrill of Defeat: Mixed Emotional Reactions to Disappointing Wins and Relieving Losses." *Psychological Science*, 15(5): 325-30.

Lazear, Edward P., and Paul Oyer. 2007. "Personnel Economics." NBER Working Paper 13480.

Loomes, Graham, and Robert Sugden. 1986. "Disappointment and Dynamic Consistency in Choice under Uncertainty." *Review of Economic Studies*, 53(2): 271-82.

MacLeod, W. Bentley. 2003. "Optimal Contracting with Subjective Evaluation." *American Economic Review*, 93(1): 216-40.

Mellers, Barbara, Alan Schwartz, and Ilana Ritov. 1999. "Emotion-Based Choice." *Journal of Experimental Psychology, General*, 128(3): 332-45.

Moynahan, John K. 1980. *Designing an Effective Sales Compensation Program.* New York: AMACOM.

Oyer, Paul. 1998. "Fiscal Year Ends and Non-Linear Incentive Contracts: The Effect on Business Seasonality." *Quarterly Journal of Economics*, 113(1): 149-88.

Oyer, Paul. 2000. "A Theory of Sales Quotas with Limited Liability and Rent Sharing." *Journal of Labor Economics*, 18(3): 405-26.

Park, Eun-Soo. 1995. "Incentive Contracting under Limited Liability." *Journal of Economics & Management Strategy*, 4(3): 477-90.

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature*, 37(1): 7-63.

Post, Thierry, Martijn J. Van den Assem, Guido Baltussen, and Richard H. Thaler. 2008. "Deal Or No Deal? Decision Making Under Risk in a Large-Payoff Game Show." *American Economic Review*, 98(1): 38-71.

Rayo, Luis, and Gary S. Becker. 2007. "Evolutionary Efficiency and Happiness." *Journal of Political Economy*, 115(2): 302-37.

Salanié, Bernard. 2003. "Testing Contract Theory." *CESifo Economic Studies*, 49(3): 461-77.

Schmidt, Ulrich. 1999. "Moral Hazard and First-Order Risk Aversion." *Journal of Economics*, Supplement 8: 167-79.

Steenburgh, Thomas J. 2008. "Effort or Timing: The Effect of Lump-Sum Bonuses." *Quantitative Marketing and Economic*, 6(3): 235-56.

Strausz, Roland. 2006. "Deterministic Versus Stochastic Mechanisms in Principal-Agent Models." *Journal of Economic Theory*, 128(1): 306-14.

Tversky, Amos, and Daniel Kahneman. 1991. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *Quarterly Journal of Economics*, 106(3): 1039-61.

Yaari, Menahem E. 1987. "The Dual Theory of Choice under Risk." *Econometrica*, 55(6): 95-115.

Zábojník, Ján. 2002. "The Employment Contract as a Lottery." http://mylaw2.usc.edu/centers/cleo/workshops/02-03/zabojnik.pdf.