# Incentive design for reference-dependent preferences*

Víctor González-Jiménez[a]

[a]*Erasmus School of Economics, Erasmus University Rotterdam*

**Abstract**

I investigate the optimal design of incentives when the agent's risk preferences exhibit reference dependence. The theoretical framework of this paper incorporates the most prominent characterizations of reference-dependent preferences and integrates frequently used reference point rules. I find that, regardless of the chosen preference specification and reference point rule, the optimal contract must include a bonus. A contract feature that allows the principal to profitably exploit the agent's irrationalities of loss aversion and diminishing sensitivity. This paper provides a rationale for incentive schemes including bonuses grounded in preference.

JEL Codes: D82, D86, J41, M52, M12
Keywords: incentive theory, reference dependence, principal-agent models, loss aversion, bonuses.

* I thank Aurelien Baillon, Han Bleichrodt, and Clement Minaudier for helpful comments. This paper benefited from comments in seminars at the University of Vienna.

## 1. Introduction

Abundant empirical evidence shows that individuals evaluate outcomes of risky alternatives relative to a reference point (Tversky and Kahneman, 1992, Abdellaoui et al., 2007, Von Gaudecker et al. 2011, Vieider et al. 2015, Baillon et al. 2020). Decision-making based on these evaluations generates major irrationalities, i.e. deviations from expected utility, that explain relevant economic phenomena. For example, the regularity that individuals are risk averse for lotteries with small stakes while not being absurdly risk averse for lotteries featuring large stakes (Rabin, 2000). [1]

That risk preference is central to incentive design is manifested in the solution to the moral hazard problem. Accordingly, optimal contracts emerge from a tradeoff faced by the principal between providing full insurance to the agent and eliciting high effort levels (Mirrlees 1976, Holmström, 1979). This property of optimal contracts conveys a powerful, yet relatively unexplored, implication. Abandoning expected utility to adopt *descriptively valid* theories of risk generates contracts that can more effectively motivate individuals in practice and that might more closely resemble the compensation practices used by organizations.

This paper incorporates reference-dependent preferences in the optimal design of incentives. Its main result is that a contract with a bonus emerges as a solution. Throughout, I refer to a bonus as a discrete jump in the agent's compensation triggered when a production threshold is met (Park 1995, Kim 1997, Oyer, 2000). According to Worldatwork (2018), 90% of American enterprises use incentive schemes that include a bonus. I provide a foundation to this widespread practice grounded in preference. Importantly, the optimality of bonuses emerges regardless of the theory of risk used to model reference dependent preferences and irrespective of the rule used to define reference points.

A bonus contract is optimal because it allows the principal to profitably exploit the agent's irrationalities. If low levels of output realize, the contract pays transfers that locate the agent in the

[1] For an extensive list of papers using reference dependence to explain economic phenomena the reader is referred to footnote 1 in Baillon et al. (2020).

domain of losses. When facing this prospect of losses, the agent is motivated to exert high effort by virtue of *loss aversion*. That is, to avoid these losses high effort will be exerted. In addition, the risk seeking attitudes from *diminishing sensitivity* imply that the risk exposure inflicted by these low transfers is tolerated by the agent. The magnitude of the bonus ensures that the agent is eventually transitioned from the domain of losses to the domain of gains. Moreover, the output level after which the bonus is awarded is chosen so that, on expectation, the losses included in the contract are offset. In this way, a contract paying low amounts at low but potentially likely performance realizations and paying a sizable bonus at high but potentially unlikely performance realizations is not only accepted by the agent but also motivates him by taking advantage of his irrationalities.

While I am not the first to incorporate reference-dependence in a principal-agent framework, existing studies differ in the way risk attitude is characterized as well as in the assumed reference point (De Meza and Webb, 2007, Dittmann et al., 2010, Herweg et al., 2010, Corgnet et al., 2018). These disagreements have led to diverse solutions and interpretations. For example, two starkly different payment modalities such as stochastic contracts—in which the principal turns a blind-eye to the agent's performance signals—and option-like contracts—in which the contract is performance-insensitive at high performance levels—are currently acknowledged solutions to the principal problem when the agent is loss averse (De Meza and Webb, 2007, Herweg et al., 2010).

The theoretical framework in this paper unifies the most prominent approaches to modeling reference-dependent preferences. The focus is on models of reference dependence with axiomatic foundations, namely prospect theory (Tversky and Kahneman, 1992), disappointment models with prior (Bell, 1985, Loomes and Sugden, 1986, Gul, 1991), and disappointment models without prior (Delquié and Cillo, 2006, Kőszegi and Rabin, 2006, 2007). The model also adopts the most prominent reference point rules, such as the status quo, max-min, goals as reference points, expectations-based reference points, and the outcomes of the contract as the reference point. To the best of my knowledge, this is the first paper to consider such a general framework. Allowing me not only to reconcile previous findings, but also to generalize them. Moreover, the proposed framework also generates novel and relevant results. For example, that a bonus contract can be first-best optimal under reference dependence, that loss aversion and diminishing sensitivity are each necessary but not sufficient conditions for the emergence of a bonus, and that it can be optimal not only to offer a bonus but to award it when a production goal is just met.

This paper contributes to previous literature in several ways. First, it provides a justification for bonuses that contrasts the standard explanation of limited liability (Park 1995, Kim 1997, Oyer, 2000). A prominent disadvantage of attributing the optimality of bonus contracts to limited liability is that such result is irreconcilable with the regularity of individuals being predominantly risk averse. In fact, introducing risk aversion to the slightest degree in those frameworks leads to solutions without bonuses (Jewitt et al., 2008). This study pursues a completely different approach. It characterizes risk preference with descriptive theories of risk, so the agent suffers from the well-established irrationalities of loss aversion and diminishing sensitivity, to subsequently find that bonuses are second-best optimal and can be first-best optimal.

Second, this is the first theoretical study to employ a general specification of reference dependence. The specification is taken from Baillon et al. (2020) and is applied to the principal-agent framework. Adopting that specification leads to solutions with properties that were deemed as desirable by previous studies. For instance, discontinuities in the pay schedule (De Meza and Webb, 2007, Dittmann et al., 2010, Herweg et al., 2010) and performance-pay for high performance levels (De Meza and Webb, 2007, Dittmann et al., 2010). Moreover, the model is flexible as it includes other important results in the literature as special cases. For example, that stochastic contracts are optimal under severe loss aversion (Herweg et al., 2010, Corgnet et al., 2019).

The generality of the present framework, its capacity to incorporate properties that were deemed desirable by previous studies, and its flexibility imply that its results can be extrapolated to other settings in which incentive design is crucial. For instance, insurance markets, corporate finance, and agricultural economics. Besides, this paper shows that practitioners can apply a general framework to model of reference dependence without having to worry about the generality of their findings due to their chosen preference specification and/or reference point rule.

## 2. General Setup

Consider a principal (she) hiring an agent (he) to produce output on a task.[2] Production output $y$ is a random variable that may take any value in the interval $[\underline{y}, \overline{y}]$, where $\underline{y} \geq 0$. The agent's action consists of choosing an effort $e \in \{e_L, e_H\}$. For simplicity, it is assumed that only high effort is costly to the agent.

**Assumption 1 (A1).** $c(e) = \begin{cases} c \text{ } if \text{ } e_H, \\ 0 \text{ } if \text{ } e_L. \end{cases}$ Where $c > 0$.

Furthermore, it is assumed that both agent and principal know that output is distributed according to the cumulative density function $F(y|e)$, which admits a probability density function $f(y|e)$. Importantly, output and effort relate according to the monotone likelihood ratio property.

**Assumption 2 (A2).** $\frac{\partial}{\partial y}\left(\frac{f(y|e_L)}{f(y|e_H)}\right) < 0 \; \forall \; y \in [\underline{y}, \overline{y}]$.

The implications of Assumption 2 are well-known: higher performance levels are more likely to be drawn from a probability density function conditional on high effort rather than from a probability density function conditional on lower effort.

To incentivize high effort, the principal offers the agent a take-it-or-leave-it contract including a transfer $w(y) \geq 0$. The agent can accept the contract and subsequently choose a level of effort $e$, or, alternatively, reject the contract and obtain his reservation utility $\overline{U} \geq 0$.[3] When the contract is accepted, payments are made according to $w(y)$ after the realization $y$ is known.

I assume that the principal is risk neutral. Intuitively, she is assumed to be able to pool multiple risks. Specifically, her objective function is:

$$\int_{\underline{y}}^{\overline{y}} \big(S(y) - w(y)\big) f(y|e) dy, \qquad (1)$$

where $S'(y) > 0$ and $S''(y) < 0$.

[2] The usage of female pronouns to refer to the principal and male pronouns to refer to the agent is standard in the field of incentive theory. I follow that convention.

[3] Since $\overline{U} \geq 0$, that $w(y)$ is nonnegative does not necessarily imply an absence of punishments as, according to Assumption 1, $c > 0$. Therefore, by setting a small enough $w(y)$ the principal can punish the agent by making him worse off than his outside option utility level.

The agent's preferences and how they dictate the way in which $w(y)$ enters the agent's utility is the main investigation of this paper. There are two ways in which transfers enter the agent's utility. First, they are evaluated in an absolute way. The following assumption formalizes that proposal.

**Assumption 3 (A3).** *The agent's consumption utility is a $C^2$ function $u: \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ with $u(0) = 0$, $u' > 0$, $u'' < 0$, and differentiable inverse $h := u^{-1}$.*

Assumption 3 posits that the agent's consumption utility exhibits diminishing returns to transfers. Within the framework of expected utility theory this property implies risk aversion.

Second, it also assumed that the agent exhibits reference-dependent preferences. The way in which reference dependence is modeled is by assuming that the consumption utility level generated by a transfer being contrasted to that generated by a reference point, $r > 0$. Transfers generating consumption utility above that generated by the reference point are classified by the agent as *gains* while transfers generating consumption utility below that level are classified as *losses*. Where, as it will be formalized below, gains and losses generate different risk preferences. The following assumption provides a formal account of the agent's value function, the utility component introducing reference dependence.

**Assumption 4 (A4).** *The agent's value function is the piece-wise function:*

$$V(w,r) = \begin{cases} v(u(w(y)) - u(r)) & \text{if} \quad w(y) \geq r, \\ -\lambda v\left(u(r) - u\left(w(y)\right)\right) & \text{if} \quad r < w(y). \end{cases}$$

*Where $\lambda > 1$ and $v: \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ is $C^2$ and exhibits*

*i)* $v(0) = 0$;
*ii)* $v' > 0$;
*iii)* $v'' < 0$;
*iv) differentiable inverse $f := v^{-1}$.*

There are two sources of risk attitude emerging from the value function described by Assumption 4. First, the curvature of $v$ is concave if $w(y) \geq r$, i.e. the domain of gains, but convex if $w(y) < r$, the domain of losses. Shapes generating risk aversion and risk seeking in gains and losses, respectively. A property known as *diminishing sensitivity*. In the principal-agent framework this property implies that the agent is willing to accept contracts that expose him to large degrees of risk while located in the domain of losses but requires contracts that protect him from risk while located in the domain of gains.

Second, the constant restriction $\lambda > 1$ implies that the agent is loss averse. That is because losses loom larger than equally sized gains. Loss aversion implies that the agent will reject contracts that expose him to losses and that do not properly compensate him for such exposure.

The agent's reference point, $r$, is assumed to be exogenous.[4] This assumption is consistent with the status quo or current welfare position as reference point (Kahneman and Tversky, 1979). In subsequent sections, I relax that assumption to gain robustness and generalizability. An implication of $r$ being exogenous is that regardless of how high or low the incentives included in the contract are, they do not affect the agent's reference point and thus what constitutes a gain or a loss for him.

All in all, the utility of an agent with reference-dependent preferences is characterized by

$$\begin{aligned} U(e, w(y), r) = \phi \int_{\underline{y}}^{\overline{y}} u\big(w(y)\big) f(y|e) \mathrm{d}y \\ + \eta \int_{\underline{y}}^{\overline{y}} \Big(\theta_{\mathbb{I}} v\big(u\big(w(y)\big) - u(r)\big) - \lambda(1-\theta_{\mathbb{I}}) v\left(u(r) - u\big(w(y)\big)\right)\Big) f(y|e) \mathrm{d}y \\ - c(e). \end{aligned} \tag{2}$$

The utility function in Eq. (2) consists of three expressions. The first one represents the agent's absolute evaluation of transfers. I refer to this expression as *expected consumption utility*. The parameter accompanying this expression, $\phi \in \{0,1\}$, determines whether the agent's preferences exhibit expected consumption utility, $\phi = 1$, or not, $\phi = 0$. As it will be explained later, some theories of risk with reference dependence require expected consumption utility. The second expression in Eq. (2) contains the value function described by Assumption 4. The parameter $\eta \geq 0$ captures the weight given to the contrasting outcomes relative to the reference point in the agent's

[4] In initial formulations of prospect theory, the *status quo* or the current welfare position of the decision-maker when making a choice was assumed to be the reference point (Kahneman and Tversky, 1979).

utility. When $\eta = 0$ the agent does not evaluate outcomes relative to a reference point. Moreover, the indicator function $\theta_{\mathbb{I}}$ takes value $\theta_{\mathbb{I}} = 1$ if the agent is in the domain of gains, $w(y) \geq r$, and becomes $\theta_{\mathbb{I}} = 0$ otherwise. Finally, the last expression in that equation captures the cost of effort.

This general representation of reference-dependent preferences is based on that of Baillon et al. (2020). However, in my model the contrasting of outcomes relative to the reference point is not linear, $w(y) - r$, but instead depends on the comparative level of utility that these two wealth levels generate, $u(w(y)) - u(r)$. This way of modeling reference dependence gives me sufficient flexibility to derive previous results in the literature. In addition, this property also enables me to show that the main result of the paper, a contract with a bonus, does not necessitate diminishing sensitivity.[5]

The shape of $U(e, w(y), r)$ is crucial to our analysis of incentives. Due to the conjunction of Assumption 3 and Assumption 4, utility can take two shapes: concave and S-shaped. The following definition formalizes the latter shape. Moreover, Lemma 1 provides a sufficient condition, in terms of the curvatures of the value function, $v$, and consumption utility, $u$, for $U(e, w(y), r)$ to be S-shaped.

**Definition.** *$U(e, w(\tilde{y}), r)$ is S-shaped if is convex in $w(\tilde{y}) < r$ and concave in $w(\tilde{y}) \geq r$, where* $\tilde{y} \in \left[\underline{y}, \overline{y}\right]$ is an output realization.

**Lemma 1.** *$U(e, w(\tilde{y}), r)$ is S-shaped if and only if* $-\frac{u''(w(\tilde{y}))}{u'(w(\tilde{y}))} \leq -\frac{u'(w(\tilde{y}))\left(\lambda\eta v''\left(u(r)-u(w(\tilde{y}))\right)\right)}{\left(\phi+\lambda\eta v'\left(u(r)-u(w(\tilde{y}))\right)\right)}$ *holds for any $\tilde{y} \in \left[\underline{y}, \overline{y}\right]$ such that $w(\tilde{y}) < r$.*

When the risk seeking implied by diminishing sensitivity in losses is stronger than the risk aversion implied by $u$, the agent's utility is S-shaped. Such an agent is risk seeking in the domain of losses while being risk averse in the domain of gains. Lemma 1 will be useful to show how the solution to the principal's problem changes with the agent's utility shape.

[5] As pointed out by Köbberling and Wakker (2005) consumption utility from Assumption 3 reflects the normative component of utility. Hence, a purely descriptive model would therefore attribute all utility curvature to diminishing sensitivity by letting $u' = 1$, as done by Baillon et al. (2020).

## 3. General solution

The principal needs to implement a contract that is both accepted by the agent and that incentivizes him to exert high effort. Formally, her program is:

$$\max_{\{\mathrm{w(y)}\}} \int_{\underline{y}}^{\overline{y}} \big(S(y) - w(y)\big) f(y|e_H) dy$$

Subject to

$$\text{PC:} \qquad U(e_H, w(y), r) \geq \overline{U}, \tag{3}$$

$$\text{IC:} \qquad U(e_H, w(y), r) \geq U(e_L, w(y), r).$$

The solution to the maximization problem in (3) is presented in Theorem 1. The most relevant property of that solution is that it is second-best optimal and can be first-best optimal to offer a contract with a bonus. The proofs of the main results of the paper are relegated to Appendix A.

**Theorem 1.** *Under A1- A4, there exist unique output levels $\hat{y}_s$, $\hat{y}_f \in (\underline{y}, \overline{y})$ such that the second-best optimal contract, $w_s^*(y)$, awards a bonus at $y = \hat{y}_s$ and either*

i. *pays the lowest possible transfer in $y < \hat{y}_s$ and increases in $y$ in $y > \hat{y}_s$ if $U(e, w_s^*(\tilde{y}), r)$ is S-shaped, or*

ii. *increases in $y$ in both $y < \hat{y}_s$ and $y > \hat{y}_s$ if $U(e, w_s^*(\tilde{y}), r)$ is concave.*

*In turn, the first-best optimal contract, $w_f^*(y)$, either*

iii. *pays the lowest possible transfer in $y < \hat{y}_f$, awards a bonus at $y = \hat{y}_f$, and pays a higher constant transfer in $y > \hat{y}_f$ if $U\big(e, w_f^*(\tilde{y}), r\big)$ is S-shaped, or*

iv. *pays the higher-than-the-lowest constant transfer for all $y$ if $U\big(e, w_f^*(\tilde{y}), r\big)$ is concave.*

The bonus included in the second-best contract allows the principal to exploit the agent's irrationalities to extract output. That is because in the segment in which the bonus is not awarded, $y < \hat{y}_s$, the contract pays transfers that locate the agent in the domain of losses. When facing this prospect, the loss averse agent is motivated to exert high effort to avoid incurring in these losses.

This motivating effect takes place even though exceptionally low financial incentives are given. Furthermore, the risk seeking from diminishing sensitivity ensures that the risk exposure inflicted by low payments is tolerated. However, the principal is aware that a contract consisting exclusively of losses would be rejected. That is why she offers a bonus designed to locate the agent in the domain of gains for all output levels $y \geq \hat{y}_s$. To make the agent indifferent between accepting or not the contract, the output level after which the bonus is awarded, $\hat{y}_s$, is such that, on expectation, the exposure to losses is offset.

That a bonus contract is optimal regardless of whether $U$ is concave or S-shaped, implies that a strong degree of diminishing sensitivity is not necessary to obtain this contract shape. In fact, the intuition given in the previous paragraph elucidates that a bonus is optimal when the agent suffers either from loss aversion or diminishing sensitivity, regardless of their degree.

Theorem 1 also presents the first-best optimal contract. A lump-sum bonus provides full insurance when $U$ is S-shaped. That is because the agent is risk seeking in losses and thus willing to be fully exposed to risk at the low end-of the output space, $y < \hat{y}_f$. These risk seeking attitudes make a strategy of offering low payments affordable for the principal. Loss aversion in turn implies that to ensure that the contract is accepted, it must also offer transfers that locate the agent in gains and that offset the losses of the contracts. This is achieved with a bonus. Instead, when $U(e, w, r)$ is concave the excessive risk exposure from low payments is not tolerated, due to risk aversion, and full insurance is attained with a constant transfer that locates the agent in the domain of gains.

To better understand the incentives imparted by the second-best contract from Theorem 1, its shape is compared to that of the first-best contract. The following corollary describes that comparison.

**Corollary 1** *(Rewards and Punishments). Punishments are imparted in both domains because the second-best optimal contract, $w_s^*(y)$, from Theorem 1,*

*i) exhibits $\hat{y}_e > \hat{y}_s \geq \hat{y}_f$ if $U(e, w_s^*(\tilde{y}), r)$ is S-shaped;*

*ii) exhibits $\hat{y}_e > \hat{y}_s$ if $U(e, w_s^*(\tilde{y}), r)$ is concave.*

*Where $\hat{y}_e$ is a unique output level satisfying $w_s^*(\hat{y}_e) = w_f^*(\hat{y}_e) > 0$.*

The second-best contract from Theorem 1 elicits high effort by imparting punishments in losses and gains. The following explanation focuses on the more interesting case in which $U(e, w_s^*(\tilde{y}), r)$ is S-shaped for any $\tilde{y} \in \left[\underline{y}, \overline{y}\right]$. In the segment $y < \hat{y}_f$, both first- and second-best contracts pay the lowest possible transfer, locating the agent in the domain of losses. Thus, $w_s^*(y)$ does not impart punishments for low performance but ensures that the agent is motivated by virtue of loss aversion. Intuitively, the agent exerts high effort to avoid incurring in losses. Moreover, the bonus feature of both contracts and the fact that $\hat{y}_s \geq \hat{y}_f$ generate sizeable punishments in $y \in (\hat{y}_f, \hat{y}_s)$. Therefore, the second-best contract includes when the agent's irrationalities become, on their own, insufficient for eliciting high effort. Finally, that the transfers of the second-best contract increase in performance in $y > \hat{y}_s$, generates incentives in the domain of gains in a standard way: punishments are given for low performance and rewards are given for high performance.[6]

The following corollary presents comparative statics that further the understanding of the influence of reference dependence on the optimal implementation of incentives. These comparative statics are invoked in the next subsections to provide intuition. Their focus is on the more complete case of moral hazard.

**Corollary 2** (Comparative statics).

*i) Higher $r$ yields a $w_s^*(y)$ with a larger bonus. The bonus is awarded at a higher $\hat{y}_s$ if $U(e, w_f^*(\tilde{y}), r)$ is S-shaped or at a lower $\hat{y}_s$ if $U(e, w_f^*(\tilde{y}), r)$ is concave.*

*ii) Higher $\lambda$ yields a $w_s^*(y)$ with a bonus awarded at a lower $\hat{y}_s$.*

A higher reference point implies that bonus needs to become larger to ensure that the agent is transitioned from the domain of losses to the domain of gains. This prospect of higher bonuses enables the principal to expose the agent to losses for a larger segment of output if he tolerates such risk exposure. Therefore, for sufficiently strong diminishing sensitivity, the threshold at which the bonus is awarded, $\hat{y}_s$, becomes higher. If the agent cannot tolerate that risk exposure,

---

[6] Corollary 1(ii) follows a similar yet simpler explanation. In that case, $w_f^*$ locates the agent in gains everywhere. So, for any $y < y_s$, contract $w_s^*$ imparts punishments in the domain of losses. Moreover, that $\tilde{y} > \hat{y}_s$ implies that $w_s^*$ punishments are also imparted in the domain of gains.

the bonus is instead given at a lower $\hat{y}_s$.[7] Moreover, a higher level of loss aversion implies that the agent experiences more disutility when exposed to losses. To keep the contract attractive to the agent, the bonus will be awarded at lower output levels.

## 4. Old and New Results

In this section, I show that Theorem 1 can be easily adapted to explain previous findings. Moreover, I demonstrate that the importance of Theorem 1 is not only restricted to explain previous findings but also to provide new insights that fill existing gaps in the literature.

### 4.1 Standard Preferences

Consider first the more traditional framework in which the evaluation of potential transfers is not performed relative to a reference point. That is a setting in which $\eta = 0$ is applied to Eq. (2). In that case, Theorem 1 yields the standard results from incentive theory.

**Corollary 3.** *Under A1-A4,* $\phi = 1$, and $\eta = 0$.

*i.* **(Borch, 1962).** *The first-best contract,* $w_f^*$*, pays a constant transfer* $w_f^* = h'\left(\frac{1}{\mu}\right)$ *where* $\mu > 0$ *is a constant.*

*ii.* **(Holmström, 1979).** *The second-best contract* $w_s^*(y)$ *is everywhere increasing in* $y$ *and satisfies* $w_s^*(y) = h'\left(\frac{1}{\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right)}\right)$ *where* $\gamma > 0$ , $\mu > 0$ *are constants.*

Without reference-dependent preferences, it is second-best optimal to offer a contract that is everywhere increasing in performance. Moreover, it is first-best optimal to pay fixed that fully insures the risk averse agent from risk.

### 4.2 Prospect Theory Preferences

---

[7] In this case a higher reference point leads to a larger segment of losses, due to the agent's low tolerance to risk it is more cost-effective for the principal to offer the bonus at a lower output levels rather than further raising the size of the bonus.

Next, I consider reference-dependent preferences within the framework of prospect theory (Tversky and Kahneman, 1992).[8] According to that theory, the agent evaluates outcomes exclusively using the value function from Assumption 4. Therefore, the preferences in Eq. (2) are modified by means of the restrictions $\phi = 0$, $\eta = 1$, and $u' = 1$.[9]

The following corollary presents the results from Theorem 1 when the agent exhibits prospect theory preferences. Throughout, $\gamma > 0$ , $\mu > 0$ are constants corresponding to the Lagrangian multipliers of the incentive and participation constraints, respectively.

**Corollary 4.** *Under A1-A4,* $\phi = 0, \eta = 1$ and $u' = 1$, *there exist unique output levels* $\hat{y}_{pf}, \hat{y}_{ps} \in (\underline{y}, \overline{y})$ *such that:*

*i)* *The first-best contract,* $w_f^*(y)$, *pays the lowest possible transfer if* $y < \hat{y}_{pf}$, *awards a bonus at* $y = \hat{y}_{pf}$, *and pays the constant transfer* $w_f^*(y) = r + f'\left(\frac{1}{\mu}\right)$ *if* $y > \hat{y}_{pf}$.

*ii)* **(Dittmann et al., 2010)** *The second-best contract,* $w_s^*(y)$, *pays the lowest possible transfer if* $y < \hat{y}_{ps}$, *exhibits a bonus at* $y = \hat{y}_{ps}$, *and increases in* $y$ *according to the schedule* $w_s^*(y) = r + f'\left(\frac{1}{\mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right)}\right)$ *if* $y > \hat{y}_{ps}$.

Under prospect theory, the second-best contract in Theorem 1 collapses to the solution found by (Dittmann et al., 2010). That is, the contract offers the lowest possible pay when low output levels realize, i.e. lower than $\hat{y}_{ps}$, followed by a bonus and a segment in which pay increases with performance.[10]

---

[8] An axiomatization of prospect theory for risk is provided by Chateauneuf and Wakker (1999).

[9] Such characterization of prospect theory abstracts from probability weighting. The main goal of this paper is to study how reference dependence, on its own, affects the optimal design of incentives. The reader interested in the contract design when the agents exhibit probability weighting is referred to González-Jiménez and Castillo (2021).

[10] In fact, the result of Corollary 4 generalizes that of Dittmann et al. (2010) in that it does not assume a functional form for $v$ and does not assume a specific reference point $r$.

The results go beyond this finding. They also show that the first-best contract is a lump-sum bonus. Hence, the optimality of bonuses does not require a setting of moral hazard and emerge in a situation in which full insurance must be provided to the agent. Also, that the first-best contract is available provides a full understanding of the incentives imparted by the second-best contract to elicit high effort when the agent has prospect theory preferences. Corollary 1 (i) and the explanation laid out thereafter provide that description.

De Meza and Webb (2007) study an agent with loss aversion and "downward risk loving". However, their representation of preference starkly differs from that considered in Eq. (2) and prospect theory. In their model, loss aversion can be non-linear and the evaluation of outcomes relative to a reference point only apply to losses. To the best of my knowledge no axiomatized theory of risk provides that preference representation. Empirical evidence also suggests loss aversion to be constant (Abdellaoui et al., 2008, Abdellaoui et al., 2016). In Appendix B, I show how the model needs to be adapted to obtain the results in De Meza and Webb (2007).

### 4.3 Prospect Theory Preferences with Endogenous Reference Points

So far, $r$ has been assumed to be exogenous. In this section, I show how the results from Theorem 1 can be easily adapted to account for endogenous reference points. Throughout, the focus is on reference point rules consistent with prospect theory and, for brevity of exposition, on the moral hazard case.

#### Salience Rules

I first consider salience-based reference points. That is rules resulting from unconscious comparisons made by the agent between outcomes or probabilities in the contract. The first rule is the *max-min* (Hershey and Schoemaker, 1985). Accordingly, the agent is cautious; he takes as reference point the maximum value from a set consisting of the minimum possible outcome of each alternative. As an example, suppose that an agent is given two contracts $w_1 := (0.5{:}\,200, 0.5{:}\,0)$ and $w_2 := 100$. The set consisting of the minimum possible outcomes of each alternative is $\{0{,}100\}$, so the agent's reference point is $100$. According to Baillon et al. (2020) the max-min rule is broadly used by individuals when making risky decisions.

The following corollary presents the solution to the moral hazard problem under the max-min rule. The resulting contract resembles that of Corollary 4 (ii). Moreover, it turns out that the agent's reference point is the first-best contract presented in Corollary 4 (ii).

**Corollary 5.** *Under A1-A4,* $\phi = 0$, $\eta = 1$, $u' = 1$ and *the max-min rule, the agent's reference point is* $r = w_f^*(y)$, *the first-best optimal contract from Corollary 4 (i). Moreover, there exists a unique output level* $\hat{y}_u \in (\underline{y}, \overline{y})$ *such that the second-best optimal contract,* $w_s^*(y)$, *pays the lowest possible transfer if* $y < \hat{y}_u$, *awards a bonus at* $y = \hat{y}_u$, *and increases in performance if* $y > \hat{y}_u$.

The agent makes two choices: accepting or not the contract and exerting high or low effort. There exist thus four candidates for reference point, corresponding to the consequences implied by each possible choice. The utility level implied by the first-best contract from Corollary 4 (i), $w_f^*(y)$, is $\overline{U}$, the welfare level obtained when the contract is rejected. Notice that that utility level is higher than that implied by the minimum amount given to the agent by the second-best contract. Otherwise, that contract would not motivate the agent to exert high effort by means punishments. Therefore, under the max-min rule, the reference point is $r = w_f^*(y)$.

Anticipating that reference point, the principal offers a contract to the agent with a shape comparable to the contracts in Theorem 1 (ii) and Corollary 4 (ii). As a contract with those incentives profitably exploit the irrationalities of the prospect theory agent to ensure motivation and participation. The differences between the contract in Corollary 5 and that in Corollary 4 (ii) have to do with the fact that the agent's reference point $w_f^*(y)$ may not need to coincide with his status quo. Consequently, the size of the bonus and the output level after which the bonus is awarded can differ across those contracts.

Two additional salience-based reference points are considered: the *min-max* rule and the $w(y)$ *at max P* (Baillon et al., 2020). The first rule implies that individuals are bold; they take as the reference point the minimum value of a set consisting of the maximum outcome of each alternative. Using the previous example, in which $w_1$ and $w_2$ are given to the agent, this rule implies that 100 is the reference point. The second rule states that the output level realizing with the highest probability is becomes the agent's reference point. Using again the previous example, the agent's reference point is $100$.

The solutions under these reference points are described by Corollary B.1. and Corollary B.2. in Appendix B. The results therein confirm the previous intuition, these rules do not change the shape of the optimal contract, which happens to be similar to that of Corollary 4, but might change the size of the bonus and the output level after which it is awarded.

**Goals as Reference Points**

There is abundant evidence showing that individuals incorporate goals as reference points (Heath et al., 1999, Larrick et al., 2009, Allen et al., 2017). For example, higher performance in cognitive and/or physical tasks is achieved when a high rather than a low goal is set (Heath et al., 1999). A result explained by a loss aversion from missing the goal—an aversion to experience psychological losses from missing the goal which motivates higher effort exertion—and a diminishing sensitivity around the goal—the willingness to exert more effort as the goal is approached as well as the willingness to exert less effort as they move away from the goal.[11]

In this section, I assume that the agent incorporates a production goal chosen by the principal as reference point. Let that goal be $g \in [\underline{y}, \overline{y}]$. Importantly, a production goal may not need to coincide with the principal's nor the agent's expectations.[12] This difference makes goals consistent with prospect theory but not necessarily with other theories of risk with reference-dependence. This claim will become evident in the next section.

The agent's preferences when the goal is taken as the reference point are given by

[11] This rationale also explains why consumers save more energy and water when a savings goal is set (Harding and Hsiaw, 2014, Tiefenbeck et al., 2018) and why college students exhibit better performance when setting a task-based goal (Clark et al., 2020).[11]

[12] Intuitively, an expectation is fully governed by probabilities associated to possible outcomes, while a goal can encompass an ambition and hope component.

$$U(e, w(y), g) = \int_{g}^{\overline{y}} v(w(y) - w(g)) f(y|e) \mathrm{d}y - \lambda \int_{\underline{y}}^{g} v\big(w(g) - w(y)\big) f(y|e) \mathrm{d}y - c(e). \quad (4)$$

Equation (4) shows that the goal divides the agent's *output space* into gains and losses, and that such division affects how the transfers in the contract are evaluated. Specifically, any transfer of the contract is contrasted to the payment given when the goal is just met, $w(g)$. Thus, while the agent's reference point is $g$, that output level enters his objective function in the form of the monetary payment. This way of modeling goals as reference points maintains the unit of the utility domain consistent.

In this setting, the principal's problem is dual. She needs to choose the production goal, $g$, and determine how performance around that goal must be incentivized, $w(g)$. Hence, her proposal consists of a tuple $(w(y), g)$ offered before the agent exerts effort on the delegated task. The steps followed in Theorem 1 along with additional assumptions and elaborations provide a solution to this problem.

**Proposition 1.** *Under A1-A4,* $\phi = 0, \eta = 1, u' = 1, \lim_{x \to 0} v'(x) = +\infty$*, and* $r = w(g)$*, the principal offers the tuple* $(w_s^*(y), g^*)$ *where:*

*i)* $w_s^*(y)$ *pays the lowest possible transfer in* $y < g^*$*, awards a bonus that increases in* $g^*$ *at* $y = g^*$*, and increases in performance in* $y > g^*$*;*

*ii)* $g^*$ *satisfies* $\mathbb{E}(w_s^*(y)) - w_s^*(g^*) = \epsilon$*, for arbitrarily small* $\epsilon > 0$.

Under an additional assumption on the agent's value function, namely the Inada condition, the principal sets a challenging but, on expectation, attainable production goal and accompanies it with contract that awards a bonus when the production goal is met. This result captures the widespread practice in organizations of paying bonuses when an individual goal is achieved. According to Worldatwork (2018) more than 75% of American firms award bonuses in such a way when compensating employees.

Proposition 1 is the first to characterize optimal contracts in a setting of moral hazard when production goals are taken by the agent as reference points. Corgnet et al. (2018) also characterize a second-best optimal contract when the agent adopts the principal's goal as reference point but restrict their investigation to linear contracts. Proposition 1 contradicts this assumption.

In addition, the present framework differs from existing models of goal setting in Economics in several ways. First, I consider a setting in which achieving the goal is uncertain for the decision maker (Wu et al., 2008, Gomez-Minambres, 2012, Corgnet et al., 2015, Dalton et al., 2016, Dalton et al., 2016b, Brookins et al., 2017). Second, I model the agent's preferences using cumulative prospect theory. In line with early representations of goals as reference points (Heath et al., 1999, Wu et al., 2008), but contrasting most approaches in the literature in which the agent's preference is modeled using disappointment models (Koch and Nafziger, 2011, 2016, Gomez-Minambres, 2012, Corgnet et al., 2015, Corgnet et al., 2018). Third, the scale of the utility domain is kept constant.

### 4.4 Disappointment Aversion

A considerable bulk of the literature is sympathetic with the intuitive idea that the expected value of an alternative should gain the status of reference point (Abeler et al., 2011, Terzi et al. 2015, Sprenger, 2015, Gneezy et al. 2017). This reference point rule is however not compatible with prospect theory. To see how, consider an agent with preferences described by Eq. $(2)$ with the restrictions $\phi = 0, \eta = 1$ and $u' = 1$, who is offered a fixed wage contract $w_k := k$, and who adopts the expectation of a contract as reference point rule. The utility of that agent is $U(e, w_k, \mathbb{E}(w_k)) = 0$ for any $k > 0$. An absurd implication.

The merit of disappointment models is to include expectations, as well as other priors that the agent might form, as reference points without incurring this problem. They do so while maintaining the descriptively relevant phenomena of loss aversion and diminishing sensitivity.[13] This is mainly achieved by requiring expected consumption utility, $\phi = 1$ is applied.

---

[13] In the jargon of disappointment models, the agent experiences *disappointment* when the contract specifies a payment that is worse than his prior. While *elation* is experienced when the payment specified by the contract is better than his prior. The agent is thus disappointment averse. To keep a consistent terminology throughout the paper, I refer to the elation and disappointment outcomes as gains and losses, respectively.

Existing disappointment models differ in the prior adopted as reference point. In Bell (1985), the expected value of a risky alternative is the reference point while in Loomes and Sugden ( 1986) is the expected consumption utility of a risky alternative. The following Corollary shows how Theorem 1 can be adapted to provide the solution to the principal's problem when the agent exhibits preferences as in Bell (1985) or Loomes and Sugden (1986).

**Corollary 6.** Let $\bar{\bar{w}} := \int_{\underline{y}}^{\overline{y}} w(y)f(y|e)dy$. *Under A1-A4,* $\phi = 1$, *and* $r = \bar{\bar{w}}$ *there exist unique output levels* $\hat{y}_{mf}, \hat{y}_{ms} \in (\underline{y}, \overline{y})$ *such that the second-best optimal contract,* $w_s^*(y)$, *awards a bonus at* $y = \hat{y}_{ms}$, *and either*

- *i)* *pays the lowest possible transfer in* $y < \hat{y}_{ms}$ *and increases in* $y$ *according to* $w_s^*(y) = \bar{\bar{w}} + f'\left(\frac{1}{\eta}\left(\frac{1}{\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right)} - 1\right)\right)$ *in* $y > \hat{y}_{ms}$ *if* $u' = 1$, *or*
- *ii)* *increases in* $y$ *according to* $w_s^*(y) = h'\left(\frac{1}{(1+\eta\lambda)\left(\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right)\right)}\right)$ *in* $y < \hat{y}_{ms}$ *and according to* $w_s^*(y) = h'\left(\frac{1}{(1+\eta)\left(\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right)\right)}\right)$ *in* $y > \hat{y}_{ms}$ *if* $v' = 1$.

*In turn, the first-best optimal contract,* $w_f^*(y)$, *either*

- *iii)* *pays the lowest possible transfer in* $y < \hat{y}_{mf}$, *exhibits a bonus at* $y = \hat{y}_{mf}$, *and pays the constant transfer* $w_f^*(y) = \bar{\bar{w}} + f'\left(\frac{1}{\eta}\left(\frac{1}{\mu} - 1\right)\right)$ *in* $y > \hat{y}_{mf}$ *if* $u' = 1$, *or*
- *iv)* *pays the constant transfer* $w_f^* = h'\left(\frac{1}{(1+\eta)\mu}\right)$ *for all* $y$ *if* $v' = 1$.

When $u' = 1$, the model of Bell (1985) is relevant: the evaluation of outcomes is performed relative to the expected vale of the contract. Instead, when $v' = 1$, the model of Loomes and Sugden (1986) becomes relevant: outcomes of the contract are evaluated relative to the expected consumption utility. The corollary shows that in both cases a contract with a bonus incentivizes the agent to exert high effort in a moral hazard setting. This result is not trivial, it shows that the optimality of a bonus contract is guaranteed with diminishing sensitivity (Corollary 6 (i)) or without it (Corollary 6 (ii)). It thus highlights the crucial role of loss aversion in the emergence of bonuses as optimal solutions.

There are two significant differences between the contracts presented in Corollary 6 and Corollary 4. First, the second-best contract in Corollary 6 (ii) is everywhere increasing, except at the bonus. This is a direct consequence of $U$ being concave, an implausible assumption under prospect theory. Second, even if $U$ is S-shaped, as in Corollary 6 (i), the fact that the reference point is endogenous entails that slight contract modifications generate changes in the size of the bonus and the point after which it is awarded. One such modification of the contract with respect to the case in which the agent has prospect theory preferences is the inclusion of expected consumption utility, $\phi = 1$. Therefore, while the shapes of the contracts in Corollary 6 (ii) and Corollary 4 are similar, the bonus size and its location in the output are distinct.

Corollary 6 is, to the best of my knowledge, the first in the literature to provide a solution to the principal's problem when the agent exhibits expectations-based reference points. A rule that has received considerable attention in behavioral economics.

Appendix B presents the optimal solution to the contracting problem when the agent exhibits Gul (1991)'s disappointment model. According to those preferences, the reference point is the agent's certainty equivalent of a risky alternative. The solution therein shows that contracts similar to those presented Corollary 6 emerge as solutions. While Meza and Webb (2007) were the first to provide a solution to the principal problem under Gul (1991)'s preferences, their preference representation suffers from the problems discussed in Section 4.2.

### 4.5 Disappointment Aversion with the Contract as Reference Point

To conclude this section, I assume that the agent exhibits preferences as in the disappointment model of Delquié and Cillo (2006) and Kőszegi and Rabin (2006, 2007)'s choice acclimating equilibria. This is arguably the most used model in economics to capture reference dependence. It starkly differs from previous models because the reference point is assumed to be stochastic.

To exemplify this rule, suppose that the worker gets $w_3 := (0.5: 200, 0.5: 100)$. First, let $r = 200$. Obtaining 100 feels like a loss that realizes with 25% probability, obtained from $0.5 * 0.5$, while obtaining 200 is a neutral outcome that also realizes with 25% probability. Let now $r = 100$. Obtaining 200 feels like a gain realizing with 25% chance and obtaining 100 is a neutral outcome. Therefore, under the considered rule $w_3$ is reframed as a 25% chance to win 100, a 25% chance to loss 100, and a 50% chance to obtain a neutral outcome.

The agent's risk preference when each possible outcome of the contract is taken as the reference point is given by

$$U\left(e, w(y), w(\tilde{\tilde{y}})\right) = \int_{\underline{y}}^{\overline{y}} u(w(y)) f(y|e) dy + \eta \int_{\tilde{\tilde{y}}}^{\overline{y}} \int_{\underline{y}}^{\overline{y}} v\left(u(w(y)) - u(w(\tilde{\tilde{y}}))\right) f(\tilde{\tilde{y}}|e) f(y|e) d\tilde{\tilde{y}} dy - \eta\lambda \int_{\underline{y}}^{\tilde{\tilde{y}}} \int_{\underline{y}}^{\overline{y}} v\left(u\left(w(\tilde{\tilde{y}})\right) - u(w(y))\right) f(y|e)\, f(\tilde{\tilde{y}}|e)\, d\tilde{\tilde{y}} dy - c(e), \quad (5)$$

for every $\tilde{\tilde{y}} \in [\underline{y}, \overline{y}]$.

The following proposition presents the solution to the principal's problem when the agent's preferences are characterized by Eq. (5). The procedure to solve the problem implements steps from Theorem 1 plus additional elaborations. The resulting optimal contract can acutely differ from those presented in Theorem 1. Primarily because it can be first-best and second-best optimal to offer a stochastic contract.

**Proposition 2.** *Under A1-A4,* $\phi = 1$*, and each outcome of the contract as the reference point, there exists a unique level* $\hat{y}_{ds} \in (\underline{y}, \overline{y})$ *such that the second-best optimal contract either*

*i)* *pays the lottery* $L_s^* = (p_s^*(b_s): b_s, 1 - p_s^*(b_s): 0)$ *with* $b_s > 0$ *and* $p_s^*(b_s) := \frac{1}{2} + \frac{u(b_s)}{2\eta\left(v(u(b_s)) - \lambda v(-u(b_s))\right)}$ *in* $y < \overline{y}$ *and* $b_s$ *in* $y = \overline{y}$ *if U is S-shaped, or*

*ii)* *awards a bonus in* $y = \hat{y}_{ds}$ *and increases in performance elsewhere if U is concave and* $-1 \leq v'(0)(1 - \lambda)$*, or* *iii)* *pays the lottery* $L_s^* = (p\colon w_s, 1-p\colon 0)$ *where* $w_s^*$ *is the contract described in part ii) and* $p \in (0,1)$ *if U is concave and* $-1 > v'(0)(1-\lambda)$.

*In turn, the first-best optimal contract either*

*iv)* *pays the lottery* $L_f^* = \left(p_f^*(\overline{y})\colon b_f, 1-p_f^*(\overline{y})\colon 0\right)$ *with* $b_f > 0$ *and* $p_s^*(\colon b_f) := \frac{1}{2} + \frac{u(b_f)}{2\eta\left(v\left(u(b_f)\right)-\lambda v\left(-u(b_f)\right)\right)}$ *in* $y < \overline{y}$ *and* $b_f$ *in* $y = \overline{y}$ *if U is S-shaped, or*

*v)* *pays a positive constant transfer if U is concave and* $-1 \leq v'(0)(1-\lambda)$*, or*

*vi)* *pays the lottery* $L_f^* = \left(p\colon w_f, 1-p\colon 0\right)$ *where* $w_f^*$ *is the contract described in part v) and* $p \in (0,1)$ *if U is concave and* $-1 > v'(0)(1-\lambda)$.

A stochastic contract can be optimal because it guarantees participation and motivation. Suppose that $w = 0$ is paid for output levels $y < \overline{y}$. Under the considered reference point rule, this strategy, followed by the principal in the solution in Theorem 1 (i), now leads to larger disutility. That is because $w = 0$ is eventually taken as the reference point, so for any $r > w > 0$, then $-\lambda\, v\big(u(-w)\big) > -\lambda\, v(u(r) - u(w))$. To avoid this pronounced exposure to losses, that could lead the agent to reject the contract, the principal offers instead a stochastic payment $L_s^*$ that includes a non-zero probability to locate the agent in the domain of gains.

Furthermore, the stochastic contract motivates the agent in two distinct ways. First, at the expense of his irrationalities in the domain of losses. The intuition of this result is akin to that given in Theorem 1. Avoiding the loss implied by the lowest outcome included in $L_s^*$ incentivizes high effort, while diminishing sensitivity implies that the agent is willing to accept the stochastic payment. Second, the large monetary incentives in $y = \overline{y}$ motivate the agent in a more traditional way.

Another reason stochastic contracts are optimal is because under the considered reference point rule, stochastic dominance can be violated (Masatlioglu and Raymond, 2016). If that is the case, if $-1 > v'(0)(1-\lambda)$, then the principal can exploit this irrationality by offering a lottery. If this is not the case, the agent would be better off with a deterministic contract. This is the result portrayed in Proposition 2 (ii) and (iii).

In a well-known paper, Herweg et al. (2010) also find that, under the considered preference representation, the second-best contract is stochastic for high levels of loss aversion, $\lambda > 2$. More emphasis is given to their result stating that for low loss aversion levels, $\lambda < 2$, the contract becomes a lump-sum bonus. This result is not only clearly captured by Proposition 2 (ii)-(iii) but also by Proposition 2 (i) under a slightly different value function specification.[14] The following Corollary presents that result.

**Corollary 7 (Herweg et al., 2010).** *Under A1-A4,* $\phi = 1$, $\lambda \leq 2$, $\eta = 1$, $u' = 1$, $v' = 1$, *and the contract as the reference point, the second-best optimal contract is binary; consists of a base wage* $w = 0$ *and a lump-sum bonus* $w = b_s$.

## 5. Extensions

This section extends the results presented in Section 3 to further gain generalizability and highlight the significance of the results in Theorem 1. I demonstrate that those results are valid beyond a set of assumptions made on the principal's preferences and knowledge. Throughout the section, I assume that the agent's reference point is exogenous, as in Theorem 1, and focus on the moral hazard case.

### 5.1 Principal with Reference Dependent Preferences

A natural extension is to consider a setting in which the principal also exhibits reference-dependent preferences. Formally, let the principal's preferences be characterized by

$$\Pi\big(S(y), r_p, w\big) = \begin{cases} S(y) - r_p - w(y) & if\ \ S(y) \geq r_p + w(y), \\ -\lambda_p \left(r_p + w(y) - S(y)\right) & if\ S(y) < r_p + w(y), \end{cases} \qquad (7)$$

[14] Specifically, let $x := w(y) - w(g)$. Replace the global concavity from Assumption 4, $v''(x) < 0$ for all $x$, for $v''(x) < 0$ if $x \geq 0$ and $u''(x) > 0$ if $x < 0$. An assumption made in Kőszegi and Rabin (2006, 2007). This assumption is accompanied with losses entering positively in the agent's utility, so instead of having $-u(-x)$ for $x < 0$, as assumed throughout this paper, such a loss enters the utility as $u(-x)$.

where $r_p \geq 0$ and $\lambda_p > 1$.

The principal is loss averse. In comparison to Assumption 4, loss aversion applies to her benefit and cost. An assumption consistent with the notion that these biases apply to monetary outcomes.[15] Furthermore, diminishing sensitivity is not included in Eq. (7). This assumption can be justified on the grounds of the principal being able to pool multiple risks and, as a result, not exhibiting utility curvature.

The solutions to the principal's program when she is loss averse are similar to those of Theorem 1. The only difference appears for the case in which output is intermediate, so that she is in the domain of losses while the agent is in the domain of gains. Proofs of the main results in this section are relegated to Appendix C.

**Proposition 3.** *Let $\hat{y}_s \in (\underline{y}, \overline{y})$ be the unique output level from Theorem 1. Under A1-A4, and that the principal's preferences are given by (7), there exists a unique output level $\hat{y}_p \in [\underline{y}, \overline{y}]$ such that the second-best contract, $w_s^*(y)$:*

*i) Is identical to the contract presented in Theorem 2 (i) and (ii) if $\hat{y}_p < \hat{y}_s$.*

*ii) Pays the minimum possible if $y < \hat{y}_s$, exhibits a bonus at $y = \hat{y}_s$, increases in performance in $y > \hat{y}_s$ but at a lower rate in the segment $y \in (\hat{y}_s, \hat{y}_p)$ if $\hat{y}_p \geq \hat{y}_s$ and $U(e, w_s^*(\tilde{y}), r)$ is S-shaped.*

*iii) Exhibits a bonus at $y = \hat{y}_s$, increases in performance in $y < \hat{y}_s$ and $y > \hat{y}_s$, but at a lower rate in the segment $y \in (\hat{y}_s, \hat{y}_p)$ if $\hat{y}_p \geq \hat{y}_s$ and $U(e, w_s^*(\tilde{y}), r)$ is concave.*

---

15 Another possible representation of reference dependence is

$$\Pi(S(y), r_p, w) = \begin{cases} P\big(S(y) - r_p\big) - w(y) & \text{if } S(y) \geq r_p , \\ -\lambda_p P\left(r_p - S(y)\right) - w(y) & \text{if } S(y) < r_p. \end{cases}$$

This representation consistent with the approach taken throughout the paper to model reference dependence for the agent. Notice that this assumption together with the assumption that the agent's preference is given by Eq. (2) imply that the contracts in Theorem 1 remain optimal. The principal's loss aversion and diminishing sensitivity do not apply to her cost component, $w(y)$, so the principal's problem is the same.

When output is high enough to ensure $S(y) \geq r_p + w(y)$, the principal is located in the domain of gains and her objective function is identical to that in the standard problem studied in Section 3. In this case loss aversion does not affect contracting. As a result, the optimal contracts are exactly those presented in Theorem 1. This solution is presented in Proposition 3 (i).

Instead, for low output levels ensuring $S(y) < r_p + w(y)$, the principal's loss aversion can affect optimal contracting. If output is sufficiently low so that $y < \hat{y}_s$ also holds, the principal transfers most of the risk to the agent by offering low payments. If $U(e, w_s^*(\tilde{y}), r)$ is S-shaped, the payment becomes the lowest possible and all risk is transferred to the risk seeking agent. Instead, if $U(e, w_s^*(\tilde{y}), r)$ is concave the risk averse agent cannot tolerate that excessive risk exposure and is given positive transfers that nevertheless locate him in the domain of losses. This shape of the optimal contract at low output levels is, again, as in Theorem 1.

In the contingency that output is intermediate, so that $S(y) < r_p + w(y)$ continues to hold but $y \geq \hat{y}_s$ is also true, the principal offers more insurance to the agent. That is why the agent is awarded a bonus that transitions him to the domain of gains as in Theorem 1. However, the loss averse principal is willing to transfer more risk to the agent as compared to the expected utility principal. This is achieved by offering lower-powered incentives in the segment $y \geq \hat{y}_s$. A differential in incentives that leads to a kink at $\hat{y}_p$. The point at which the principal herself transitions from losses to gains, making her willing to offer incentives that are as high-powered as in Theorem 1. This slight modification to Theorem 1 is reflected in the second and third part of Proposition 3.

### 5.2 Adverse Selection followed by Moral Hazard

The assumption that the principal is fully informed about the agent's risk preferences is typically made in models of moral hazard. However, this assumption becomes more prominent and stringent in the framework of this paper. This extension considers a setting in which she does not exactly know the agent's risk preferences.

Suppose that the principal is informed about the agent's utility shape but that his loss aversion is unknown. For simplicity, assume that she can contract with agents with either high or low degrees of loss aversion, $\lambda_i \in \{\lambda_L, \lambda_H\}$ where $\lambda_H > \lambda_L > 1$. Contracting with an agent with $\lambda_H$ occurs with probability $\omega$, while contracting with an agent with $\lambda_L$ occurs with the complement probability, $1-\omega$.

The timing of the interaction between agent and principal is as follows. First, nature moves and determines $\lambda_i$ which is private information to the agent. Second, the principal offers a menu of contracts $w(y)^i$ with $i = \{L, H\}$. Third, the agent self-selects into a contract. Fourth, $e$ is chosen by the agent. Finally, $y$ realizes and the agent is paid according to the transfers specified in the selected contract.

The principal's objective is to offer a menu of contracts that motivates all agents to exert high effort, participate, and that enables her to screen agents according to their degree of loss aversion. The following proposition shows that the optimal menu of contracts $w(y)^i$ consists of the bonus contracts from Theorem 1 complemented with informational rents.

**Proposition 4.** *Under A1-A4 and that $\lambda_i \in \{\lambda_L, \lambda_H\}$ is unknown to the principal, the optimal menu of contracts is $\{w_S^*(y)^H, w_S^*(y)^L\}$ such that*

*i) $w_S^*(y)^H$ is the optimal second-best contract from Theorem 1 (i) - (ii),*

*ii) $w_S^*(y)^L$ is the second-best optimal contract from Theorem 1 (i) - (ii) satisfying $U(e_H, w_S^*(y)^L, r, \lambda_L) = \overline{U} + (\lambda_H - \lambda_L)\int_{\underline{y}}^{\hat{y}_H} v(u(r) - u(w_S^*(y)^H)) f(y|e_H)\mathrm{d}y$.*

*Where $\hat{y}_H \in (\underline{y}, \overline{y})$ is the output level after which contract $w_S^*(y)^H$ awards the bonus.*

The optimal menu of contracts consists of a tuple of bonus contracts. According to Theorem 1, this contract shape guarantees that both types of agents exert high effort at the lowest possible cost for the principal. Importantly, to screen agents with different levels of loss aversion, the principal offers informational rents to more efficient agents, that is agents with lower loss aversion. These rents have the goal of deterring these agents from mimicking agents with high loss aversion. This is achieved by offering a rent that makes them indifferent between engaging in a strategy of mimicking or not. To conclude, note that agents with high loss aversion are not willing to mimic agents with low loss aversion, since, as suggested by Corollary 2, choosing such a contract would lead to sizeable disutility.

## 6. Conclusion

This paper provided a preference foundation for the conventional compensation practice of offering bonuses. A contract with a bonus exploits the agent's loss aversion and diminishing sensitivity in a way that allows the principal to offer insurance and generate motivation in a cost-effective way. I also demonstrated that regardless of the theory of risk chosen to characterize reference-dependent preferences, the rule chosen to define a reference point, and the set of assumptions made about the principal's preferences and knowledge, the bonus feature of the contract emerges as optimal solution to the contracting problem.

## References

Abdellaoui, M., Bleichrodt, H., & l'Haridon, O. (2008). A tractable method to measure utility and loss aversion under prospect theory. *Journal of Risk and uncertainty*, *36*(3), 245-266.

Abdellaoui, M., Bleichrodt, H., l'Haridon, O., & Van Dolder, D. (2016). Measuring loss aversion under ambiguity: A method to make prospect theory completely observable. *Journal of Risk and uncertainty*, *52*(1), 1-20.

Abdellaoui, M., Bleichrodt, H., & Paraschiv, C. (2007). Loss aversion under prospect theory: A parameter-free measurement. *Management Science*, *53*(10), 1659-1674.

Abeler, J., Falk, A., Goette, L., & Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, *101*(2), 470-92.

Allen, E. J., Dechow, P. M., Pope, D. G., & Wu, G. (2017). Reference-dependent preferences: Evidence from marathon runners. *Management Science*, *63*(6), 1657-1672.

Baillon, A., Bleichrodt, H., & Spinu, V. (2020). Searching for the reference point. *Management Science*, *66*(1), 93-112.

Bell, D. E. (1985). Disappointment in decision making under uncertainty. *Operations research*, *33*(1), 1-27.

Borch K.H. (1962). Equilibrium in a Reinsurance Market. *Econometrica, 30*, 424-444.

Brookins, P., Goerg, S., & Kube, S. (2017). Self-chosen goals, incentives, and effort. *Unpublished manuscript*.

Chateauneuf, A., & Wakker, P. (1999). An axiomatization of cumulative prospect theory for decision under risk. *Journal of Risk and Uncertainty*, *18*(2), 137-145.

Clark, D., Gill, D., Prowse, V., & Rush, M. (2020). Using goals to motivate college students: Theory and evidence from field experiments. *Review of Economics and Statistics*, *102*(4), 648-663.

Corgnet, B., Gómez-Miñambres, J., & Hernán-Gonzalez, R. (2015). Goal setting and monetary incentives: When large stakes are not enough. *Management Science*, *61*(12), 2926-2944

Corgnet, B., Gómez-Miñambres, J., & Hernán-Gonzalez, R. (2018). Goal setting in the principal–agent model: Weak incentives for strong performance. *Games and Economic Behavior*, *109*, 311-326.

Corgnet, B., & Hernán-González, R. (2019). Revisiting the trade-off between risk and incentives: The shocking effect of random shocks?. *Management Science*, *65*(3), 1096-1114.

Dalton, P. S., Ghosal, S., & Mani, A. (2016)a. Poverty and aspirations failure. *The Economic Journal*, *126*(590), 165-188.

Dalton, P. S., Gonzalez, V., & Noussair, C. N. (2016)b. Self-chosen goals: Incentives and gender differences.

De Meza, D., & Webb, D. C. (2007). Incentive design under loss aversion. *Journal of the European Economic Association*, *5*(1), 66-92.

Delquié, P., & Cillo, A. (2006). Disappointment without prior expectation: a unifying perspective on decision under risk. *Journal of Risk and Uncertainty*, *33*(3), 197-215.

Dittmann, I., Maug, E., & Spalt, O. (2010). Sticks or carrots? Optimal CEO compensation when managers are loss averse. *The Journal of Finance*, *65*(6), 2015-2050.

Gneezy, U., Goette, L., Sprenger, C., & Zimmermann, F. (2017). The limits of expectations-based reference dependence. *Journal of the European Economic Association*, *15*(4), 861-876.

González-Jiménez, V, Castillo G. (2021) Optimal incentives without expected utility. Unpublished.

Gómez-Miñambres, J. (2012). Motivation through goal setting. *Journal of Economic Psychology*, *33*(6), 1223-1239.

Gul, F. (1991). A theory of disappointment aversion. *Econometrica: Journal of the Econometric Society*, 667-686.

Harding, M., & Hsiaw, A. (2014). Goal setting and energy conservation. *Journal of Economic Behavior & Organization*, *107*, 209-227.

Heath, C., Larrick, R. P., & Wu, G. (1999). Goals as reference points. *Cognitive psychology*, *38*(1), 79-109.

Hershey, J. C., & Schoemaker, P. J. (1985). Probability versus certainty equivalence methods in utility measurement: Are they equivalent?. *Management Science*, *31*(10), 1213-1231

Herweg, F., Müller, D., & Weinschenk, P. (2010). Binary payment schemes: Moral hazard and loss aversion. *American Economic Review*, *100*(5), 2451-77.

Holmström, B. (1979). Moral hazard and observability. *The Bell journal of economics*, 74-91.

Jewitt, I., Kadan, O., & Swinkels, J. M. (2008). Moral hazard with bounded payments. *Journal of Economic Theory*, *143*(1), 59-82.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291.

Kim, S. K. (1997). Limited liability and bonus contracts. *Journal of Economics & Management Strategy*, *6*(4), 899-913.

Koch, A. K., & Nafziger, J. (2011). Self‐regulation through goal setting. *Scandinavian Journal of Economics*, *113*(1), 212-227.

Koch, A. K., & Nafziger, J. (2016). Goals and bracketing under mental accounting. *Journal of Economic Theory*, *162*, 305-351.

Köbberling, V., & Wakker, P. P. (2005). An index of loss aversion. *Journal of Economic Theory*, *122*(1), 119-131.

Kőszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, *121*(4), 1133-1165.

Kőszegi, B., & Rabin, M. (2007). Reference-dependent risk attitudes. *American Economic Review*, *97*(4), 1047-1073.

Larrick, R. P., Heath, C., & Wu, G. (2009). Goal-induced risk taking in negotiation and decision making. *Social Cognition*, *27*(3), 342-364.

Loomes, G., & Sugden, R. (1986). Disappointment and dynamic consistency in choice under uncertainty. *The Review of Economic Studies*, *53*(2), 271-282.

Markle, A., Wu, G., White, R., & Sackett, A. (2018). Goals as reference points in marathon running: A novel test of reference dependence. *Journal of Risk and Uncertainty*, *56*(1), 19-50.

Mirrlees, J. A. (1976). The optimal structure of incentives and authority within an organization. *The Bell Journal of Economics*, 105-131.

Oyer, P. (2000). A theory of sales quotas with limited liability and rent sharing. *Journal of labor Economics*, *18*(3), 405-426.

Park, E. S. (1995). Incentive contracting under limited liability. *Journal of Economics & Management Strategy*, *4*(3), 477-490.

Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, *68*(5), 1281-1281.

Masatlioglu, Y., & Raymond, C. (2016). A behavioral analysis of stochastic reference dependence. *American Economic Review*, *106*(9), 2760-82.

Sprenger, C. (2015). An endowment effect for risk: Experimental tests of stochastic reference points. *Journal of Political Economy*, *123*(6), 1456-1499.

Terzi, A., Koedijk, K., Noussair, C. N., & Pownall, R. (2016). Reference point heterogeneity. *Frontiers in psychology*, *7*, 1347.

Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R., & Staake, T. (2018). Overcoming salience bias: How real-time feedback fosters resource conservation. *Management science*, *64*(3), 1458-1476.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), 297-323.

Vieider, F. M., Lefebvre, M., Bouchouicha, R., Chmura, T., Hakimov, R., Krawczyk, M., & Martinsson, P. (2015). Common components of risk and uncertainty attitudes across contexts and domains: Evidence from 30 countries. *Journal of the European Economic Association*, *13*(3), 421-452.

Von Gaudecker, H. M., Van Soest, A., & Wengstrom, E. (2011). Heterogeneity in risky choice behavior in a broad population. *American Economic Review*, *101*(2), 664-94.

Worldatwork. (2018). *2018 Incentive Pay Practices for Publicly-Traded Companies*.

Wu, G., Heath, C., & Larrick, R. (2008). A prospect theory model of goal behavior. *Unpublished manuscript*.

# Appendix A

**Lemma 1**

According to Definition 1 that $U$ is S-shaped or concave is determined by its shape in the domain of losses. Hence, consider a realization $\tilde{y} \in [\underline{y}, \overline{y}]$ such that $w(\tilde{y}) < r$. Eq. (2) becomes:

$$U(e, w(\tilde{y}), r) = \int_{\underline{y}}^{\overline{y}} \left(\phi u\big(w(\tilde{y})\big) - \lambda\eta v\left(u(r) - u\big(w(\tilde{y})\big)\right)\right) f(y|e)\mathrm{d}y - c(e). \qquad (A1)$$

The second derivative of (A1) with respect to $w$ is

$$U''(e, w(\tilde{y}), r) = u''\big(w(\tilde{y})\big)\left(\phi + \lambda\eta v'\left(u(r) - u\big(w(\tilde{y})\big)\right)\right) - \left(u'\big(w(\tilde{y})\big)\right)^2 \left(\lambda\eta v''\left(u(r) - u\big(w(\tilde{y})\big)\right)\right). \qquad (A2)$$

A sufficient and necessary condition for $U(e, w(\tilde{y}), r)$ to be convex in the domain of losses is $U''(e, w(\tilde{y}), r) \geq 0$. That condition can be rewritten using Eq. (A2) as follows

$$-\frac{u''\big(w(\tilde{y})\big)}{u'\big(w(\tilde{y})\big)} \leq -\frac{u'\big(w(\tilde{y})\big)\left(\lambda\eta v''\left(u(r) - u\big(w(\tilde{y})\big)\right)\right)}{\left(\phi + \lambda\eta v'\left(u(r) - u\big(w(\tilde{y})\big)\right)\right)} \qquad (A3)$$

for any $\tilde{y} \in [\underline{y}, \overline{y}]$ such that $w(\tilde{y}) < r$. ■

**Theorem 1**

The proof follows four parts. In part $i$) the Lagrangian and first-order conditions are presented. Part $ii$) shows that when $U(e_H, w(\tilde{y}), r)$ is S-shaped for any $\tilde{y} \in [\underline{y}, \overline{y}]$, the solutions from the first-order conditions are not valid, and an alternative solution is given. Part $iii$) presents the properties of the second-best optimal contract for the two relevant cases, when $U(e_H, w(\tilde{y}), r)$ is concave and S-shaped. Finally, iv) presents the properties of the first-best optimal contract.

*i) Lagrangian and first-order conditions*

Denote by $\mu \geq 0$ and $\gamma \geq 0$ the Lagrangian multipliers of the agent's participation and incentive compatibility constraints, respectively. The Lagrangian of the principal's maximization problem writes as follows

$$
\begin{aligned}
\mathcal{L}(w,e) = {} & \big(S(y) - w(y)\big) f(y|e_H) \\
& + \mu \Big[\phi u\big(w(y)\big) f(y|e_H) + \theta_{\mathbb{I}} \eta v\Big(u\big(w(y)\big) - u(r)\Big) f(y|e_H) \\
& - \lambda(1-\theta_{\mathbb{I}}) \eta v\Big(u(r) - u\big(w(y)\big)\Big) f(y|e_H) - c\Big] \\
& + \gamma \Big[\phi u\big(w(y)\big) f(y|e_H) + \theta_{\mathbb{I}} \eta v\Big(u\big(w(y)\big) - u(r)\Big) f(y|e_H) \\
& - \lambda(1-\theta_{\mathbb{I}}) \eta v\Big(u(r) - u\big(w(y)\big)\Big) f(y|e_H) - c - \phi u\big(w(y)\big) f(y|e_L) \\
& - \theta_{\mathbb{I}} \eta v\Big(u\big(w(y)\big) - u(r)\Big) f(y|e_L) \\
& + \lambda(1-\theta_{\mathbb{I}}) \eta v\Big(u(r) - u\big(w(y)\big)\Big) f(y|e_L)\Big]. \qquad (A4)
\end{aligned}
$$

Pointwise optimization with respect to $w(y)$ gives

$$
\begin{aligned}
-f(y|e_H) + \mu \Big[\phi f(y|e_H) + \theta_{\mathbb{I}} \eta v'\Big(u\big(w(y)\big) - u(r)\Big) f(y|e_H) \\
+ \lambda(1-\theta_{\mathbb{I}}) \eta v'\Big(u(r) - u\big(w(y)\big)\Big) f(y|e_H)\Big] u'\big(w(y)\big) \\
+ \gamma \Big[\phi\big(f(y|e_H) - f(y|e_L)\big) + \theta_{\mathbb{I}} \eta v'\Big(u\big(w(y)\big) - u(r)\Big)\big(f(y|e_H) - f(y|e_L)\big) \\
+ \lambda(1-\theta_{\mathbb{I}}) \eta v'\Big(u(r) - u\big(w(y)\big)\Big)\big(f(y|e_H) - f(y|e_L)\big)\Big] u'\big(w(y)\big) = 0 \\
(A5)
\end{aligned}
$$

Denote by $w_s^F(y)$ the transfer satisfying $(A5)$. The following expressions are obtained after some manipulations:

$$
\frac{1}{u'(w_s^F(y))\Big(\phi + \eta v'\big(u(w_s^F(y))\big) - u(r)\big)\Big)} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right) \qquad (A6)
$$

if $\theta_{\mathbb{I}} = 1$, and

$$
\frac{1}{u'(w_s^F(y))\Big(\phi + \lambda\eta v'\big(u(r) - u(w_s^F(y))\big)\Big)} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right) \qquad (A7)
$$

if $\theta_{\mathbb{I}} = 0$. The derivative of $(A6)$ with respect to $y$ gives

$$
\frac{dw_s^F(y)}{dy} = \frac{\gamma \frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right)\Big(u'(w_s^F(y))\Big(\phi + \eta v'\Big(u\big(w_s^F(y)\big) - u(r)\Big)\Big)\Big)^2}{\Big(u''\big(w_s^F(y)\big)\Big(\phi + \eta v'\Big(u\big(w_s^F(y)\big) - u(r)\Big)\Big) + \Big(u'\big(w_s^F(y)\big)\Big)^2\Big(\phi\eta v''\Big(u\big(w_s^F(y)\big) - u(r)\Big)\Big)\Big)}, \qquad (A8)
$$

Since $u'' < 0$ and $v'' < 0$ and because $\frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right) \leq 0$ (Assumption 2), Eq. $(A8)$ shows that $\frac{dw_s^F(y)}{dy} \geq 0$. In the domain of gains, $w_s^F(y)$ is non-decreasing in performance.

The derivative of $(A7)$ with respect to $y$ gives:

$$\frac{dw_s^F(y)}{dy} = \frac{\gamma \frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right)\left(u'(w_s^F(y))\left(\phi + \lambda\eta v'\ \left(u(r) - u\left(w_s^F(y)\right)\right)\right)\right)^2}{\left(u''(w_s^F(y))\left(\phi + \lambda\eta v'\ \left(u(r) - u\left(w_s^F(y)\right)\right)\right) - \left(u'\left(w_s^F(y)\right)\right)^2\left(\phi + \lambda\eta v''\left(u(r) - u\left(w_s^F(y)\right)\right)\right)\right)}, \quad (A9)$$

Due to Assumption 2, $\frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right) \leq 0$, the numerator of the right-hand side of $(A9)$ is negative. Lemma 1 shows that the denominator of Eq. $(A9)$ is also negative if $U(e_H, w(\tilde{y}), r)$ is concave for any $\tilde{y} \in [\underline{y}, \overline{y}]$. In that case $\frac{dw_s^F(y)}{dy} \geq 0$. Instead, if $U(e_H, w(\tilde{y}), r)$ is S-shaped, as defined by Definition 1, Lemma 1 gives $\frac{dw_s^F(y)}{dy} \leq 0$. In the domain of losses, the behavior of $w_s^F(y)$ depends on the shape of $U(e_H, w(\tilde{y}), r)$. If that shape is concave, $w_s^F(y)$ is non-decreasing in performance, if S-shaped, $w_s^F(y)$ can decrease in performance.

*ii) Solution when $U(e_H, w(\tilde{y}), r)$ is S-shaped*

It is well-known that if $U(e_H, w(\tilde{y}), r)$ is concave for any $\tilde{y} \in [\underline{y}, \overline{y}]$, the solutions $w_s^F(y)$ are necessary and sufficient to solve the maximization problem given in Eq. (3). However, these solutions are not optimal when $U(e_H, w(\tilde{y}), r)$ is S-shaped. In that case, the principal is better off offering a lottery $L := (p: r, 1 - p: 0)$. Since $\phi u(w_s^F(y)) - \lambda v\left(u(r) - u(w_s^F(y))\right)$ increases in $w_s^F(y)$, there must exist a number $p \in (0,1)$ such that:

$$\phi u(w_s^F(y)) - \lambda v\left(u(r) - u\left(w_s^F(y)\right)\right) = \phi p u(r) - (1 - p)\lambda v\left(u(r)\right). \qquad (A10)$$

Therefore, replacing $w_s^F(y)$ from $(A4)$ by $L$ with a probability satisfying $(A10)$ leaves the agent's participation and incentive compatibility constraints unchanged. Equation $(A10)$ and the convexity of $U(e_H, w(\tilde{y}), r)$ in the domain of losses (Lemma 1), imply:

$$\phi\left(u\left(w_{sb}^F(y)\right) - pu(r)\right) = \lambda v\left(u(r) - u(w_s^F(y))\right) - (1 - p)\lambda v\left(u(r)\right)$$

$$\leq \lambda v\left(u(r) - u\left(w_s^F(y)\right)\right) - \lambda v\left((1 - p)u(r)\right). \qquad (A11)$$

Since $v' > 0$, the last inequality in $(A11)$ implies

$$w_s^F(y) > pr. \qquad (A12)$$

Hence, $L$ is more cost-effective for the principal than the solution implied by $(A7)$.

I turn to analyze the incentives of $L$ for a given $p \in (0,1)$. Let $\overline{L} := pr$. Substituting $\overline{L}$ into the agent's utility gives:

$$U(e_H, L, r) = \left(\frac{\overline{L}}{r}\right)\phi u(r) - \left(1 - \frac{\overline{L}}{r}\right)\lambda\eta v\left(u(r)\right) - c, \qquad (A13)$$

which is linear in $\overline{L}$. Therefore, changes in $\overline{L}$, through adjustments of $p$, do not alter the agent's marginal expected utility. To analyze whether and in which output segments these changes in $p$ apply, denote by $\hat{y}_s \in [\underline{y}, \overline{y}]$ the performance level satisfying:

$$\frac{1}{\frac{\phi u(r) + \lambda\eta v\big(u(r)\big)}{r}} = \nu + \gamma\left(1 - \frac{f(\hat{y}_s|e_L)}{f(\hat{y}_s|e_H)}\right). \qquad (A14)$$

The existence and uniqueness of $\hat{y}_s$ is guaranteed by the fact that the left-hand side of $(A14)$ is positive and constant in $y$, while the right-hand side of that equation increases with $y$ in the domain $[0, \infty)$ (Assumption 2).

Eq. $(A14)$ implies that the marginal incentives of $L$ for a given $p \in (0,1)$ either do not incentivize high effort or lead to non-binding constraints in $y \in [\underline{y}, \overline{y}] \backslash \{\hat{y}_s\}$. If $y < \hat{y}_s$, the right-hand side of Eq. $(A14)$ is larger than the left-hand side of that equation. In that case, the expected value of the lottery, $\overline{L}$, can be reduced by decreasing $p$. Eq. $(A13)$ shows that a reduction of $\overline{L}$ does not change the agent's marginal utility. Hence, is optimal to set $p = 0$. Instead, for $y > \hat{y}_s$ the expected value of the lottery $\overline{L}$ should be increased. Again, since $U(e_H, L, r)$ is linear in $\overline{L}$, is optimal to set $p = 1$.

*iii) Properties of the optimal contract*

Let $U(e_H, w(\tilde{y}), r)$ be S-shaped for any realization $\tilde{y} \in [\underline{y}, \overline{y}]$. Since the agent is loss averse, $\lambda > 1$, $w_s^*(y) = 0$, which is equivalent to $\hat{y}_s = \underline{y}$, cannot be a solution as it induces considerable disutility, leading the agent to reject the contract. Also, the solution given by Eq. $(A6)$, which is equivalent to $\hat{y}_s = \overline{y}$, cannot be optimal on its own because the principal could deviate from that solution by offering $w_s^*(y) = 0$ at the low end of the output space. That deviation is not only more cost-effective but also motivates the agent to exert high effort to avoid the disutility from experiencing losses. Hence, $\hat{y}_s \in \left(\underline{y}, \overline{y}\right)$.

The optimal contract is given by:

$$w_s^*(y) = \begin{cases} 0 & if\ y < \hat{y}_s, \\ w_s^F(y) \text{ satisfying } (A6) & if\ y \geq \hat{y}_s. \end{cases} \qquad (A15)$$

Notice that this solution exhibits a discrete jump at $y = \hat{y}_s$ since $\lim_{y \to \hat{y}_s^+} w_s^*(y) > r$, and $\lim_{y \to \hat{y}_s^-} w_s^*(y) = 0$. This proves the first part of the Theorem.

Let $U(e_H, w(\tilde{y}), r)$ be concave for any $\tilde{y} \in [\underline{y}, \overline{y}]$. The optimal contract, $w_s^*(y)$, consists of two components: $w_s^F(y)$ satisfying $(A6)$, which implies $w_s^F(y) \geq r$, and $w_s^F(y)$ satisfying $(A7)$, which implies $w_s^F(y) < r$. Showing that $\hat{y}_s \in \left(\underline{y}, \overline{y}\right)$ for this case follows a similar rationale as the one applied above. The transfer $w_f^F(y)$ satisfying $(A7)$ cannot be a solution on its own as leads to disutility and will be rejected by the agent. Also, the solution given by Eq.$(A6)$ is not optimal on its own, as the principal would be better off exposing the agent to losses at low output levels. Hence, the optimal contract combines the first-order conditions $(A6)$ and $(A7)$. The transition is given by the output level $\hat{y}_s \in (\underline{y}, \overline{y})$ satisfying:

$$\phi \int_{\hat{y}_s}^{\overline{y}} u(w_s^F(y)) f(y|e_H)\,\mathrm{d}y + \eta \int_{\hat{y}_s}^{\overline{y}} v\big(u(w_s^F(y)) - u(r)\big) f(y|e_H)\,\mathrm{d}y + \phi \int_{\underline{y}}^{\hat{y}_s} u(w_s^F(y)) f(y|e_H)\mathrm{d}y$$
$$- \lambda\eta \int_{\underline{y}}^{\hat{y}_s} v\big(u(r) - u(w_s^F(y))\big) f(y|e_H)\mathrm{d}y - c = \overline{U}. \qquad (A16)$$

The existence and uniqueness of $\hat{y}_s$ is guaranteed by the fact that the left-hand side of Eq. $(A16)$ can be negative if $\hat{y}_s = \overline{y}$, increases as $\hat{y}_s$ decreases, and is positive if $\hat{y}_s = \underline{y}$.

As a result, the optimal incentive scheme is given by:

$$w_s^*(y) = \begin{cases} w_s^F(y) & satisfying\ (A6)\ \ if\ y \geq \hat{y}_s\,, \\ w_s^F(y) & satisfying\ (A7)\ \ if\ y < \hat{y}_s. \end{cases} \qquad (A17)$$

Notice that this solution exhibits a discrete jump at $y = \hat{y}_s$ since $\lambda > 1$ appears in the denominator of the right-hand side of Eq. $(A7)$ but this coefficient does not enter in Eq. $(A6)$. This proves the second part of the Theorem.

*iv) Optimal first-best contract*

Let $\gamma = 0$. Denote by $w_f^F(y)$ the candidate solution from the first-order approach under this restriction. Eq. $(A5)$ collapses to

$$\frac{1}{u'\big(w_f^F(y)\big)\Big(\phi + \eta v'\Big(u\big(w_f^F(y)\big)\big) - u(r)\Big)\Big)} = \mu, \qquad (A18)$$

if $\theta_{\mathbb{I}} = 1$, and

$$\frac{1}{u'\big(w_f^F(y)\big)\Big(\phi + \lambda\eta v'\ \Big(u(r) - u\big(w_f^F(y)\big)\Big)\Big)} = \mu, \qquad (A19)$$

if $\theta_{\mathbb{I}} = 0$. Equations $(A18)$ and $(A19)$ show that $\frac{dw_f^F(y)}{dy} = 0$. Hence, $w_f^F(y)$ is performance insensitive.

As in the derivation of the second-best contract, it can be shown that if $U(e_H, w(\tilde{y}), r)$ is S-shaped for any $\tilde{y} \in [\underline{y}, \overline{y}]$, the principal is better off offering a lottery $L = (p{:}\,r, 1 - p{:}\,0)$ rather than $w_f^F(y)$ satisfying $(A19)$. That lottery can be offered to the agent with a $p \in (0,1)$ that satisfies:

$$\phi u\left(w_f^F(y)\right) - \lambda v\left(u(r) - u\left(w_f^F(y)\right)\right) = \phi p u(r) - (1-p)\lambda v\big(u(r)\big). \qquad (A20)$$

The existence of the $p \in (0,1)$ satisfying Eq. $(A20)$ is guaranteed by the fact that $\phi u\left(w_f^F(y)\right) - \lambda v\left(u(r) - u\left(w_f^F(y)\right)\right)$ increases in $w_f^F(y)$ and but $\phi p u(r) - (1-p)\lambda v\big(u(r)\big)$ is constant. Therefore, replacing $w_f^F(y)$ from $(A19)$ by $L$ leaves the agent's participation constraint unchanged.

Equation $(A20)$ and the convexity of $U(e_H, w(\tilde{y}), r)$ for any output realization $\tilde{y}$ such that $w(\tilde{y}) < r$, imply:

$$\phi\left(u\left(w_f^F(y)\right) - pu(r)\right) \leq \lambda v\left(u(r) - u\left(w_f^F(y)\right)\right) - \lambda v\big((1-p)u(r)\big) \qquad (A21)$$

Since $v' > 0$, the last inequality implies $w_f^F(y) > pr$. Hence, $L$ is more cost-effective for the principal than the candidate solution from Eq. $(A19)$.

When $L$ is offered, the Lagrangian multiplier, $\mu$, can be large enough to ensure:

$$\frac{1}{\frac{\phi u(r) + \lambda\eta v\big(u(r)\big)}{r}} = \mu. \qquad (A22)$$

However, if $\mu$ is smaller, so that the left-hand side of $(A22)$ is larger than the right-hand side of the same equation, $\overline{L}$ should be reduced to ensure that the participation constraint binds with equality. Eq. $(A13)$ shows that a reduction of $\overline{L}$ does not change the agent's marginal utility. Thus, is optimal to set $p = 0$. In contrast, if $\mu$ is large enough, such that the left-hand side of $(A22)$ is smaller than the right-hand side of that equation, then $\overline{L}$, should be increased. To ensure participation. Again, since $U(e_H, L, r)$ is linear in $\overline{L}$, Eq. $(A13)$, then is optimal to set $p = 1$.

Therefore, when $U(e_H, w(\tilde{y}), r)$ is S-shaped for any $\tilde{y} \in \left[\underline{y}, \overline{y}\right]$, the optimal first-best contract, $w_f^*(y)$, consists of two components: $w_f^*(y) = 0$ and $w_f^*(y) = w_f^F(y)$, where $w_f^F(y)$ satisfies $(A18)$. These components cannot be implemented on their own. Suppose that $w_f^*(y) = 0$ is implemented on its own. That contract induces considerable disutility, and the agent would reject it. Next, suppose that only $w_f^*(y) = w_f^F(y)$ is offered. The principal can profitably deviate from that solution by paying $w_f^*(y) = 0$ for the lowest output levels. Due to the risk seeking attitudes of the agent in the domain of losses, i.e. convexity of $-v(\cdot)$, this contract will not be rejected. Therefore, the optimal contract consists of a combination of the schedules $w_f^*(y) = 0$, and $w_f^*(y) = w_f^F(y)$.

The transition from $w_f^*(y) = 0$ to $w_f^*(y) = w_f^F(y)$ satisfying Eq. $(A18)$ is defined next. Let $\hat{y}_f \in (\underline{y}, \overline{y})$ be the output level satisfying

$$\phi\int_{\hat{y}_f}^{\overline{y}} u\big(w_f^F(y)\big)f(y|e_H)\,\mathrm{d}y+\eta\int_{\hat{y}_f}^{\overline{y}} v\left(u\big(w_f^F(y)\big)-u(r)\right)f(y|e_H)\,\mathrm{d}y+\phi\int_{\underline{y}}^{\hat{y}_f} u(r)f(y|e_H)\mathrm{d}y$$
$$-\lambda\eta\int_{\underline{y}}^{\hat{y}_f} v\big(u(r)\big)f(y|e_H)\mathrm{d}y-c\ =\overline{U}. \qquad (A23)$$

The existence of $\hat{y}_f$ is guaranteed by the fact that the left-hand side of Eq. $(A23)$ can be negative if $\hat{y}_f=\overline{y}$ , increases as $\hat{y}_f$ decreases, and is positive if $\hat{y}_f=\underline{y}$ . As a result, the optimal incentive scheme is given by:

$$w_f^*(y)=\begin{cases}0 & if\ y<\hat{y}_f,\\ w_f^F(y) & if\ y\geq\hat{y}_f.\end{cases} \qquad (A24)$$

This solution exhibits a discrete jump at $y=\hat{y}_f$ since $\lim_{y\to\hat{y}_f^{\,+}} w_f^*(y)>r$, and $\lim_{y\to\hat{y}_f^{\,-}} w_f^*(y)=0$. This proves the third part of the Theorem.

When $U(e_H,w(\tilde{y}),r)$ is concave for any $\tilde{y}\in\left[\underline{y},\overline{y}\right]$, the solution also consists of two components: $w_f^F(y)$ satisfying $(A18)$, which implies $w_f^F(y)\geq r$, and $w_f^F(y)$ satisfying $(A19)$, which implies $w_f^F(y)<r$. Since the agent is loss averse, $\lambda>1$, $w_f^F(y)$ satisfying $(A19)$ cannot be a solution on its own as it induces losses. This leads the agent to reject the contract. Moreover, a combination of these two components exposes the agent to the risk of experiencing losses. Since $U(e_H,w(y),r)$ is concave, such a combination does not provide full insurance to this risk averse agent. Hence, it must be that the schedule $w_f^F(y)$ satisfying $(A18)$ is offered. This proves the last part of the Theorem. ■

**Corollary 1**

Part $i$). Let $U(e_H,w(\tilde{y}),r)$ be S-shaped for any $\tilde{y}\in\left[\underline{y},\overline{y}\right]$. I proceed by contradiction by supposing that $\hat{y}_f>\hat{y}_s$. If the bonus of $w_s^*(y)$ is at least as large than that of $w_f^*(y)$, then $w_s^*(y)$ does not impart punishments. That is because $w_s^*(y)=w_f^*(y)=0$ if $y<\hat{y}_s$ and $w_s^*(y)>w_f^*(y)$ if $y>\hat{y}_s$. The second-best does not impart incentives to exert high effort at low output levels. A contradiction since at the optimum the incentive compatibility constraint binds. A similar rationale applies when the bonus of $w_s^*(y)$ is larger than that of $w_f^*(y)$.

If the bonus of $w_s^*(y)$ is smaller than that of $w_f^*(y)$, rewards are imparted non-monotonically. Denote by $\hat{y}_e$ the output level in $y>\hat{y}_f$ satifying $w_s^*(\hat{y}_e)=w_f(\hat{y}_e)$. The existence and uniqueness of that output level is ensured by the properties $w_s^*(\hat{y}_s)<w_f^*(\hat{y}_s)$, $w_s^*(y)$ increasing in performance in $y>\hat{y}_s$ and $w_f^*(y)$ being constant in performance in that segment. In $y\in(\hat{y}_s,\hat{y}_f)$, $w_s^*(y)$ implements rewards with respect to $w_f^*(y)$, followed by punishments in $y\in(\hat{y}_f,\hat{y}_e)$, to subsequently exhibit again rewards in $y>\hat{y}_e$. This implementation of incentives does not incentivize high effort in $y\in(\hat{y}_s,\hat{y}_e)$. A contradiction since at the optimum the incentive compatibility constraint binds. Hence, it must be that $\hat{y}_e > \hat{y}_s \geq \hat{y}_f$. Punishments are imparted in losses and gains.

Part $ii$). Let $U(e_H, w(\tilde{y}), r)$ be concave for any realization $\tilde{y} \in \left[\underline{y}, \overline{y}\right]$. The optimal first-best contract $w_f^*(y)$ from Theorem 1, offers full insurance by bringing the agent to the domain of gains. The optimal contract $w_s^*(y)$ from Theorem 1, brings the agent to the domain of gains in $y > \hat{y}_s$. Hence, it must be that $\hat{y}_e$ , the unique output level that satisfies $w_s^*(\hat{y}_e) = w_f(\hat{y}_e)$ is in $y > \hat{y}_s$. Consequently, punishments are imparted in losses and gains. ■

**Corollary 2.**

Part $i$). Let $U(e_H, w(\tilde{y}), r)$ be S-shaped for any realization $\tilde{y} \in \left[\underline{y}, \overline{y}\right]$. Implicit differentiation of Eq. $(A6)$ gives

$$\frac{dw_s^*(y)}{dr} = \frac{u'\left(w_s^*(y)\right)u'(r)\left(\eta v''\left(u(w_s^*(y)))\right) - u(r)\right)}{u''\left(w_s^*(y)\right)\left(\phi + \eta v'\left(u(w_s^*(y))\right) - u(r)\right)\right) + \left(u'\left(w_s^F(y)\right)\right)^2\left(\eta v''\left(u(w_s^*(y))\right) - u(r)\right)\right)}. \quad (A25)$$

Since $u'' < 0$ and $v'' < 0$, in the domain of gains $\frac{dw_s^*(y)}{dr} > 0$. The bonus increases with $r$.

To investigate the influence of changes in $r$ in the location of $\hat{y}_s$, compute the derivative of $r$ with respect to the left-hand side of Eq. $(A14)$ to obtain

$$\frac{d}{dr}\left(\frac{r}{\phi u(r) + \lambda\eta v\big(u(r)\big)}\right) = \frac{\left(\phi u(r) - \phi u'(r)r + \lambda\eta v\big(u(r)\big) - \lambda\eta v'\big(u(r)\big)u'(r)r\right)}{\left(\phi u(r) + \lambda\eta v\big(u(r)\big)\right)^2}. \quad (A26)$$

Using the Taylor theorem around zero gives $v(0) = 0 = \phi u(r) + \lambda\eta v\big(u(r)\big) - \phi u'(r)r - \lambda\eta v'\big(u(r)\big)u'(r)r + \frac{\phi u''(r)r^2}{2} + \frac{\lambda\eta\left(v''\left(u(r)\right)u'(r) + v'\left(u(r)\right)u''(r)\right)r^2}{2}$. Hence, Eq. $(A26)$ demonstrates that $\frac{d}{dr}\left(\frac{r}{\phi u(r)+\lambda\eta v(u(r))}\right) > 0$ since $u'' < 0$ and $v'' < 0$. Therefore, the equality in $(A14)$ is maintained for a higher output level under a higher $r$. The bonus is given at a higher threshold $\hat{y}_s$.

Let $U(e_H, w(\tilde{y}), r)$ be concave. In that case, the result from Eq. $(A25)$ that $\frac{dw_s^*(y)}{dr} > 0$ in the domain of gains continues to hold. Furthermore, implicit differentiation of Eq. $(A7)$ gives.

$$\frac{dw_s^*(y)}{dr} = \frac{-u'\left(w_s^*(y)\right)u'(r)\left(\lambda\eta v''\left(u(r) - u\left(w_s^*(y)\right)\right)\right)}{u''\left(w_s^*(y)\right)\left(\phi + \lambda\eta v'\left(u(r) - u\left(w_s^*(y)\right)\right)\right) - \left(u'\left(w_s^*(y)\right)\right)^2\left(\lambda\eta v''\left(u(r) - u\left(w_s^*(y)\right)\right)\right)}. \quad (A27)$$

Lemma 1 shows that the denominator of Eq. $(A27)$ is negative for any $\tilde{y} \in [\underline{y}, \overline{y}]$. Since $v'' < 0$, then in the domain of losses $\frac{dw_S^*(y)}{dr} < 0$. The bonus increases with $r$.

To investigate the influence of changes in $r$ on the location of the bonus, compute the derivative of Eq. $(A16)$ with respect to $r$ to obtain:

$$-\eta \int_{\hat{y}_S}^{\overline{y}} u'(r)\, v'\left(u\left(w_S^F(y)\right) - u(r)\right) f(y|e_H)\, \mathrm{d}y - \lambda\eta \int_{\underline{y}}^{\hat{y}_S} u'(r) v'\left(u(r) - u\left(w_S^F(y)\right)\right) f(y|e_H) \mathrm{d}y < 0 \qquad (A28)$$

Hence, for the equality $(A16)$ to hold under higher $r$, $\hat{y}_S$ needs to be smaller. The bonus is given at a lower threshold $\hat{y}_S$.

Part $ii$). Let $U(e_H, w(\tilde{y}), r)$ be S-shaped for any output realization $\tilde{y} \in [\underline{y}, \overline{y}]$. The left-hand side of Eq. $(A14)$ decreases as $\lambda$ increases. Hence, to maintain that equality, and due to Assumption 2, $\hat{y}_S$ must decrease. Hence, for larger $\lambda$, $w_S^*(y)$ exhibits a smaller segment in which $w_S^*(y) = 0$ is the solution. Finally, notice that $\lambda$ does not enter in Eq.$(A6)$, so changes in that parameter do not influence the shape of $w_S^*(y)$ in the domain of gains nor in the magnitude of the bonus.

Let $U(e_H, w(\tilde{y}), r)$ be concave. Implicit differentiation of Eq. $(A7)$ gives

$$\frac{dw_S^*(y)}{d\lambda} = \frac{-u'\left(w_S^F(y)\right)\left(\eta v'\left(u(r) - u\left(w_S^F(y)\right)\right)\right)}{u''\left(w_S^F(y)\right)\left(\phi + \lambda\eta v'\left(u(r) - u\left(w_S^F(y)\right)\right)\right) - \left(u'\left(w_S^F(y)\right)\right)^2 \left(\lambda\eta v''\left(u(r) - u\left(w_S^F(y)\right)\right)\right)}. \qquad (A29)$$

Lemma 1 shows that the denominator of Eq. $(A29)$ is negative. Since $u' > 0$, then $\frac{dw_S^*(y)}{d\lambda} > 0$. Moreover, note that $\lambda$ does not enter in Eq.$(A6)$. The bonus shrinks.

To investigate the influence of changes in $\lambda$ on the location of the bonus, note that the derivative of $(A14)$ with respect to $\lambda$ gives $-\eta \int_{\underline{y}}^{\hat{y}_S} v\left(u(r) - u(w_S^F(y))\right) f(y|e_H) \mathrm{d}y$, a negative expression. Hence, for the equality in $(A14)$ to hold under higher $\lambda$, $\hat{y}_S$ needs to be smaller. In that way the exposition of the agent to losses is reduced. ■

**Corollary 3.**

Let $\eta = 0$ and $\phi = 1$. Under these restrictions, $U(e_H, w(\tilde{y}), r)$ is concave for any output realization $\tilde{y} \in [\underline{y}, \overline{y}]$. Since $-\frac{u''(w(\tilde{y}))}{u'(w(\tilde{y}))} \geq 0$, the necessary and sufficient condition in Eq. $(A3)$ cannot hold. Therefore, the solutions satisfying Eq. $(A6)$ and Eq. $(A7)$ are necessary and sufficient to solve the principal's maximization problem.

Under the assumed restrictions, these two equations become:

$$\frac{1}{u'(w_S^*(y))} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right). \qquad \text{(A30)}$$

Rearranging the above expression leads to:

$$w_s^*(y) = h'\left(\frac{1}{\mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right)}\right). \qquad \text{(A31)}$$

Moreover, equation (A8) becomes:

$$\frac{dw_s^*(y)}{dy} = \frac{\gamma \frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right)\left(u'(w_s^F(y))\right)^2}{u''\left(w_s^F(y)\right)}. \qquad \text{(A32)}$$

Since $\frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right) \leq 0$ (Assumption 2) and $u'' < 0$, then $\frac{dw_s^*(y)}{dy} \geq 0$. The optimal second-best contract, given in $(A31)$, is everywhere increasing in performance. This proves the second part of the corollary.

We turn to study the first-best optimal contract. Therefore, in addition to $\eta = 0$ and $\phi = 1$, let $\gamma = 0$. In that case, $(A31)$ becomes

$$w_f^*(y) = h'\left(\frac{1}{\mu}\right). \qquad \text{(A33)}$$

Equation $(A32)$ shows that $\frac{dw_f^*(y)}{dy} = 0$. Full insurance is given to the agent with a performance-insensitive contract. This proves the first part of the corollary.

**Corollary 4.**

Let $\eta = 1$, $\phi = 0$, and $u' = 1$. Under these restrictions, $U(e_H, w(\tilde{y}), r)$ is S-shaped for any realization $\tilde{y} \in \left[\underline{y}, \overline{y}\right]$ since in the domain of losses, that is when $\theta_{\mathbb{I}} = 0$, $-\frac{v''\left(u(r)-u\left(w(\tilde{y})\right)\right)}{v'\left(u(r)-u\left(w(\tilde{y})\right)\right)} \geq 0$, corroborating the necessary and sufficient condition in Eq. $(A3)$. Hence, only the candidate solution from the first-order condition for the domain of gains is sufficient and necessary for solving the principal's problem. Under the assumed restrictions that solution, given by Eq. $(A6)$, becomes

$$\frac{1}{v'(w_s^*(y) - r)} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right). \qquad \text{(A34)}$$

Rearranging the above expression one obtains,

$$w_s^*(y) \ = r + f'\left(\frac{1}{\mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right)}\right). \qquad \text{(A35)}$$

Moreover, Eq. (A8) becomes

$$\frac{dw_s^*(y)}{dy} = \frac{\gamma \frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right)\left(v'\ (w_s^F(y) - r)\right)^2}{v''(w_s^F(y) - r)}. \qquad (A36)$$

Since $\frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right) \leq 0$ (Assumption 2) and $v'' < 0$, then $\frac{dw_s^*(y)}{dy} \geq 0$ if $\theta_{\mathbb{I}} = 1$.

Eq. $(A14)$ becomes:

$$\frac{1}{\frac{\lambda v(u(r))}{r}} = \mu + \gamma\left(1 - \frac{f(\hat{y}_{ps}|e_L)}{f(\hat{y}_{ps}|e_H)}\right). \qquad (A37)$$

So, the transition from losses to gains is given by the $\hat{y}_{ps}$ that satisfies $(A37)$. The existence and uniqueness of that output level is guaranteed by $\frac{1}{\frac{\lambda v(u(r))}{r}} > 0$, $\frac{1}{\frac{\lambda v(u(r))}{r}}$ being constant in performance, and $\frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right) \leq 0$ (Assumption 2). Furthermore, Theorem 1 shows that it is optimal to pay the lowest possible in the domain of losses, that is $w_s^*(y) = 0$ in $y < \hat{y}_{ps}$.

All in all, the optimal contract is given by $w_s^*(y) = \begin{cases} 0 & if\ y < \hat{y}_{ps}, \\ r + f'\left(\frac{1}{\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right)}\right) & if\ y \geq \hat{y}_{ps}. \end{cases}$ This proves the second part of the corollary.

We turn to study the first-best contract. Therefore, in addition to $\eta = 1$, $\phi = 0$, and $u' = 1$, let $\gamma = 0$. In that case, Eq. $(A35)$ becomes

$$w_s^*(y) \ = r + f'\left(\frac{1}{\mu}\right). \qquad \text{(A38)}$$

Eq. $(A36)$ shows that $\frac{dw_f^*(y)}{dy} = 0$ if $\theta_{\mathbb{I}} = 1$. Moreover, Eq. $(A23)$ becomes under the assumed restrictions

$$\int_{\hat{y}_{p1}}^{\overline{y}} v\left(w_f^*(y) - r\right) f(y|e_H)\, \mathrm{dy} - \lambda \int_{\underline{y}}^{\hat{y}_{pf}} v(r) f(y|e_H) \mathrm{dy} - c \ = \overline{U}. \qquad (A39)$$

The transition from losses to gains is given by the $\hat{y}_f$ that satisfies $(A39)$. The existence of $\hat{y}_{pf}$ is guaranteed by the fact that $\int_{\hat{y}_{pf}}^{\overline{y}} v\left(w_f^*(y) - r\right) f(y|e_H)\, \mathrm{d}y$ is positive, while $-\lambda \int_{\underline{y}}^{\hat{y}_{pf}} v(r) f(y|e_H) \mathrm{d}y$ is negative. The limit of those integrals, $\hat{y}_{pf}$, can be adjusted to obtain a positive expression in the left-hand side of $(A39)$ equal to $\overline{U} \geq 0$. That both $w_f^*(y)$ satisfying Eq. (A35) and the lottery payment $L$ make the participation constraint of the agent bind, guarantee that such $\hat{y}_{pf}$ exists.

Finally, Theorem 1 shows that $w_s^*(y) = 0$ in $y < \hat{y}_{pf}$. Therefore, the optimal contract is given by $w_f^*(y) = \begin{cases} 0 & if\ y < \hat{y}_{pf}, \\ r + f'\left(\frac{1}{\mu}\right) & if\ y \geq \hat{y}_{pf}. \end{cases}$ This proves the first part of the corollary.

**Corollary 5.**

The agent makes, at most, two choices: accepting the contract, or not, and choosing an effort level. These choices generate three candidates for reference point: rejecting the contract and obtaining $\overline{U} \geq 0$, obtaining $\min\{w_s^*(y)\}$ after choosing $e_H$, and obtaining $\min\{w_s^*(y)\}$ after choosing $e_L$.

Denote the second-best optimal contract by $w_s^*(y)$. Notice that $\min\{w_s^*(y)\} = 0$. Also, note that an optimal contract generating utility $\overline{U}$ and that is compatible with the agent's preferences is $w_f^*(y)$ from Corollary 4 (i). Since $c > 0$ and $\overline{U} \geq 0$, it must be that $\mathbb{E}\left(w_f^*(y)\right) > 0$. Consequently, under the max-min rule, $r = w_f^*(y)$.

Corollary 4 presented the solution to the principal's problem when the agent exhibits prospect theory preferences with an exogenous $r$. Since the agent's preferences are still characterized by prospect theory, the solution presented therein remains optimal once $r = w_f^*(y)$ is accounted for.

To that end, I first define the output level after which the agent is awarded the bonus. Let $\hat{y}_u$ be the output level satisfying:

$$\frac{1}{\frac{\lambda v\left(w_f^*(y)\right)}{w_f^*(y)}} = \mu + \gamma\left(1 - \frac{f(\hat{y}_u|e_L)}{f(\hat{y}_u|e_H)}\right). \qquad (A40)$$

The above condition is given by Eq. $(A37)$ when $r = w_f^*(y)$. The existence and uniqueness of that output level is guaranteed by $\frac{1}{\frac{\lambda v\left(w_f^*(y)\right)}{w_f^*(y)}} > 0$, $\frac{1}{\frac{\lambda v\left(w_f^*(y)\right)}{w_f^*(y)}}$ being constant in performance in the domain of losses, and $\frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right) \leq 0$ (Assumption 2).

According to Corollary 4 if $y \geq \hat{y}_u$, the agent must be transitioned to the domain of gains with a contract satisfying Eq. (A34). Adapting that incentive schedule to account for the considered reference point rule gives:

$$\frac{1}{v'\left(w_s^*(y) - w_f^*(y)\right)} = \mu + \gamma\left(1 - \frac{f(y|e_H)}{f(y|e_L)}\right). \qquad (A41)$$

Also, Corollary 4 states that $w_s^*(y) = 0$ if $y < \hat{y}_u$. The reference point rule does not affect that level of payment. Therefore, the optimal contract is given by $w_s^*(y) =$

$$\begin{cases} 0 & if\ y < \hat{y}_u, \\ w_f^*(y) + f'\left(\frac{1}{\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right)}\right) & if\ y \geq \hat{y}_u. \end{cases} \blacksquare$$

To proof Proposition 1, a preliminary result is proved next. Under certain conditions on the agent's incentive scheme, goals motivate high effort at the cost of generating disutility.

**Lemma 2.** *Under A1-A4, and that the agent's preferences are characterized by Eq.* (4)*, then higher goals:*

i. *generate disutility if* $w'(g) > 0$*;*
ii. *motivate high effort if* $w'(g) > 0$*,* $w'(y) > 0$ *for* $y > g$ *, and* $w'(y) \leq 0$ *for* $y \leq g$.

*Proof.* Compute the derivative of Eq. (4), with respect to $g$ to obtain

$$\frac{dU(e, w(y), g)}{dg} = -\int_g^{\overline{y}} w'(g)v'\big(w(y) - w(g)\big)f(y|e)dy - \lambda\int_{\underline{y}}^{g} w'(g)v'\big(w(g) - w(y)\big)f(y|e)\mathrm{d}y. \qquad (A42)$$

Since $v' > 0$ and $\lambda > 1$, Eq. $(A42)$ is negative if $w'(g) > 0$. In that case, higher goals induce disutility. This proves part (i) of the lemma.

Integration by parts applied to Eq. (4) gives:

$$U(e, w(y), g) = v\big(w(\overline{y}) - w(g)\big) - \int_g^{\overline{y}} w'(y)v'\big(w(y) - w(g)\big)F(y|e_H)\mathrm{dy} - \lambda\int_{\underline{y}}^{g} w'(y)u'\big(w(g) - w(y)\big)\,F(y|ee_H)dy. \qquad (A43)$$

The agent's incentive compatibility constraint can be rewritten using Eq.$(A43)$ as

$$-\int_{g}^{\overline{y}} w'(y)v'\big(w(y)-w(g)\big)\big(F(y|e_H)-F(y|e_L)\big)\mathrm{dy} \\ -\lambda\int_{\underline{y}}^{g} w'(y)v'\big(w(g)-w(y)\big)\big(F(y|e_H)-F(y|e_L)\big)\mathrm{dy} \geq c. \quad (A44)$$

To investigate whether higher goals incentivize higher effort, derive $(A44)$ with respect to $g$ to obtain:

$$-(\lambda-1)w'(g)v'\big(w(g)-w(g)\big)\big(F(g|e_H)-F(g|e_L)\big) \\ +\int_{g}^{\overline{y}} w'(y)w'(g)v''\big(w(y)-w(g)\big)\big(F(y|e_H)-F(y|e_L)\big)\mathrm{dy} \\ -\lambda\int_{\underline{y}}^{g} w'(y)w'(g)v''\big(w(g)-w(y)\big)\big(F(y|e_H)-F(y|e_L)\big)\mathrm{dy}. \quad (A45)$$

Since $v'>0$, $v''<0$ and $F(y|e_L)\geq F(y|e_H)$, a consequence of Assumption 2, Eq.$(A45)$ shows that higher goals generate higher effort if $w'(g)>0$, $w'(y)>0$ in $y\geq g$ and $w'(y)=0$ in $y<g$.

**Proposition 1.**

*Part i).* Let $\eta=1$, $\phi=0$, $u'=1$, and $r=w(g)$. Under these restrictions, $U(e_H, w(\tilde{y}), w(g))$ is S-shaped for any realization $\tilde{y}\in\left[\underline{y},\overline{y}\right]$ since in the domain of losses $-\frac{v''\left(u(w(g))-u(w(\tilde{y}))\right)}{v'\left(u(w(g))-u(w(\tilde{y}))\right)}\geq 0$, corroborating the condition in Eq. $(A3)$. Therefore, the solution from the first-order condition is sufficient and necessary for the domain of gains only. That condition, given by Eq. $(A6)$, rewrites under the considered restrictions as:

$$\frac{1}{v'(w_S^*(y)-w_S^*(g))}=\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right). \quad \text{(A46)}$$

Rearranging (A46) gives

$$w_S^*(y)=w_S^*(g)+f'\left(\frac{1}{\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right)}\right). \quad \text{(A47)}$$

Eq. $(A47)$ shows that implementing the requirement $w'(g)>0$ from Lemma 2, implies $\frac{dw_S^*(y)}{dg}>0$.

Moreover, the derivative of Eq. (A46) with respect to $y$ gives $\frac{dw_S^F(y)}{dy} = \frac{\gamma \frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right)\left(v'\left(w_S^F(y)-w_S^F(g)\right)\right)^2}{v''\left(w_S^F(y)-w_S^F(g)\right)} \geq 0$. Therefore, in the domain of gains the second-best optimal contract increases in performance.

To study the location of the bonus, Eq. $(A14)$ is rewritten to account for the considered restrictions. I obtain the following condition

$$\frac{1}{\frac{\lambda v(w_S^*(g))}{w_S^*(g)}} = \nu + \gamma\left(1 - \frac{f(\hat{y}_g|e_L)}{f(\hat{y}_g|e_H)}\right). \qquad (A48)$$

Eq. $(A48)$ shows that the output level $\hat{y}_g$ that satisfies $(A48)$ transitions the agent from losses to gains. The existence and uniqueness of that output level is guaranteed by $\frac{1}{\frac{\lambda v(w_S^*(g))}{w_S^*(g)}} > 0$, $\frac{d}{dy}\left(\frac{1}{\frac{\lambda v(w_S^*(g))}{w_S^*(g)}}\right) = 0$, and Assumption 2.

Theorem 1 also demonstrates that for an S-shaped utility function, $w_S^*(y) = 0$ in $y < \hat{y}_g$ is second-best optimal. Accordingly, the second-best contract is

$$w_S^*(y) = \begin{cases} 0 & \text{if } y < \hat{y}_g, \\ w_S^*(g) + f'\left(\frac{1}{\mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right)}\right) & \text{if } y \geq \hat{y}_g. \end{cases}$$

This solution that exhibits a discrete jump at $y = \hat{y}_g$ as $\lim_{y \to \hat{y}_g^+} w_S^*(y) = w_S^*(g)$ and $\lim_{y \to \hat{y}_g^-} w_S^*(y) = 0$. In addition, a direct consequence of the condition $w_S^{*\prime}(g) > 0$ from Lemma 2, necessary for goals to be motivating, is that the needs to become larger the larger $g$ is.

Finally, I demonstrate that $\hat{y}_g = g$. Proceed by contradiction by assuming that $\hat{y}_g < g$. In that case, the principal overinsures the agent in $y \in [\hat{y}_g, g]$, a segment where he is risk seeking due to $U(e_H, w(y), r)$ being S-shaped. The principal could increase profits by setting $w_S^*(y) = 0$. In addition, due to the convexity of $U(e_H, w(\tilde{y}), r)$ in losses the agent would be willing to accept that contract. Now suppose $\hat{y}_g > g$. The agent is overexposed to risk in $y \in [g, \hat{y}_g]$, where he is risk averse. This leads the agent to reject the contract. To ensure that the contract is accepted, the principal offers $w_S^*(y)$ given by $(A47)$ for any $y \geq g$. Then it must be that $\hat{y}_g = g$.

The result $\hat{y}_g = g$ has important implications. First, since $\lim_{y \to g^+} w_S^*(y) = w_S^*(g)$ and $\lim_{y \to g^-} w_S^*(y) = 0$, a bonus of size $w_S^*(g)$ is given at $y = g$. Second, to fulfill $w_S^{*\prime}(g) > 0$ from

Lemma 2, that bonus increases with the size of $g$. This is consistent with the result of Corollary 2. Lastly, the conditions $w_s^{*\prime}(y) > 0$ if $y \geq g$ and $w_s^{*\prime}(y) = 0$ if $y < g$ from Lemma 2 are met. As a result, it is second-best optimal to offer

$$w_s^*(y) = \begin{cases} 0 & if\ y < g, \\ w_s^*(g) + f'\left(\dfrac{1}{\mu + \gamma\left(1 - \dfrac{f(y|e_L)}{f(y|e_H)}\right)}\right) & if\ y \geq g. \end{cases}$$

*Part ii).* Under $\eta = 1$, $\phi = 0$, $u' = 1$, and $r = w(g)$ the derivative of Eq. $(A4)$ with respect to $g$ gives:

$$\begin{aligned} &\mu\left(-w'(g)\theta_{\mathbb{I}}v'\big(w(y) - w(g)\big)f(y|e) - \lambda(1-\theta_{\mathbb{I}})w'(g)v'\left(w(g) - w(y)\right)f(y|e)\right) \\ &\quad + \gamma\left(-w'(g)\theta_{\mathbb{I}}v'\big(w(y) - w(g)\big)\big(f(y|e_H) - f(y|e_L)\big)\right. \\ &\quad \left. - \lambda w'(g)(1-\theta_{\mathbb{I}})v'\big(w(g) - w(y)\big)\big(f(y|e_H) - f(y|e_L)\big)\right) = 0. \qquad (A49) \end{aligned}$$

Denoting by $g^*$ the goal level that satisfies the condition $(A49)$, the following conditions are obtained:

$$-w'(g)\left(\mu v'\big(w(y) - w(g^*)\big)f(y|e) + \gamma u'\big(w(y) - w(g^*)\big)\big(f(y|e_H) - f(y|e_L)\big)\right) = 0, \quad (A50)$$

if $\theta_{\mathbb{I}} = 1$, and

$$-\lambda w'(g)\left(\mu v'\big(w(g^*) - w(y)\big)f(y|e) + \gamma v'\big(w(g^*) - w(y)\big)\big(f(y|e_H) - f(y|e_L)\big)\right) = 0. \quad (A51)$$

if $\theta_{\mathbb{I}} = 0$. Eqs.$(A50)$ and $(A51)$ imply:

$$\mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right) = 0. \qquad (A52)$$

Eqs. $(A46)$ and $(A52)$ are used to obtain

$$v'(w_s^*(y) - w_s^*(g^*)) = +\infty. \qquad (A53)$$

Since $\lim_{y \to g^+} w_s^*(y) = w_s^*(g)$ then $\lim_{y \to g^+} v'\big(w_s^*(y) - w_s^*(g^*)\big) = +\infty$, a direct consequence of the Inada condition, $\lim_{x \to 0} v'(x) = +\infty$. However, the realization of $y$ is ex-ante unknown so the principal cannot set a goal that is just met. Instead, she can set a goal such that $(A53)$ holds on expectation. Using the fact that $v'(\cdot)$ is decreasing and convex

$$\mathbb{E}\big(v'(w_s^*(y) - w_s^*(g)\big) \geq v'(\mathbb{E}(w_s^*(y)) - w_s^*(g)). \quad (A54)$$

So $g^*$ is set such that $\mathbb{E}(w_s^*(y)) - w_s^*(g^*) = \epsilon$ for arbitrarily small $\epsilon > 0$. This gives $v'\left(\mathbb{E}(w_s^*(y)) - w_s^*(g^*)\right) = +\infty$. The goal $g^*$ is set so that is, on expectation, attainable.■

**Corollary 6.**
Let $\phi = 1$, $u' = 1$, and $r = \bar{\bar{w}}$. The utility $U(e_H, w(\tilde{y}), r)$ is S-shaped for any $\tilde{y} \in [\underline{y}, \overline{y}]$ since $-\frac{v''\left(u(\bar{\bar{w}})-u(w(\tilde{y}))\right)}{v'\left(u(\bar{\bar{w}})-u(w(\tilde{y}))\right)} \geq 0$ in the domain of losses, corroborating Eq. $(A3)$. Therefore, the solution from the first-order condition is necessary and sufficient only for the domain of gains.

That condition, given in Eq. $(A6)$, becomes under the assumed restrictions equal to

$$\frac{1}{\left(1 + \eta v'\left(w_s^*(y) - \bar{\bar{w}}\right)\right)} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right). \qquad (A55)$$

Algebraic manipulations of the above equation give $w_s^*(y) = \bar{\bar{w}} + f'\left(\frac{1}{\eta}\left(\frac{1}{\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right)} - 1\right)\right)$. This is the optimal solution in the domain of gains. Moreover, the derivative of Eq. $(A55)$ with respect to $y$ gives

$$\frac{dw_s^*(y)}{dy} = \frac{\gamma\frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right)\left(\left(1+\eta v'\left(w_s^*(y)-\bar{\bar{w}}\right)\right)\right)^2}{\left(\left(1+\eta v''\left(w_s^*(y)-\bar{\bar{w}}\right)\right)\right)}. \qquad (A56)$$

Since $\frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right) \leq 0$ and $v'' < 0$, then $\frac{dw_s^*(y)}{dy} \geq 0$. The contract is everywhere increasing in the domain of gains.

The transition from losses to gains is given by Eq. $(A14)$, which becomes under the assumed restrictions equal to:

$$\frac{1}{\frac{1+\lambda v(\bar{\bar{w}})}{\bar{\bar{w}}}} = \mu + \gamma\left(1 - \frac{f(\hat{y}_{ms}|e_L)}{f(\hat{y}_{ms}|e_H)}\right). \qquad (A57)$$

The bonus is given when output is larger than the output level $\hat{y}_{ms}$ satisfying Eq. $(A57)$. According to Theorem 1, it is second-best optimal to set $w_s^*(y) = 0$ in $y < \hat{y}_{ms}$. Hence, the optimal contract is $w_s^*(y) = \begin{cases} 0 & if\ y < \hat{y}_{ms}, \\ \bar{\bar{w}} + f'\left(\frac{1}{\eta}\left(\frac{1}{\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right)} - 1\right)\right) & if\ y \geq \hat{y}_{ms}. \end{cases}$ This proves the first part of the corollary.

We turn to study the first-best contract. Hence, consider $\gamma = 0$ in addition to $\phi = 1, u' = 1$, and $r = \bar{\bar{w}}$. Eq. $(A55)$ becomes $\frac{1}{\left(1+\eta v'\left(w_f^*(y)-\bar{\bar{w}}\right)\right)} = \mu$, which after some manipulations yields $w_f^*(y) = \bar{\bar{w}} + f'\left(\frac{1}{\eta}\left(\frac{1}{\mu} - 1\right)\right)$. According to Eq. $(A56)$, that transfer exhibits $\frac{dw_f^*(y)}{dy} = 0$.

Also, Theorem 1 shows that in the domain of losses it is also first-best optimal to offer $w_f^*(y) = 0$. The transition from $w_f^*(y) = 0$ to $w_f^*(y) = \bar{\bar{w}} + f'\left(\frac{1}{\eta}\left(\frac{1}{\mu} - 1\right)\right)$ is given by the following equality, which adapts Eq. $(A23)$ to account for the considered restrictions,

$$\int_{\hat{y}_{mf}}^{\bar{y}} w_f^*(y) f(y|e_H)\,\mathrm{d}y + \eta \int_{\hat{y}_{mf}}^{\bar{y}} v\left(w_f^*(y) - \bar{\bar{w}}\right) f(y|e_H)\,\mathrm{d}y - \lambda \int_{\underline{y}}^{\hat{y}_{mf}} v(\bar{\bar{w}}) f(y|e_H)\mathrm{d}y - c = \bar{U}. \quad (A58)$$

Hence, the output $\hat{y}_{mf}$ that satisfies Eq.$(A58)$ provides that transition. Theorem 1 shows that this output level is unique and interior. Hence, the optimal contract is $w_f^*(y) = \begin{cases} 0 & if\ y < \hat{y}_{mf}, \\ \bar{\bar{w}} + f'\left(\frac{1}{\eta}\left(\frac{1}{\mu} - 1\right)\right) & if\ y \geq \hat{y}_{mf}. \end{cases}$ This proves the third part of the corollary.

Now let $\phi = 1$, $v' = 1$, and $r = \bar{\bar{w}}$. Under these restrictions $U(e_H, w(\tilde{y}), \bar{\bar{w}})$ is concave for any output realization $\tilde{y} \in \left[\underline{y}, \bar{y}\right]$ since $-\frac{u''(w(\tilde{y}))}{u'(w(\tilde{y}))} \geq 0$, a contradiction of the condition in Eq. $(A3)$. Hence, the solutions from the first-order conditions are necessary and sufficient to solve the maximization problem of the principal.

The first order conditions from Eqs. $(A6)$ and $(A7)$ become

$$\frac{1}{u'(w_s^*(y))(1+\eta)} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right), \qquad (A59)$$

and

$$\frac{1}{u'(w_s^*(y))(1+\eta\lambda)} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right), \qquad (A60)$$

respectively. Eq. $(A59)$ can be expressed as $w_s^*(y) = h'\left(\frac{1}{(1+\eta)\left(\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right)\right)}\right)$ and Eq. $(A60)$ as $w_s^F(y) = h'\left(\frac{1}{(1+\eta\lambda)\left(\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right)\right)}\right)$. These two solutions exhibit $\frac{dw_s^*(y)}{dy} > 0$, as shown by Eqs. $(A8)$ and $(A9)$.

The transition from $w_s^*(y)$ satisfying $(A59)$ to $w_s^F(y)$ satisfying $(A60)$ is given by the unique output level $\hat{y}_{ms} \in (\underline{y}, \overline{y})$ that satisfies:

$$\int_{\hat{y}_{ms}}^{\overline{y}} u(W_S^*(y)) f(y|e_H)\,\mathrm{d}y + \eta \int_{\hat{y}_{ms}}^{\overline{y}} \left(u(W_S^*(y)) - u(\overline{\overline{w}})\right) f(y|e_H)\,\mathrm{d}y + \int_{\underline{y}}^{\hat{y}_{ms}} u(W_S^*(y) f(y|e_H)\mathrm{d}y$$
$$- \lambda\eta \int_{\underline{y}}^{\hat{y}_{ms}} \left(u(\overline{\overline{w}}) - u\left(w_S^*(y)\right)\right) f(y|e_H)\mathrm{d}y - c \; = \overline{U}. \qquad (A61)$$

The existence of $\hat{y}_{ms}$ is guaranteed by the fact that the solutions from Eqs. $(A59)$ and $(A60)$ make the participation constraint bind for gains and losses, respectively. As a result, the optimal incentive scheme is given by:

$$w_f^*(y) = \begin{cases} h'\left(\dfrac{1}{(1+\eta)\left(\mu + \gamma\left(1 - \dfrac{f(y|e_L)}{f(y|e_H)}\right)\right)}\right) & if\ y \geq \hat{y}_{ms}\,, \\ h'\left(\dfrac{1}{(1+\eta\lambda)\left(\mu + \gamma\left(1 - \dfrac{f(y|e_L)}{f(y|e_H)}\right)\right)}\right) & if\ y < \hat{y}_{ms}. \end{cases} \qquad (A62)$$

This proves the second part of the corollary.
To conclude, we analyze the first-best contract. Hence, consider $\gamma = 0$ in addition to $\phi = 1, v' = 1$, and $r = \overline{\overline{w}}$. The first-order condition in $(A59)$ becomes $\frac{1}{u'\left(w_f^*(y)\right)(1+\eta)} = \mu$, which after algebraic manipulations gives $w_f^*(y) = h'\left(\frac{1}{(1+\eta)\mu}\right)$. Eq. $(A8)$ shows that $\frac{dw_f^*(y)}{dy} = 0$ under the considered restrictions. Finally, Theorem 1 shows that $w_f^*(y) = h'\left(\frac{1}{(1+\eta)\mu}\right)$ is first-best optimal when $U(e_H, w(y), r)$ is concave. ■

**Proposition 2.**

*i) Lagrangian and first-order conditions*

Denote by $\mu \geq 0$ and $\gamma \geq 0$ the Lagrangian multipliers of the agent's participation and incentive compatibility constraints. The Lagrangian of the principal's maximization problem writes as

$$
\begin{aligned}
\mathcal{L} = &\left(S(y) - w(y)\right)f(y|e_H) \\
&+ \mu\left(u(w(y))f(y|e_H) + \eta\theta_{\mathbb{I}}\int_{\underline{y}}^{\overline{y}} v\left(u(w(y) - u(w(\tilde{\tilde{y}}))f(y|e_H)f(\tilde{\tilde{y}}|e_H)\mathrm{d}\tilde{\tilde{y}}\right. \\
&\left. - \eta\lambda(1-\theta_{\mathbb{I}})\int_{\underline{y}}^{\overline{y}} v\left(u(w(\tilde{\tilde{y}})) - (u(w(y))\right)f(y|e_H)f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}} - c\right) \\
&+ \gamma\left((u(w(y))\left(f(y|e_H) - f(y|e_L)\right)\right. \\
&+ \theta_{\mathbb{I}}\eta\int_{\underline{y}}^{\overline{y}} v\left((u(w(y)) - u(w(\tilde{\tilde{y}}))\right)\left(f(y|e_H) - f(y|e_L)\right)f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}} \\
&\left. - \lambda\eta(1-\theta_{\mathbb{I}})\int_{\underline{y}}^{\overline{y}} v\left(u(w(\tilde{\tilde{y}}) - (u(w(y))\right)\left(f(y|e_H) - f(y|e_L)\right)f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}} - c\right).
\end{aligned}
\tag{A63}
$$

Pointwise optimization with respect to $w(y)$ gives

$$
\begin{aligned}
-f(y|e_H) + \mu&\left(u'(w(y))f(y|e_H) + \eta\theta_{\mathbb{I}}\int_{\underline{y}}^{\overline{y}} v'(u(w(y) - u(w(\tilde{\tilde{y}}))u'\,(w(y)f(y|e_H)f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}\right. \\
&\left. + \eta\lambda(1-\theta_{\mathbb{I}})\int_{\underline{y}}^{\overline{y}} v'\left(u\left(w(\tilde{\tilde{y}})\right) - (u(w(y))\right)u'\,(w(y)f(y|e_H)f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}\right) \\
&+ \gamma\left((u'(w(y))\left(f(y|e_H) - f(y|e_L)\right)\right. \\
&+ \theta_{\mathbb{I}}\eta\int_{\underline{y}}^{\overline{y}} v'\left((u(w(y)) - u(w(\tilde{\tilde{y}}))\right)u'(w(y)\left(f(y|e_H) - f(y|e_L)\right)f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}} \\
&\left. + \lambda\eta(1-\theta_{\mathbb{I}})\int_{\underline{y}}^{\overline{y}} v'\left(u(w(\tilde{\tilde{y}}) - (u(w(y))\right)\left(f(y|e_H) - f(y|e_L)\right)u'\left(w(y)\right)f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}\right) = 0.
\end{aligned}
\tag{A64}
$$

Denoting by $w_s^F(y)$ the transfer satisfying $(A64)$, the following expressions are obtained after algebraic manipulations:

$$
\frac{1}{u'\left(w_s^F(y)\right) + \eta\int_{\underline{y}}^{\overline{y}} v'\left(u(w_s^F(y) - u(w_s^F(\tilde{\tilde{y}}))u'\,(w_s^F(y))f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right), \tag{A65}
$$

if $\theta_{\mathbb{I}} = 1$, and

$$
\frac{1}{u'\left(w_s^F(y)\right) + \eta\lambda\int_{\underline{y}}^{\overline{y}} v'\left(u(w_s^F(\tilde{\tilde{y}}) - (u(w_s^F(y))\right)u'\,(w_s^F(y))f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right), \tag{A66}
$$

if $\theta_{\mathbb{I}} = 0$. Due to $u'' < 0$, $v'' < 0$ and $\frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right) \leq 0$ (Assumption 2), the derivative of $(A65)$ with respect to $y$ exhibits $\frac{dw_s^F(y)}{dy} \geq 0$. Similarly, the derivative of $(A66)$ with respect to y exhibits

$\frac{dw_s^F(y)}{dy} \geq 0$ if $U\left(e_H, w(\tilde{y}), w(\tilde{\tilde{y}})\right)$ is concave for any realization $\tilde{y} \in [\underline{y}, \overline{y}]$ but becomes $\frac{dw_s^F(y)}{dy} \leq 0$ if $U\left(e_H, w(\tilde{y}), w(\tilde{\tilde{y}})\right)$ is S-shaped.

*ii) Solution when $U(e_H, w(y), r)$ is S-shaped*

Let $U\left(e_H, w(\tilde{y}), w(\tilde{\tilde{y}})\right)$ be S-shaped for any output realization $\tilde{y} \in [\underline{y}, \overline{y}]$. If $w_s^F(y)$ satisfying Eq. $(A66)$ exhibits $0 < w_s^F(y) < w_s^F(\tilde{\tilde{y}})$, the principal is better off offering $L_s := (p: w_s^F(\tilde{\tilde{y}}), 1-p: 0)$ for given $\tilde{\tilde{y}}$ and $p \in [0,1]$. Since the agent's utility increases in $w_s^F(y)$, there must exist a $p \in [0,1]$ such that:

$$\begin{aligned}
&\mathbb{E}_y\left(u(w_s^F(y))\right) - \eta\lambda \int_{\underline{y}}^{\tilde{y}} \int_{\underline{y}}^{\overline{y}} v\left(u\left(w_s^F(\tilde{\tilde{y}})\right) - u(w_s^F(y))\right) f(\tilde{\tilde{y}}|e)\, \mathrm{d}\tilde{\tilde{y}} f(y|e) \mathrm{d}y \\
&\qquad + \eta \int_{\tilde{y}}^{\overline{y}} \int_{\underline{y}}^{\overline{y}} v\left(u(w_s^F(y)) - u\left(w_s^F(\tilde{\tilde{y}})\right)\right) f(\tilde{\tilde{y}}|e)\, \mathrm{d}\tilde{\tilde{y}} f(y|e) \mathrm{d}y \\
&\qquad = p\, \mathbb{E}_y\left(u\left(w_s^F(\tilde{\tilde{y}})\right)\right) - p(1-p)\eta\lambda\left(\int_{\underline{y}}^{\overline{y}} v\left(-u\left(w_s^F(\tilde{\tilde{y}})\right)\right) f(\tilde{\tilde{y}}|e) \mathrm{d}\tilde{\tilde{y}}\right) \\
&\qquad + p(1-p)\eta\left(\int_{\underline{y}}^{\overline{y}} v\left(u\left(w_s^F(\tilde{\tilde{y}})\right)\right) f(\tilde{\tilde{y}}|e) \mathrm{d}\tilde{\tilde{y}}\right). \qquad (A67)
\end{aligned}$$

Hence, replacing $w_s^F(y)$ from Eq.$(A66)$ by $L_s$ leaves the agent's participation incentive compatibility constraints unchanged. Eq. $(A67)$ and the convexity of $v$ imply:

$$\begin{aligned}
&\mathbb{E}_y\left(u(w_s^F(y))\right) - p\, \mathbb{E}_y\left(u\left(w_s^F(\tilde{\tilde{y}})\right)\right) \\
&\qquad \leq \eta\lambda \int_{\underline{y}}^{\tilde{y}} \int_{\underline{y}}^{\overline{y}} v\left(u\left(w_s^F(\tilde{\tilde{y}})\right) - u(w_s^F(y))\right) f(\tilde{\tilde{y}}|e)\, \mathrm{d}\tilde{\tilde{y}} f(y|e) \mathrm{d}y \\
&\qquad - \eta \int_{\tilde{y}}^{\overline{y}} \int_{\underline{y}}^{\overline{y}} v\left(u(w_s^F(y)) - u\left(w_s^F(\tilde{\tilde{y}})\right)\right) f(\tilde{\tilde{y}}|e)\, \mathrm{d}\tilde{\tilde{y}} f(y|e) \mathrm{d}y \\
&\qquad - p\eta\lambda\left(\int_{\underline{y}}^{\overline{y}} v\left(-(1-p)u\left(w_s^F(\tilde{\tilde{y}})\right)\right) f(\tilde{\tilde{y}}|e) \mathrm{d}\tilde{\tilde{y}}\right) \\
&\qquad + p\eta\left(\int_{\underline{y}}^{\overline{y}} v\left((1-p)u\left(w_s^F(\tilde{\tilde{y}})\right)\right) f(\tilde{\tilde{y}}|e) \mathrm{d}\tilde{\tilde{y}}\right)
\end{aligned}$$

$(A68)$

Since $\mathbb{E}_y\left(u(w_s^F(y))\right) - p\, \mathbb{E}_y\left(u(w_s^F(y))\right) \geq 0$ and $v' > 0$ the inequality in Eq. $(A68)$ implies

$$w_s^F(y) > p w_s^F(\tilde{\tilde{y}}). \qquad (A69)$$

Thus, $L_s$ is more cost-effective for the principal than the solution implied by $(A66)$.

Next, I investigate the marginal incentives from offering $L_s$. Denote by $\overline{L_s}$ its expected value and substitute it in the agent's expected utility to obtain:

$$U\left(e_H, L, w_s^F(\tilde{\tilde{y}})\right) = \left(\frac{\overline{L}_s}{w_s^F(\tilde{\tilde{y}})}\right) u\left(w_s^F(\tilde{\tilde{y}})\right) + \frac{\overline{L}_s}{w_s^F(\tilde{\tilde{y}})}\left(1 - \frac{\overline{L}_s}{w_s^F(\tilde{\tilde{y}})}\right)\eta\left(v\left(u\left(w_s^F(\tilde{\tilde{y}})\right)\right) - \lambda v\left(-u\left(w_s^F(\tilde{\tilde{y}})\right)\right)\right).$$

$$(A70)$$

The above expression is not linear in $\overline{L}_s$. Hence, changes in $\overline{L}_s$, via changes in $p$ affect the agent's marginal utility. So unlike Theorem 1 there is an interior probability $p_s^*(y)$ that maximizes the agent's utility. The first-order condition of the utility in Eq. $(A70)$ with respect to $p$ is:

$$u\left(w_s^F(\tilde{\tilde{y}})\right) + (1-2p)\eta\left(v\left(u\left(w_s^F(\tilde{\tilde{y}})\right)\right) - \lambda v\left(-u\left(w_s^F(\tilde{\tilde{y}})\right)\right)\right) = 0. \qquad (A71)$$

The second derivative of Eq. $(A70)$ with respect to $p$ is $-2\eta\left(v\left(u\left(w_s^F(\tilde{\tilde{y}})\right)\right) - \lambda v\left(-u(w_s^F(\tilde{\tilde{y}}))\right)\right) < 0$. Therefore, the optimal probability is interior and has closed-form solution $p_s^*(\tilde{\tilde{y}}) = \frac{1}{2} + \frac{u\left(w_s^F(\tilde{\tilde{y}})\right)}{2\eta\left(v\left(u\left(w_s^F(\tilde{\tilde{y}})\right)\right) - \lambda v\left(-u\left(w_s^F(\tilde{\tilde{y}})\right)\right)\right)}$. The principal pays lottery $L_s^* = (p_s^*(\tilde{\tilde{y}}): w_s^F(\tilde{\tilde{y}}), 1 - p_s^*(\tilde{\tilde{y}}): 0)$.

It is next shown that $\hat{y}_{ds} = \overline{y}$ . If there were a unique threshold output below which $L_s^*$ is paid and above which $w_s^F(y)$ satisfying Eq. $(A65)$ is paid, it must satisfy:

$$\frac{1}{\frac{u\left(w_s^F(\tilde{\tilde{y}})\right)}{w_s^F(\tilde{\tilde{y}})} + \frac{1}{w_s^F(\tilde{\tilde{y}})}\left(1 - \frac{2\overline{L}}{w_s^F(\tilde{\tilde{y}})}\right)\eta\left(v\left(u\left(w_s^F(\tilde{\tilde{y}})\right)\right) - \lambda v\left(-u\left(w_s^F(\tilde{\tilde{y}})\right)\right)\right)} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right)$$

$$(A72)$$

Since $p_s^*(\tilde{\tilde{y}})$ satisfies Eq. $(A71)$ the numerator of the left-hand side of Eq. $(A72)$ is zero. The solution to Eq. $(A72)$ is at the boundary of the output space. Due to Assumption 2, the solution is $\hat{y}_{ds} = \overline{y}$

*iii) Properties of the optimal contract.*

Let $U\left(e_H, w(\tilde{y}), w_s^F(\tilde{\tilde{y}})\right)$ be S-shaped for any $\tilde{y} \in [\underline{y}, \overline{y}]$. Denote by $b_s > 0$ the pay level $w_s^F(y)$ satisfying Eq. $(A65)$ evaluated at $y = \overline{y}$. It is second-best optimal to offer $L_s^* = (p_s^*(\overline{y}): b_s, 1 - p_s^*(\overline{y}): 0)$ if $y < \overline{y}$ and $b_s$ paid at $y = \overline{y}$ proving the first part of the Proposition.

Let $U\left(e_H, w(\tilde{y}), w(\tilde{\tilde{y}})\right)$ be concave for any $\tilde{y} \in [\underline{y}, \overline{y}]$. The optimal contract, $w_s^*(y)$, consists of two components: $w_s^F(y)$ satisfying Eq. $(A65)$, which implies $w_s^F(y) \geq w_s^F(\tilde{\tilde{y}})$, and $w_s^F(y)$

satisfying Eq. $(A66)$, entailing $w_s^F(y) < w_s^F(\tilde{\tilde{y}})$. The optimal contract combines these two first-order conditions. A contract paying only $w_s^F(y)$ satisfying Eq. $(A66)$ is rejected as it generates losses. Also, a contract $w_s^F(y)$ satisfying Eq. $(A65)$ only, is dominated by a combination of these two conditions.

The transition from losses to gains is defined next. Let $\hat{y}_{ds} \in [\underline{y}, \overline{y}]$ be the output level satisfying:

$$\int_{\underline{y}}^{\hat{y}_{ds}} u(w_s^F(y))f(y|e)dy + \int_{\hat{y}_{ds}}^{\overline{y}} u(w_s^F(y))f(y|e)dy + \eta \int_{\hat{y}_{ds}}^{\overline{y}} \int_{\underline{y}}^{\overline{y}} v\left(u(w_s^F(y)) - u\left(w_s^F(\tilde{\tilde{y}})\right)\right) f(\tilde{\tilde{y}}|e)f(y|e)d\tilde{\tilde{y}}dy$$
$$- \eta\lambda \int_{\underline{y}}^{\hat{y}_{ds}} \int_{\underline{y}}^{\overline{y}} v\left(u\left(w_s^F(\tilde{\tilde{y}})\right) - u(w_s^F(y))\right) f(y|e)\, f(\tilde{\tilde{y}}|e)\, d\tilde{\tilde{y}}dy - c(e) = \overline{U} \qquad (A73)$$

The existence of $\hat{y}_{ds}$ is guaranteed because the left-hand side of Eq. $(A73)$ can be negative when $\hat{y}_{ds} = \overline{y}$, increases as $\hat{y}_{ds}$ increases and becomes positive when $\hat{y}_{ds} = \underline{y}$.

Hence, a candidate for optimal contract is given by

$$w_s^*(y) = \begin{cases} w_s^F(y) & \text{satisfying } (A65) & \text{if } y \geq \hat{y}_{ds}, \\ w_s^F(y) & \text{satisfying } (A66) & \text{if } y < \hat{y}_{ds}. \end{cases} \qquad (A74)$$

That solution exhibits a discrete jump at $y = \hat{y}_{ds}$ since $\lambda > 1$ appears in the denominator of the right-hand side of $(A66)$ but this coefficient does not enter in $(A65)$. Masatlioglu and Raymond (2016) show that if $-1 > (1-\lambda)v'(0)$, the agent does not respect first-order stochastic dominance. In that case, the Eq. $(A74)$ is dominated by a contract paying $L_s = (p: w_s^*(y), 1 - p: 0)$ for any $p$. This proves the second part of the Theorem.

*iv) Optimal first-best contract*

Consider now $\gamma = 0$ on top of $\phi = 1$. Denote by $w_f^F(y)$ the candidate solution from the first-order approach under these restrictions. Eq. $(A65)$ collapses to

$$\frac{1}{u'(w(y)) + \eta \int_{\underline{y}}^{\overline{y}} v'(u(w(y) - u(w(\tilde{\tilde{y}}))u'\,(w(y)f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}} = \mu, \qquad (A75)$$

if $\theta_{\mathbb{I}} = 1$, and Eq. $(A66)$ collapses to

$$\frac{1}{u'(w(y)) + \eta\lambda \int_{\underline{y}}^{\overline{y}} v'\left(u(w(\tilde{\tilde{y}}) - (u(w(y))\right) u'\,(w(y)f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}} = \mu, \qquad (A76)$$

if $\theta_{\mathbb{I}} = 0$. Eqs. $(A75)$ and $(A76)$ show that $\frac{dw_f^F(y)}{dy} = 0$ under $\gamma = 0$. Hence, $w_f^F(y)$ is performance insensitive.

As in the derivation of the second-best contract, it can be shown that if $U\left(e_H, w(\tilde{y}), w(\tilde{\tilde{y}})\right)$ is S-shaped for any $\tilde{y}$, the principal is better off paying a lottery $L_f = (p\colon w_f^F(\tilde{\tilde{y}}), 1-p\colon 0)$ instead of $w_f^F(y)$ satisfying Eq. $(A76)$. That lottery, $L$, can be offered to the agent with a $p \in (0,1)$ that satisfies:

$$\begin{aligned}
\mathbb{E}_y\left(u\left(w_f^F(y)\right)\right) - \eta\lambda \int_{\underline{y}}^{\tilde{\tilde{y}}} \int_{\underline{y}}^{\overline{y}} v\left(u\left(w_f^F(\tilde{\tilde{y}})\right) - u\left(w_f^F(y)\right)\right) f(\tilde{\tilde{y}}|e)\, \mathrm{d}\tilde{\tilde{y}} f(y|e)\mathrm{d}y \\
+ \eta \int_{\tilde{\tilde{y}}}^{\overline{y}} \int_{\underline{y}}^{\overline{y}} v\left(u\left(w_f^F(y)\right) - u\left(w_f^F(\tilde{\tilde{y}})\right)\right) f(\tilde{y}|e)\, \mathrm{d}\tilde{\tilde{y}} f(y|e)\mathrm{d}y \\
= p\, \mathbb{E}_y\left(u\left(w_f^F(\tilde{\tilde{y}})\right)\right) - p(1-p)\eta\lambda\left(\int_{\underline{y}}^{\overline{y}} v\left(-u\left(w_f^F(\tilde{\tilde{y}})\right)\right) \mathrm{d}y f(y|e)\mathrm{d}y\right) \\
+ p(1-p)\eta\left(\int_{\underline{y}}^{\overline{y}} v\left(u\left(w_f^F(\tilde{\tilde{y}})\right)\right) f(y|e)\mathrm{d}y\right). \qquad (A77)
\end{aligned}$$

Therefore, replacing $w_f^F(y)$ from Eq. $(A76)$ by $L_f$ leaves the agent's participation constraint unchanged. Eq. $(A77)$ and the convexity of $U\left(e_H, w_f^F(\tilde{y}), w_f^F(\tilde{\tilde{y}})\right)$ for $\tilde{\tilde{y}}$ such that $w_f^F(\tilde{\tilde{y}}) < w_f^F(y)$, imply:

$$\begin{aligned}
\mathbb{E}_y\left(u\left(w_f^F(y)\right)\right) - p\, \mathbb{E}_y\left(u\left(w_f^F(\tilde{\tilde{y}})\right)\right) \\
\leq \eta\lambda \int_{\underline{y}}^{\tilde{\tilde{y}}} \int_{\underline{y}}^{\overline{y}} v\left(u\left(w_f^F(\tilde{\tilde{y}})\right) - u\left(w_f^F(y)\right)\right) f(\tilde{\tilde{y}}|e)\, \mathrm{d}\tilde{\tilde{y}} f(y|e)\mathrm{d}yy \\
- \eta \int_{\tilde{\tilde{y}}}^{\overline{y}} \int_{\underline{y}}^{\overline{y}} v\left(u\left(w_f^F(y)\right) - u\left(w_f^F(\tilde{\tilde{y}})\right)\right) f(\tilde{y}|e)\, \mathrm{d}\tilde{\tilde{y}} f(y|e)\mathrm{d}y \\
- p\eta\lambda\left(\int_{\underline{y}}^{\overline{y}} v\left(-(1-p)u\left(w_f^F(\tilde{\tilde{y}})\right)\right) \mathrm{d}y f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}\right) \\
+ p\eta\left(\int_{\underline{y}}^{\overline{y}} v\left((1-p)u\left(w_f^F(\tilde{\tilde{y}})\right)\right) f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}\right). \qquad (A78)
\end{aligned}$$

Since $v' > 0$, the last inequality implies $w_f^F(y) > p w_f^F(\tilde{\tilde{y}})$. Hence, $L_f$ is more cost-effective for the principal than the candidate solution given by Eq. $(A78)$.

To investigate the incentives of $L_f$, denote by $\overline{L_f}$ its expected value and substitute it in the agent's expected utility to obtain:

$$U\left(e_H, L, w_f^F(\tilde{\tilde{y}})\right) = \left(\frac{\overline{L_f}}{w_f^F(\tilde{\tilde{y}})}\right) u\left(w_s^F(\tilde{\tilde{y}})\right) + \frac{\overline{L_f}}{w_f^F(\tilde{\tilde{y}})}\left(1 - \frac{\overline{L_f}}{w_f^F(\tilde{\tilde{y}})}\right)\eta\left(v\left(u\left(w_f^F(\tilde{\tilde{y}})\right)\right) - \lambda v\left(-u\left(w_f^F(\tilde{\tilde{y}})\right)\right)\right),$$

$$(A79)$$

an expression that is not linear in $\overline{L}_f$. Hence, changes in $\overline{L}_f$ affect the agent's marginal utility. As it was the case with the second-best contract, the probability that maximizes the agent's utility can be found via the first order condition of Eq. $(A79)$ with respect to $p$. The resulting probability is $p_f^*(\tilde{\tilde{y}}) = \frac{1}{2} + \frac{u\left(w_f^F(\tilde{\tilde{y}})\right)}{2\eta\left(v\left(u\left(w_f^F(\tilde{\tilde{y}})\right)\right) - \lambda v\left(-u\left(w_f^F(\tilde{\tilde{y}})\right)\right)\right)}$. Hence, lottery $L_f^* := (p_f^*(\tilde{\tilde{y}}): w_s^F(\tilde{\tilde{y}}), 1 - p_f^*(\tilde{\tilde{y}}): 0)$ is proposed by the principal.

When $U\left(e_H, w(\tilde{y}), w(\tilde{\tilde{y}})\right)$ is S-shaped for any output realization $\tilde{y}$, the first-best contract, $w_f^*(y)$, consists of $L_f^*$. Suppose instead that $w_f^F(y)$ satisfying Eq. $(A77)$ is offered. Since $\overline{L}_f < w_f^F(y)$, the principal can profitably deviate from that solution by paying $L_f$ at the lower end of the output space. While $L_f^*$ is evaluated as a sizeable loss when $w_f^F(y)$ satisfying $(A77)$ is taken as reference point, that $U\left(e_H, w(\tilde{y}), w(\tilde{\tilde{y}})\right)$ is S-shaped ensures a tolerance to that risk. The principal can increase the segment for which $L_f^*$ is the solution until achieving the boundary $\hat{y}_{df} = \overline{y}$. This strategy is cost-effective. Also, proceeding in such way would not fully locate the agent in losses since gains are experienced when the outcome of the lottery $w = 0$ is taken as the reference point.

Denote by $b_f > 0$ the pay level $w_f^F(y)$ satisfying Eq. $(A75)$ evaluated at $y = \overline{y}$. It is second-best optimal to offer $L_s^* = \left(p_f^*(\overline{y}): b_f, 1 - p_{s=f}^*(\overline{y}): 0\right)$ if $y < \overline{y}$ and $b_s$ paid at $y = \overline{y}$ proving the first part of the Proposition.

When $U\left(e_H, w(\tilde{y}), w(\tilde{\tilde{y}})\right)$ is concave for any $\tilde{y} \in [\underline{y}, \overline{y}]$, the optimal contract also consists of two components: $w_f^F(y)$ satisfying Eq. $(A77)$, which implies $w_f^F(y) \geq w_s^F(\tilde{\tilde{y}})$, and $w_f^F(y)$ satisfying Eq. $(A78)$, which implies $w_f^F(y) < w_s^F(\tilde{\tilde{y}})$. Since the agent is loss averse, $\lambda > 1$, $w_f^F(y)$ satisfying $(A78)$ cannot be a solution on its own as it induces considerable disutility, leading the agent to reject the contract. Moreover, a combination of these two components exposes the agent to the risk of experiencing losses. Since $U\left(e_H, w(\tilde{y}), w(\tilde{\tilde{y}})\right)$ is concave, such a combination does not provide full insurance. Hence, it must be $w_f^F(y)$ satisfying Eq. $(A77)$ is offered. This proves the last part of the Proposition. ■

**Corollary 7.**

Let $x := u(w(y)) - u(w(g))$. Replace $u'' < 0$ from Assumption 4 for $u''(x) < 0$ if $x \geq 0$ and $u''(x) > 0$ if $x < 0$. Under such modification, receiving lottery $L := (1 - p: w_s^*(y), p: 0)$ generates the following utility:

$$KR(L) = (1 - p)u\left(w_s^*(y)\right) + p(1 - p)\eta\left(v\left(u\left(w_s^*(y)\right)\right) + \lambda v\left(-u\left(w_s^*(y)\right)\right)\right). \qquad (A80)$$

Since the second derivative of Eq. $(A80)$ is $-2\eta\left(v(u(w_s^*(y)\,))+\lambda v\left(u\big(-w_s^*(y)\big)\right)\right)$, a negative expression, the probability that maximizes utility is given by the closed-form solution $p^*(y)=\frac{1}{2}-\frac{u\left(w_S^*(y)\right)}{2\eta\left(\lambda v(-u(w_S^*(y)))+v(u(w_S^*(y)))\right)}$.

Proposition 2 (i) shows that $\hat{y}_{ds}=\overline{y}$, so it is optimal to offer $L_k{:}=(1-\ p^*(y){:}\,b_s,p^*(y){:}\,0)$ for any $y<\overline{y}$. Consider $\eta=1$ and $v'=1$. The optimal probability becomes $p^*=\frac{1}{2}+\frac{1}{2(\lambda-1)}$, where it is clear that $p^*\in(0,1)$ if $\lambda>2$ but $p^*=1$ if $\lambda\leq 2$. Consequently, $w_s^*(y)=\begin{cases}0\ if\ y<\overline{y}\\ b_s\ if\ y=\overline{y}\end{cases}$ where $b>0$ is optimal if $\lambda\leq 2$.

An alternative proof uses Proposition 2 (ii) and (iii). Under $\eta=1$ and $v'=1$, the condition for respecting first-order dominance becomes $\lambda\leq 2$. If that condition is respected Eqs. $(A65)$ and $(A66)$ become

$$\frac{1}{2u'\left(w_S^F(y)\right)}=\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right), \qquad (A81)$$

if $\theta_{\mathbb{I}}=1$, and

$$\frac{1}{(1+\lambda)u'\left(w_S^F(y)\right)}=\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right), \qquad (A82)$$

if $\theta_{\mathbb{I}}=0$. Eqs. $(A81)$ and $(A82)$ show that under $u'=1$ the optimal contract becomes a binary pay scheme. If $\lambda>2$ that binary paid becomes the outcome of a lottery realizing with $p\in(0,1)$ and paying nothing with the complement. ■

# Appendix B

**B.1 Other salience-based reference points.**

**Corollary B.1** *Under A1- A4 and the max-min rule, the agent's reference point and the second-best contract are identical to those presented in Corollary 5.*

*Proof.* The agent makes, at most, two choices: accepting the contract, or not, and choosing an effort level. These choices generate three candidates for reference point under the min-max rule: rejecting the contract and obtaining $\overline{U} \geq 0$, obtaining $\max\{w(y)\}$ after choosing $e_H$, and obtaining $\max\{w(y)\}$ after choosing $e_L$.

Denote the second-best optimal contract by $w_s^*(y)$. Notice that an optimal contract generating utility $\overline{U}$ and that is compatible with the agent's preferences is $w_f^*(y)$ from Corollary 4 (i). It is shown next that $w_f^*(y) \leq \max\{w_{SB}(y)\}$ for any $e$. We proceed by contradiction by supposing that $w_f^*(y) > \max\{w_s^*(y)\}$. In that case the agent would not accept the contract since $U(e_H, w_f^*(y), r) = \overline{U} > U(e_H, \max\{w_s^*(y)\}, r)$. A contradiction since at the optimum the participation constraint binds. Hence, it must be $w_f^*(y) \leq \max\{w_s^*(y)\}$. Moreover, since the incentive compatibility constraint binds at the optimum it must be that $w_f^*(y) < \max\{w_s^*(y)\}$, otherwise the second-best contract would not provide rewards to exert high effort. As a result, the min-max rule thus implies that $r = w_f^*(y)$.

Since the agent's preferences are characterized by prospect theory and his reference point is $r = w_f^*(y)$, the optimal incentive scheme is identical to that presented in Corollary 5.■

**Corollary B.2** *Under A1-A4 and the $w(y)$ at max P rule, the agent's reference point is $r = w_s^*(y_p)$, where $y_p$ satisfies $f(y_p|e_H) = \max\{f(y|e_H)\}$, and the second-best contract, $w_s^*(y)$, pays the lowest possible in $y < \hat{y}_p$, exhibits a bonus at $y = y_p$, and increases in performance in $y > \hat{y}_p$.*

*Proof.* Let $y_p \in [\underline{y}, \overline{y}]$ be a performance level satisfying $f(y_p|e) = \max\{f(y|e)\}$. If $f(y_p|e)$ is multimodal, define $y_p$ as the smallest output level satisfying $f(y_p|e) = \max\{f(y|e)\}$. That $[\underline{y}, \overline{y}] \subseteq \mathbb{R}^+$ , implies that the point $y_p$ attains the highest probability as compared to any $y \in [\underline{y}, \overline{y}] \setminus \{y_p\}$.

Since the agent's preferences are characterized by prospect theory, a contract with the same shape as that described in Corollary 4 remains to be optimal. Denote that contract by $w_s^*(y)$. The $w(y)$ *at max P* reference point rule entails that $r = w_s^*(y_p)$. Hence, the output level after which the bonus is awarded might be different than that given in Corollary 4. That point is defined next. Let $\hat{y}_p \in [\underline{y}, \overline{y}]$ satisfy:

$$\frac{1}{\frac{\lambda v\left(w_s^*(\hat{y}_p)\right)}{w_s^*(\hat{y}_p)}} = \nu + \gamma\left(1 - \frac{f(\hat{y}_p|e_L)}{f(\hat{y}_p|e_H)}\right). \qquad (B1)$$

According to Corollary 4, the optimal contract should pay $w_s^*(y)$ satisfying the following first-order condition

$$\frac{1}{v'\left(w_s^*(y) - w_s^*(y_p)\right)} = \nu + \gamma\left(1 - \frac{f(y|e_H)}{f(y|e_L)}\right), \qquad (B2)$$

if $y > \hat{y}_p$, and $w_s^* = 0$ if $y < \hat{y}_p$.

Finally, I demonstrate that $\hat{y}_p = y_p$. I proceed by contradiction. Suppose instead that $\hat{y}_p < y_p$. In that case, the principal is overinsuring the agent from risk in $y \in [\hat{y}_p, y_p]$ by offering $w_s^*(y)$ satisfying $(B2)$ in a segment where he is risk seeking due to diminishing sensitivity. The principal could increase profits by exposing the agent to large amounts of risk by setting $w_s^*(y) = 0$ for all $y < y_p$, including $y \in [\hat{y}_p, y_p]$, and the agent would accept such contract.

Next, suppose that $\hat{y}_p > y_p$. In that case the agent is being exposed to large amounts of risk in $y \in [y_p, \hat{y}_p]$, a segment where he is risk averse (Assumption 4). This incentivizes the agent to reject the contract. The principal anticipates this and provides insurance offering the payment scheme $w_s^*(y)$ satisfying $(B2)$ for $y \geq y_p$. Hence, it must be that $\hat{y}_p = y_p$. ■

**B.2 Gul's (1991) model**

The disappointment model of Gul (1991) differs from those of Bell (1985) and Loomes and Sugden (1986) in that the agent's reference point is his certainty equivalent. Importantly, the certainty equivalent includes the agent's psychological utility component.

More formally, consider the general specification of reference dependence given in $(2)$. As with the other previous disappointment models allow for expected consumption utility by letting $\phi = 1$. Under these restrictions, the agent's certainty equivalent is the level $CE \in \mathbb{R}$ that satisfies $U(e_H, w(y), CE) = u(CE)$ for a given incentive scheme $w(y)$.

The first-best and second-best optimal contracts under Gul's (1991) preferences are presented in the following next corollary.

**Corollary B.3** *Under assumptions A1-A4, $\phi = 1$, and $r = CE$, there exist unique output levels $y_{cf}, y_{cs} \in [\underline{y}, \overline{y}]$ such that:*

i) *The first-best contract, $w_f^*(y)$, is equal to that given in Corollary 6 with $y_{mf}$ replaced by $y_{cf}$.*

ii) *The second-best contract, $w_s^*(y)$, is equal to that given in Corollary 6 with $y_{mf}$ replaced by $y_{cs}$.*

*Proof.* Let $\phi = 1$, $u' = 1$, and $r = CE$. Under the assumed restrictions, $U(e_H, w(\tilde{y}), CE)$ is S-shaped since $-\frac{v''\left(u(r)-u(w(y))\right)}{v'\left(u(r)-u(w(y))\right)} > 0$ in the domain of losses, corroborating equation $(A3)$. According to Theorem 1, it is optimal to set $w_s^*(y) = 0$ for low output levels. Moreover, the first-order condition is necessary and sufficient only in the domain of gains.

The point at which the bonus of the second-best contract is awarded is defined by the following output level, which adapts Eq. $(A14)$ to account for the assumed restrictions,

$$\frac{1}{\frac{1+\lambda v(CE)}{CE}} = \mu + \gamma\left(1 - \frac{f(\hat{y}_{cs}|e_L)}{f(\hat{y}_{cs}|e_H)}\right). \qquad (B3)$$

The bonus is awarded when the unique output level $\hat{y}_{cs}$ that satisfies $(B3)$ is surpassed. Hence, the optimal contract is given by $w_s^*(y) = \begin{cases} 0 & if\ y < \hat{y}_{cs}, \\ CE + f'\left(\frac{1}{\eta}\left(\frac{1}{\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right)} - 1\right)\right) & if\ y \geq \hat{y}_{cs}. \end{cases}$ This proves the first part of the corollary.

We turn to study the first-best contract. Hence, consider $\gamma = 0$ in addition to $\phi = 1$, $u' = 1$, and $r = CE$. Eq. $(A55)$ becomes $\frac{1}{\left(1+\eta v'\left(w_f^*(y)-CE\right)\right)} = \mu$, which after some manipulations yields $w_f^*(y) = CE + f'\left(\frac{1}{\eta}\left(\frac{1}{\mu} - 1\right)\right)$. According to Eq. $(A56)$, that transfer exhibits $\frac{dw_f^*(y)}{dy} = 0$.

Also, Theorem 1 shows that in the domain of losses it is also first-best optimal to offer $w_f^*(y) = 0$. The transition from $w_f^*(y) = 0$ to $w_f^*(y) = CE + f'\left(\frac{1}{\eta}\left(\frac{1}{\mu} - 1\right)\right)$ is given by the following equality, which adapts Eq. $(A23)$ to account for the considered restrictions,

$$\int_{\hat{y}_{cf}}^{\overline{y}} w_f^*(y) f(y|e_H)\, \mathrm{d}y + \eta \int_{\hat{y}_{cf}}^{\overline{y}} v\left(w_f^*(y) - CE\right) f(y|e_H)\, \mathrm{d}y - \lambda \int_{\underline{y}}^{\hat{y}_{cf}} v(CE) f(y|e_H) \mathrm{d}y - c = \overline{U}. \quad (B4)$$

Hence, the output $\hat{y}_{cf}$ that satisfies Eq.$(B4)$ provides that transition. Theorem 1 shows that this output level is unique and interior. Hence, the optimal contract is $w_f^*(y) = \begin{cases} 0 & if\ y < \hat{y}_{cf}, \\ CE + f'\left(\frac{1}{\eta}\left(\frac{1}{\mu} - 1\right)\right) & if\ y \geq \hat{y}_{cf}. \end{cases}$ This proves the third part of the corollary.

Next let $\phi = 1$, $v' = 1$, and $r = CE$. Under these restrictions $U(e_H, w(\tilde{y}), CE)$ is concave since $-\frac{u''(w(y))}{u'(w(y))} > 0$, a contradiction of the condition in Eq. $(A3)$. Hence, the solutions from the first-order conditions are necessary and sufficient to solve the maximization problem of the principal.

The first order conditions given by Eqs. $(A59)$and $(A60)$ provide the solution. The transition from $w_s^*(y)$ satisfying $(A59)$ to $w_s^F(y)$ satisfying $(A60)$ is given by the unique output level $\hat{y}_{cs} \in (\underline{y}, \overline{y})$ that satisfies:

$$\int_{\hat{y}_{cs}}^{\overline{y}} u(w_S^*(y))f(y|e_H)\,\mathrm{d}y + \eta \int_{\hat{y}_{cs}}^{\overline{y}} \left(u(w_S^*(y)) - u(CE)\right)f(y|e_H)\,\mathrm{d}y + \int_{\underline{y}}^{\hat{y}_{cs}} u(w_S^*(y)f(y|e_H)\mathrm{d}y$$
$$- \lambda\eta \int_{\underline{y}}^{\hat{y}_{cs}} \left(u(CE) - u\left(w_s^*(y)\right)\right) f(y|e_H)\mathrm{d}y - c = \overline{U}. \qquad (B5)$$

The existence of $\hat{y}_{cs}$ is guaranteed by the fact that the solutions from Eqs. $(A59)$ and $(A60)$ make the participation constraint bind for gains and losses, respectively. As a result, the optimal incentive scheme is given by:

$$w_f^*(y) = \begin{cases} h'\left(\dfrac{(1+\eta)}{\mu + \gamma\left(1 - \dfrac{f(y|e_L)}{f(y|e_H)}\right)}\right) & if\ y \geq \hat{y}_{cs}\,, \\ h'\left(\dfrac{(1+\eta\lambda)}{\mu + \gamma\left(1 - \dfrac{f(y|e_L)}{f(y|e_H)}\right)}\right) & if\ y < \hat{y}_{cs}. \end{cases} \qquad (B6)$$

This proves the second part of the corollary.

To conclude, we analyze the first-best contract. Hence, consider $\gamma = 0$ in addition to $\phi = 1$, $v' = 1$, and $r = CE$. The first-order condition in $(A59)$ becomes $\frac{1}{u'\left(w_f^*(y)\right)(1+\eta)} = \mu$, which after algebraic manipulations gives $w_f^*(y) = h'\left(\frac{(1+\eta)}{\mu}\right)$. Eq. $(A8)$ shows that $\frac{dw_f^*(y)}{dy} = 0$ under the considered restrictions. Finally, Theorem 1 shows that $w_f^*(y) = h'\left(\frac{(1+\eta)}{\mu}\right)$ is first-best optimal when $U(e_H, w(y), r)$ is concave. ■

An agent with disappointment averse preferences and who adopts his certainty equivalent as reference point is insured and motivated with contracts that greatly resemble those described in Corollary 6. Therefore, contracts with a bonus enable the principal to exploit the agent's irrationalities of loss aversion and diminishing sensitivity in an optimal way.

However, the reference point rule specified by Gul (1991)'s mode generates potential differences in the location and magnitude of the bonus. Intuitively, a (globally) risk averse agent must exhibit $CE < \mathbb{E}(w(y))$. Hence, to guarantee that the contract is accepted, the principal protects this agent from risk by awarding the bonus at lower output levels as compared to the hypothetical case in which the agent was risk neutral, $CE = \mathbb{E}(w(y))$. Hence, $y_{cf} < y_{mf}$ and $y_{cs} < y_{ms}$. These more lenient threshold levels come at the cost of the magnitude of the bonus included in each contract, which becomes smaller as compared to the risk neutral case. In that way, the principal keeps the agent just indifferent between accepting or rejecting the contract. A similar intuition leads to the conclusion that for a globally risk seeking agent $y_{c1} > y_{m1}$, $y_{c2} > y_{m2}$, and both contracts include a larger bonus. A result that is consistent with the comparative static presented in Corollary 2.

### B.3 Adapting the model to accommodate De Meza and Webb (2007)

The model can be adapted to allow for the results of De Meza and Webb (2007). The following adaptation of Assumption 4 is considered.

**Assumption B1 (AB1)**. *The agent's value function $V$ is the piece-wise function:*

$$V(w,r) = \begin{cases} 0 & \text{if } w(y) > r, \\ -v\left(u(r) - u\big(w(y)\big)\right) & \text{if } r \leq w(y). \end{cases}$$

*with properties:*

- $v: \mathbb{R}^+ \longrightarrow \mathbb{R}^+$;
- $v$ *is* $C^2$;
- $v(0) = 0$;
- $v' > 0 \; \forall \, y \in \left[\underline{y}, \overline{y}\right]$;
- $v'' \leq 0$;
- $v$ *has a differentiable inverse* $f := v^{-1}$.

There are three key differences between A4 and AB1. First, loss aversion in the usual sense, i.e. sign dependence, is abandoned. In other words, $\lambda = 1$. Second, outcome comparisons relative to the reference point are restricted to the domain of losses. A consequence of that assumption is that diminishing sensitivity applies only to losses. This effect is referred as loss aversion by De Meza and Webb (2007). However, this way of modeling loss aversion contradicts standard definitions when $v'' < 0$. (Tversky and Kahneman, 1992, Köbberling and Wakker, 2005). Third, the transition from gains to losses is not given at the reference point but once that value is surpassed, that is for $w(y) > r$.

Moreover, let $\phi = 1$ and $\eta = 1$. All in all, the decision-maker's preferences are given by

$$U(e, w(y), r) = \int_{\underline{y}}^{\overline{y}} u\big(w(y)\big) f(y|e) \mathrm{d}y - \int_{\underline{y}}^{\overline{y}} \left((1 - \theta_{\mathbb{I}}) v\left(u(r) - u\big(w(y)\big)\right)\right) f(y|e) \mathrm{d}y - c(e) \qquad (B7)$$

Under preferences as in Eq. $(B7)$, the results of De Meza and Webb (2007) follow. When $v' = l > 0$, that is when diminishing sensitivity is assumed to be piece-wise linear, then the results of their Proposition 1 follow. Throughout, they interpret $l > 0$ as loss aversion.

In that case, $U(e, w(\tilde{y}), r)$ for any $\tilde{y} \in [\underline{y}, \overline{y}]$ is concave. The first order conditions from Eqs. $(A6)$ and $(A7)$, given in the Proof of Theorem 1, become

$$\frac{1}{u'(w_s^*(y))} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right), \qquad (B8)$$

and

$$\frac{1}{u'(w_s^*(y))(1+l)} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right), \qquad (B9)$$

respectively. These two equations are increasing in performance. Also, these equations are equal to Eq.(8) in their paper, crucial for their proof of Proposition 1.

When $l$ is small enough, the principal can offer $w_s^*(y)$ satisfying $(B9)$ while still guaranteeing $U(e, w_s^*(y), r) = \overline{U}$. That is because expected consumption utility, $\phi = 1$, ensures that the participation constraint binds even though the contract locates the agent in the domain of losses. Since $r$ is still part of the domain of losses, due to $w < r$, the optimal incentive scheme can be insensitive at high output levels, i.e. $w_s^*(y) = r$. This leads to Proposition 1 (ii) and Figure 1a in De Meza and Webb (2007).

For larger $l$, the agent needs to be translated to the domain of gains to guarantee $U(e, w_s^*(y), r) = \overline{U}$. In that case, expected consumption utility does not fully outweigh losses in the contract and at high ouput levels $w(y) > r$ must be ensured. Since $\lambda = 1$, there is no jump or kink when the transition from $w_s^*(y) = r$ to $w_s^*(y)$ satisfying Eq. $(B8)$ takes place. This case generates Proposition 1 (iv) and Figure 1c in De Meza and Webb (2007). For even higher $l$, the exposure of the agent in the domain of losses is reduced by paying $w_s^*(y) = r$ for low output levels and $w_s^*(y)$ satisfying Eq. $(B8)$ being paid at intermediate and high output levels. This covers their Proposition 1 (iii) and Figure 1b.

Finally, when $v'' < 0$, which De Meza and Webb (2007) denote as non-linear loss aversion, the results of their Proposition 2 follow. Let $U(e, w(\tilde{y}), r)$ be S-shaped. The proof of Theorem 1 shows that in the domain of losses either $w_s^*(y) = 0$ or $w_s^*(y) = r$ must be given. Hence, the optimal incentive scheme is given by $w_s^*(y) = 0$, $w_s^*(y) = r$, and $w_s^*(y)$ satisfying Eq. $(B8)$. As above, the magnitude of $v$ determines the shape of the incentive scheme. When $v$ is small enough, then a combination of $w_s^*(y) = 0$ and $w_s^*(y) = r$ is given to the agent. This case is given in Figure 2c in Meza and Webb (2007). When $v$ is larger, the agent needs to be transitioned in the domain of gains for high output levels. Then, the optimal incentive scheme is a combination of $w_s^*(y) = 0$, $w_s^*(y) = r$, and $w_s^*(y)$ satisfying Eq. $(B8)$. Finally, a large enough $v$ leads to an optimal incentive scheme that combines $w_s^*(y) = r$ and $w_s^*(y)$ satisfying Eq. $(B8)$. This case is depicted in Figure 2a.

# Appendix C

**Proposition 3.**

Denote by $\mu \geq 0$ and $\gamma \geq 0$ the Lagrangian multipliers of the agent's participation and incentive compatibility constraints. First, let $S(y) < r_p + w(y)$. In that case, the Lagrangian of the principal's maximization program writes as:

$$\mathcal{L} = \left(-\lambda_p \left(r_p + w(y) - S(y)\right)\right) f(y|e_H)$$

$$+\mu\left[\phi u(w(y) + \theta_{\mathbb{I}}\eta v(u(w(y)) - u(r))f(y|e_H) - \lambda(1-\theta_{\mathbb{I}})\eta v(u(r) - u(w(y)))f(y|e_H) - c - \overline{U}\right] \quad (C1)$$

$$+\gamma\left[\phi u(w(y)(f(y|e_H) - f(y|e_L)) + \eta\theta_{\mathbb{I}} v\left(u(w(y)) - u(r)\right)(f(y|e_H) - f(y|e_L)) - \lambda(1-\theta_{\mathbb{I}})\eta v(u(r) - u(w(y)))(f(y|e_H) - f(y|e_L)) - c\right].$$

Pointwise optimization with respect to $w(y)$ gives

$$-f(y|e_H)\lambda_p + u'(w(y)\mu\left[\phi + \eta\theta_{\mathbb{I}} v'\left(u(w(y)) - u(r)\right) f(y|e_H) + \eta\lambda(1-\theta_{\mathbb{I}})v'\left(u(r) - u(w(y))\right) f(y|e_H)\right]$$
$$+ u'(w(y)\gamma\left[\phi + \eta\theta_{\mathbb{I}} v'\left(u(w(y)) - u(r)\right)(f(y|e_H) - f(y|e_L)) + \eta\lambda(1-\theta_{\mathbb{I}})v'\left(u(r) - u(w(y))\right)(f(y|e_H) - f(y|e_L))\right] = 0 \quad (C2)$$

Denote by $w_s^F(y)$ the transfer satisfying $(C2)$. After algebraic manipulations, I find the following expressions

$$\frac{\lambda_p}{u'(w(y)\left(\phi + \eta\theta_{\mathbb{I}} v'\left(u(w(y)) - u(r)\right)\right)} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right), \quad (C3)$$

if $\theta_{\mathbb{I}} = 1$, and

$$\frac{\lambda_p}{u'(w(y)\left(\phi + \lambda\eta\theta_{\mathbb{I}} v'\left(u(r) - u(w(y))\right)\right)} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right), \quad (C4)$$

and $\theta_{\mathbb{I}} = 0$. As in Theorem 1, if $U(e_H, w(\tilde{y}), r)$ is concave, the conditions given in $(C3)$ and $(C4)$ are necessary and sufficient to solve the maximization problem. Instead, if $U(e_H, w(\tilde{y}), r)$ for any $\tilde{y} \in [\underline{y}, \overline{y}]$ is S-shaped, the principal is better of offering lottery $L = (p{:}\, r, 1-p{:}\, 0)$ with a probability $p \in (0,1)$ that satisfies:

$$\phi u(w_s^F(y)) - \lambda v\left(u(r) - u\left(w_s^F(y)\right)\right) = \phi p u(r) - (1-p)\lambda v\left(u(r)\right). \quad (C5)$$

Hence, paying $L$ does not change participation and incentive compatibility constraint. Also, offering that lottery is more cost effective for the principal since from $(C5)$:

$$\phi\left(u\big(w_s^F(y)\big)-pu(r)\right)\leq \lambda v\left(u(r)-u\big(w_s^F(y)\big)\right)-\lambda v\big((1-p)u(r)\big),$$

implying $w_s^F(y) > pr$.

Denote by $\hat{y}_s$ the output level satisfying:

$$\frac{1}{\frac{\phi u(r)+\lambda\eta v\big(u(r)\big)}{r}}=\mu+\gamma\left(1-\frac{f(\hat{y}_s|e_L)}{f(\hat{y}_s|e_H)}\right). \qquad (C6)$$

That output level is unique since the left-hand side of Eq.$(C6)$ is constant in $y$ and is positive, while the right-hand side of that equation increases in $y$ over the domain $[0,\infty)$.

When $y < \hat{y}_s$, the scheme pays $w_s^F(y) = 0$. That is because when offered $L$ the agent's utility can be expressed as

$$U(e_H,L,r)=-\lambda\left(1-\frac{\overline{L}}{r}\right)u(r)-c, \qquad (C7)$$

where $\overline{L}$ is the expected value of $L$. Notice that Eq. $(C7)$ is linear in $\overline{L}$. Hence, changes in $\overline{L}$ do not affect the agent's marginal utility and the principal can afford to set $p = 0$. Instead, if $\hat{y}_s < y$, the agent's payment can be set $p = 1$, which brings him to the domain of gains. In that domain, the principal should be paid $w_s^F(y)$ satisfying $(C3)$. Therefore, the solution to the principal's problem is

$$w_{SB}(y)=\begin{cases}0 & if\ y<\hat{y}_s,\\ w_s^F(y)\text{ from }(C3) & if\ y\geq\hat{y}_s.\end{cases} \quad (C8)$$

If $U(e_H, w(\tilde{y}), r)$ for any $\tilde{y} \in [\underline{y}, \overline{y}]$ is concave $w_s^*(y)$, consists of two components: $w_s^F(y)$ satisfying $(C3)$, which implies $w_s^F(y) \geq r$, and $w_s^F(y)$ satisfying $(C4)$, which implies $w_s^F(y) < r$. Because the agent is loss averse, $\lambda > 1$, $w_f^F(y)$ satisfying $(C4)$ cannot be a solution on its own as it induces considerable disutility, leading the agent to reject the contract. Also $(C3)$ on its own is not optimal, as the principal would be fully protecting the agent from losses, demotivating him to exert high effort to avoid the disutility from experiencing losses. Hence, the optimal contract combines the first-order conditions $(C3)$ and $(C4)$. The transition from $w_s^F(y)$ satisfying $(C3)$ to $w_s^F(y)$ satisfying $(C4)$ is defined next. Let $\hat{y}_s \in (\underline{y}, \overline{y})$ be the output level satisfying:

$$\phi\int_{\hat{y}_s}^{\overline{y}}u(w_s^F(y))f(y|e_H)\,\mathrm{d}y+\eta\int_{\hat{y}_s}^{\overline{y}}v\big(u(w_s^F(y))-u(r)\big)f(y|e_H)\,\mathrm{d}y+\phi\int_{\underline{y}}^{\hat{y}_s}u(w_s^F(y)f(y|e_H)\mathrm{d}y$$
$$-\lambda\eta\int_{\underline{y}}^{\hat{y}_s}v\left(u(r)-u\big(w_s^F(y)\big)\right)f(y|e_H)\mathrm{d}y-c\ =\overline{U}. \qquad (C9)$$

The existence of $\hat{y}_s$ is guaranteed by the fact that the two solutions given by Eqs. $(C3)$ and $(C4)$ make the participation constraint bind for gains and losses. Uniqueness of $\hat{y}_s$ is because the magnitude of the first four expressions depends on $\hat{y}_s$. The first three expressions in the left-hand side of Eq. $(C9)$ are positive and become larger as $\hat{y}_s$ decreases, while the fourth expression is negative and becomes larger as $\hat{y}_s$ increases. Since $\overline{U}$ is constant, there exists a unique $\hat{y}_s$ that satisfies $(C9)$.

As a result, the optimal incentive scheme is given by:

$$w_s^*(y) = \begin{cases} w_s^F(y) & satisfying\ (A6)\ \ if\ y \geq \hat{y}_s, \\ w_s^F(y) & satisfying\ (A7)\ \ if\ y < \hat{y}_s. \end{cases} \qquad (C10)$$

Notice that this solution exhibits a discrete jump at $y = \hat{y}_s$ since $\lambda > 1$ appears in the denominator of the right-hand side of $(C4)$ but this coefficient does not enter in $(C3)$.

Now suppose that $S(y) \geq r_p + w(y)$. Since principal's and agent's objective functions are identical to thos studied in Theorem 1, that solution remains optimal.

Denote by $\hat{y}_p \in [\underline{y}, \overline{y}\,]$ the output level satisfying $S(\hat{y}_p) - r_p - w_s^F(\hat{y}_p) = 0$. The existence of that output level is guaranteed by $S' > 0$, $S(0) = 0$, $w_s^{F\prime}(y) > 0$ in $y > \hat{y}_s$, and $w_{SB} = 0$ in $y < \hat{y}_s$. There follow two relevant cases. Namely, $\hat{y}_s < \hat{y}_p$ and $\hat{y}_s > \hat{y}_p$.

Let $\hat{y}_s < \hat{y}_p$. If $y < \hat{y}_s < \hat{y}_p$, both agent and principal are in the domain of losses. Then, $w_s^*(y) = 0$ is given when $U(e_H, w(\tilde{y}), r)$ is S-shaped while $w_s^*(y)$ satisfying $(C4)$ is given when $U(e_H, w(y), r)$ is concave. If $\hat{y}_s < y < \hat{y}_p$, the principal is in the domain of losses, while the agent is in the domain of gains. In that case, the principal offers insurance to the agent by paying $w_s^F(y)$ satisfying $(C3)$. Finally, for $\hat{y}_s < \hat{y}_p < y$ both principal and agent are in the domain of gains, so the principal offers $w_s^F(y)$ satisfying $(A5)$. Since $\lambda_p > 2$ is absent in $(A5)$ but present in $(C3)$, the agent's compensation exhibits a kink at $y = \hat{y}_p$.

Let $\hat{y}_p < \hat{y}_s$. If $y < \hat{y}_p < \hat{y}_s$, both agent and principal are in the domain of losses. Again, $w_s^*(y) = 0$ is given when $U(e_H, w(\tilde{y}), r)$is S-shaped and $w_s^*(y)$ satisfying $(C4)$ is given when $U(e_H, w(\tilde{y}), r)$ is concave. If $\hat{y}_p < y < \hat{y}_s$, the agent is in the domain of losses, while the principal is in the domain of gains. The solution is in that case identical to Theorem 1. Namely, the principal offers $w_s^*(y) = 0$ when $U(e_H, w(\tilde{y}), r)$is S-shaped and $w_s^*(y)$ satisfying $(C4)$ is given when $U(e_H, w(\tilde{y}), r)$ is concave. Finally, for $\hat{y}_p < \hat{y}_s < y$ both are in the domain of gains, the principal offers $w_s^F(y)$ satisfying $(A3)$. There is no kink in that case. ■

**Proposition 7.**

The agent with $\lambda_H$ faces the following adverse selection constraint,

$$\int_{\underline{y}}^{\overline{y}} u(w(y)^H) f(y|e_H)\mathrm{d}y + \int_{\hat{y}_H}^{\overline{y}} v(u(w(y)^H) - u(r)) f(y|e_H)\mathrm{d}y - \lambda_H \int_{\underline{y}}^{\hat{y}_H} v(r - w(y)^H) f(y|e_H)\mathrm{d}y - c$$
$$\geq \max_{e\in\{e_L,e_H\}} \left\{ \int_{\underline{y}}^{\overline{y}} u(w(y)^L)\, f(y|e)\mathrm{d}y + \int_{\hat{y}_L}^{\overline{y}} v(w(y)^L - r) f(y|e)\mathrm{d}y - \lambda_H \int_{\underline{y}}^{\hat{y}_L} v(r - w(y)^L) f(y|e)\mathrm{d}y - c(e) \right\}, \qquad (C11)$$

moral hazard incentive constraint,

$$\int_{\underline{y}}^{\overline{y}} u(w(y)^H) f(y|e_H)\mathrm{d}y + \int_{\hat{y}_H}^{\overline{y}} v(u(w(y)^H) - u(r)) f(y|e_H)\mathrm{d}y - \lambda_H \int_{\underline{y}}^{\hat{y}_H} u(r - w(y)^H) f(y|e_H)\mathrm{d}y - c$$
$$\geq \int_{\underline{y}}^{\overline{y}} u(w(y)^H) f(y|e_L)\mathrm{d}y + \int_{\hat{y}_H}^{\overline{y}} v(u(w(y)^H) - u(r)) f(y|e_L)\mathrm{d}y - \lambda_H \int_{\underline{y}}^{\hat{y}_H} v\big(u(r) - u(w(y)^H)\big) f(y|e_L)\mathrm{d}y, \qquad (C12)$$

and participation constraint

$$\int_{\underline{y}}^{\overline{y}} u(w(y)^H) f(y|e_H)\mathrm{d}y + \int_{\hat{y}_H}^{\overline{y}} v(u(w(y)^H) - u(r)) f(y|e_H)\mathrm{d}y - \lambda_H \int_{\underline{y}}^{\hat{y}_H} v\big(u(r) - u(w(y)^H)\big) f(y|e_H)\mathrm{d}y - c \geq \overline{U}. \qquad (C13)$$

Similarly, the agent with $\lambda_L$ faces the following adverse selection constraint,

$$\int_{\underline{y}}^{\overline{y}} u(w(y)^L) f(y|e_H)\mathrm{d}y + \int_{\hat{y}_L}^{\overline{y}} v(u(w(y)^L) - u(r)) f(y|e_H)\mathrm{d}y - \lambda_L \int_{\underline{y}}^{\hat{y}_L} v\big(u(r) - u(w(y)^L)\big) f(y|e_H)\mathrm{d}y - c$$
$$\geq \max_{e\in\{e_L,e_H\}} \left\{ \int_{\underline{y}}^{\overline{y}} u(w(y)^H) f(y|e_H)\mathrm{d}y + \int_{\hat{y}_H}^{\overline{y}} v(u(w(y)^H) - u(r)) f(y|e)\mathrm{d}y - \lambda_L \int_{\underline{y}}^{\hat{y}_H} v\big(u(r) - u(w(y)^H)\big) f(y|e)\mathrm{d}y - c(e) \right\}, \qquad (C14)$$

moral hazard incentive constraint,

$$\int_{\underline{y}}^{\overline{y}} u(w(y)^L) f(y|e_H)\mathrm{d}y + \int_{\hat{y}_L}^{\overline{y}} v(u(w(y)^L) - u(r)) f(y|e_H)\mathrm{d}y - \lambda_L \int_{\underline{y}}^{\hat{y}_L} v(u(r) - u(w(y)^L)) f(y|e_H)\mathrm{d}y - c$$
$$\geq \int_{\underline{y}}^{\overline{y}} u(w(y)^L) f(y|e_L)\mathrm{d}y + \int_{\hat{y}_L}^{\overline{y}} v\big(u(w(y)^L) - u(r)\big) f(y|e_L)\mathrm{d}y - \lambda_L \int_{\underline{y}}^{\hat{y}_L} v\big(u(r) - u(w(y)^L)\big) f(y|e_L)\mathrm{d}y, \qquad (C15)$$

and participation constraint

$$\int_{\underline{y}}^{\overline{y}} u(w(y)^H) f(y|e_H) \mathrm{d}y + \int_{\hat{y}_L}^{\overline{y}} v(u(w(y)^L) - u(r)) f(y|e_H) \mathrm{d}y - \lambda_L \int_{\underline{y}}^{\hat{y}_L} v\big(u(r) - u(w(y)^L)\big) f(y|e_H) \mathrm{d}y - c \geq \overline{U}. \quad (C16)$$

The agent with $\lambda_L$ mimicking the agent with $\lambda_H$ derives the following utility $U(\hat{e}, w_H, r, \lambda_L)$ for a given effort level $\hat{e}$,

$$\begin{aligned} U(\hat{e}, w(y)^H, r, \lambda_L) &= \int_{\underline{y}}^{\overline{y}} u(w(y)^H) f(y|\hat{e}) \mathrm{d}y + \int_{\hat{y}_H}^{\overline{y}} v(u(w(y)^H) - u(r)) f(y|\hat{e}) \mathrm{d}y \\ &- \lambda_H \int_{\underline{y}}^{\hat{y}_H} v\big(u(r) - u(w(y)^H)\big) f(y|\hat{e}) \mathrm{d}y \\ &+ (\lambda_H - \lambda_L) \int_{\underline{y}}^{\hat{y}_H} v\big(u(r) - u(w(y)^H)\big) f(y|\hat{e}) \mathrm{d}y - c(\hat{e}) \\ &= U(\hat{e}, w_H, r, \lambda_H) + (\lambda_H - \lambda_L) \int_{\underline{y}}^{\hat{y}_H} v(u(r) - u(w(y)^H)) f(y|\hat{e}) \mathrm{d}y. \end{aligned} \quad (C17)$$

Since $\lambda_H > \lambda_L$ and $r > w(y)^H$ in $y \in (\underline{y}, \hat{y}_H)$, the agent derives informational rents. The agent with $\lambda_H$ mimicking the agent with $\lambda_L$ derives the following utility for a given effort level $\hat{e}$,

$$\begin{aligned} U(\hat{e}, w(y)^L, r, \lambda_H) &= \int_{\hat{y}_L}^{\overline{y}} v(u(w(y)^L) - u(r)) f(y|\hat{e}) \mathrm{d}y - \lambda_L \int_{\underline{y}}^{\hat{y}_L} v\big(u(r) - u(w(y)^L)\big) f(y|\hat{e}) \mathrm{d}y - c(\hat{e}) \\ &- (\lambda_H - \lambda_L) \int_{\underline{y}}^{\hat{y}_L} v\big(u(r) - u(w(y)^L)\big) f(y|\hat{e}) \mathrm{d}y - c(\hat{e}) \\ &= U(\hat{e}, w(y)^L, r, \lambda_L) - (\lambda_H - \lambda_L) \int_{\underline{y}}^{\hat{y}_L} v(u(r) - u(w(y)^L)) f(y|\hat{e}) \mathrm{d}y. \end{aligned} \quad (C18)$$

Eq.$(C18)$ shows that engaging in that strategy is not profitable. Next, use Eqs. $(C17)$ and $(C18)$ to rewrite the adverse selection constraints in Eqs. $(C11)$ and $(C14)$ as follows:

$$\int_{\hat{y}_H}^{\overline{y}} u(w(y)^H - r) f(y|e_H) \mathrm{d}y - \lambda_H \int_{\underline{y}}^{\hat{y}_H} u(r - w(y)^H) f(y|e_H) \mathrm{d}y - c \geq \max_{e \in \{e_L, e_H\}} \left\{ U(e, w(y)^L, r, \lambda_L) - (\lambda_H - \lambda_L) \int_{\underline{y}}^{\hat{y}_L} u(r - w(y)^L) f(y|e) \mathrm{d}y \right\}, \quad (C19)$$

and

$$\int_{\hat{y}_L}^{\overline{y}} u(w(y)^L - r)f(y|e_H)\mathrm{d}y - \lambda_L \int_{\underline{y}}^{\hat{y}_L} u(r - w(y)^L)f(y|e_H)\mathrm{d}y - c$$

$$\geq \max_{e\in\{e_L,e_H\}} \left\{U(e, w(y)^H, r, \lambda_H) + (\lambda_H - \lambda_L)\int_{\underline{y}}^{\hat{y}_H} u(r - w(y)^H)f(y|e)\mathrm{d}y\right\}, \qquad (C20)$$

respectively.

From the above equations it can be concluded that $(C13)$ and $(C19)$ imply $(C16)$, so it must be that $(C16)$ slacks at the optimum while $(C13)$ binds. Moreover, since $(C17)$ and $(C18)$ show that only the agent with $\lambda_L$ derives profits when mimicking, then $(C19)$ and is strictly satisfied and $(C20)$ binds at the optimum. Denote by $w_S^*(y)^i$ the contract from Theorem 1. From the proof of that Theorem, it is known that $e_H$ generates high effort. This reduces the number of constraints to two, namely:

$$U(e_H, w_S^*(y)^H, r, \lambda_H) = \overline{U}, \qquad (C21)$$

and

$$U(e_H, w_S^*(y)^L, r, \lambda_L) = U(e_H, w_S^*(y)^H, r, \lambda_H) + (\lambda_H - \lambda_L)\int_{\underline{y}}^{\hat{y}_H} v\big(u(r) - u(w_S^*(y)^H)\big)f(y|e_H)\mathrm{d}y. \quad (C22)$$

Solving the above equations yields that $w_S^*(y)^H$ must satisfy $U(e_H, w_S^*(y)^H, r, \lambda_H) = \overline{U}$, and $w_S^*(y)^L$ must yield $U(e_H, w_S^*(y)^L, r, \lambda_L) = \overline{U} + (\lambda_H - \lambda_L)\int_{\underline{y}}^{\hat{y}_H} v\big(u(r) - u(w_S^*(y)^H)\big)f(y|e_H)\mathrm{d}y$. ■