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 exhibits reference dependence. The theoretical framework of this paper incorporates the most prominent characterizations of reference-dependent preferences and integrates the most 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. This contract shape allows the principal to profitably exploit the agent's irrationalities of loss aversion and diminishing sensitivity to extract output. 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 risky outcomes 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). This type of evaluation generates irrationalities— deviations from expected utility— that explain important 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 crucial to the design of incentives is manifest in the solution to the classical moral hazard problem. There, optimal contracts emerge from a tradeoff between providing full *insurance* to the agent and eliciting high effort (Mirrlees 1976, Holmström, 1979); incentives are thus designed according to the agent's risk tolerance. Consequently, abandoning expected utility to adopt *descriptively valid* theories of risk sets the stage for obtaining theoretical predictions that more closely resemble the compensation practices used by organizations, addressing the paradox put forward by Salanie (2003), and that effectively motivate individuals in practice.

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 to the principal-agent problem. Importantly, this finding takes place regardless of the theory of risk used to model reference dependent preferences and irrespective of the rule used to define reference points.

Throughout, a bonus is defined 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 including bonuses. This paper provides a foundation to this widespread practice grounded in preference.

A bonus contract is optimal because it allows the principal to profitably exploit the agent's irrationalities. For low output levels, the bonus is not awarded, and the contract specifies modest

[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).

transfers that are evaluated as losses because they fall short of the agent's reference point. When facing this prospect, the agent is motivated by virtue of *loss aversion*. That is, he will exert high effort to avoid experiencing such losses. In addition, the agent's risk-seeking attitudes from *diminishing sensitivity* imply that the risk exposure inflicted by these low transfers will be tolerated. The bonus' magnitude and location is crucial for the contract to be accepted. Its magnitude ensures that the agent is transitioned from the domain of losses to the domain of gains. Also, the output level after which the bonus is awarded is chosen so that the losses included in the contract are, on expectation, offset, and the agent just meets his reservation utility. In this way, the contract is accepted and motivates the agent at the expense of his irrationalities.

This paper contributes to the literature of incentive theory in several ways. First, it provides an explanation for bonuses other than 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 empirical regularity of individuals being predominantly risk averse (Holt and Laury, 2002, Wakker, 2010). 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, and subsequently finds that bonuses are optimal.

Second, it unifies and generalizes results that were hitherto scattered in the literature. Existing studies incorporating reference dependence in a principal-agent framework differ in the way risk attitude is characterized and 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 different solutions and interpretations. For example, two distinct payment modalities such as stochastic contracts—in which the principal turns a blind-eye to the agent's performance signals—and option-like contracts—a contract that is performance-insensitive at low performance levels—are proposed solutions to the principal problem when the agent is loss averse (De Meza and Webb, 2007, Herweg et al., 2010).

The theoretical framework of this paper integrates the most prominent 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). Also, the model adopts the most well-known reference point rules, such as the status quo, max-min (Hershey and Schoemaker, 1985), goals as reference points (Heath et al., 1999), expectations-based reference points (Bell, 1985), and the outcomes of the contract as the reference point (Delquie and Cillo, 2006).

The generality of the theoretical framework enables me to obtain the remarkable result that bonuses are optimal regardless of the assumed theory of risk and irrespective of the reference point. This is the result presented in Theorem 1. Notably, this finding emerges without making restrictions about the agent's degree of loss aversion and utility curvature, as Herweg et al. (2010) do, without committing to a functional form of utility or to exogenous reference points, contrasting Dittmann et al. (2010), and using models with axiomatic foundations, unlike De Meza and Webb (2007).[2] Moreover, the model is flexible enough as to include well-known results as special cases. For example, that stochastic contracts can be optimal under severe loss aversion (Herweg et al., 2010, Corgnet and Hernán-González 2019).

Third, the paper provides novel and interesting results that fill existing gaps in the. For example, the theoretical framework reconciles prospect theory with Kőszegi and Rabin (2006, 2007)'s choice acclimating equilibrium. When the agent's behavior is assumed to be guided by salience considerations, the reference point turns out to be the potential outcomes of the contract. This approach to modeling stochastic reference points is preferable not only because they emerge as endogenous outcomes of the model, instead of being assumed, but also because absurd implications, typically encountered for prospect theory with expectations-based reference points,

[2] These are not trivial restrictions. First, many studies found loss aversion to be larger than 2 (Tversky and Kahneman, 1992, Abdellaoui et al., 2007, 2008, 2016). According to Herweg et al. (2010), the result that a bonus is optimal cannot emerge under those levels of loss aversion. Second, the interpretation of loss aversion dramatically changes with the assumed form of the value function (Köbberling and Wakker, 2005). A power utility function, as used in Dittmann et al. (2010), encounters difficulties for the modeling and interpretation of loss aversion. Finally, the intended interpretations of loss aversion and diminishing sensitivity remain valid when the model has an axiomatic foundation. The preference representation of De Meza and Webb (2007) lacks such foundation.

are avoided. This approach also circumvents the violations of first-order stochastic dominance inherent to Kőszegi and Rabin (2006, 2007).

Moreover, it is shown that loss aversion and diminishing sensitivity are, each, sufficient conditions for the emergence of a bonus contract. However, omitting any of them, as it is common practice in applied work, would lead to incorrect predictions about the magnitude and location of incentives of the contract. Finally, the model characterizes first-best contracts. A solution that has been neglected even though it crucially enhances the understanding of incentives, as shown by Corollary 1. I find the surprising result that a bonus can be first-best optimal under reference dependence. Therefore, bonuses not only generate incentives, as most of previous literature emphasizes, but also provide insurance to the agent with reference dependence.

2. General Setup

Consider a principal (she) hiring an agent (he) to produce output on a task.[3] Production output y is a random variable that may take any value in the compact 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.

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

Furthermore, 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.

[3] 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.

To incentivize high effort, the principal offers the agent a take-it-or-leave-it contract including a transfer. Let the transfer be a function $w(y)\colon [\underline{y},\overline{y}] \to \Delta(\mathbb{R}_+)$, where $\Delta = \{p \in \mathbb{R}_+^2 \colon p_1 + p_2 = 1\}$. The principal can thus choose from the set of binary lotteries over positive payments. This definition includes transfers in the traditional sense, i.e. $w(y)\colon [\underline{y},\overline{y}] \to \mathbb{R}_+$, as the subset of degenerate lotteries, i.e $p_1 = 1$ or $p_2 = 0$.[4]

The agent can accept the proposed contract and subsequently choose a level of effort e, or, alternatively, reject it and obtain his reservation utility $\overline{U} \geq 0$.[5] 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. 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 $S''(y) < 0$.

The agent's preferences, and how they affect incentives, is the main topic of investigation in this paper. I assume that there are two ways in which transfers enter the agent's utility. First, they are evaluated according to the absolute level of consumption that they yield. This traditional evaluation of transfers is formalized next.

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

---

[4] Extending in this way the set of transfers is not immaterial. It is key for the proof Theorem 1, which requires finding a lottery improvement with respect to the pointwise optimization solution, and is required for characterizing the optimal solution in Proposition 2 as well as those in previous studies (Herweg et al., 2010, Corgnet and Hernán-González, 2019).

[5] 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.

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

Second, the agent exhibits reference dependence. That is, the consumption utility level generated by a transfer is contrasted to the consumption utility level generated by a reference point, $r > 0$. Transfers yielding consumption utility above the consumption utility of the reference point count as *gains*, while transfers yielding consumption utility below that level count as *losses*. The following assumption formalizes reference dependence.

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

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

Where $\lambda > 1, r > 0$ *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 in Assumption 4. First, V is concave if $w(y) \geq r$, i.e. in the domain of gains, but convex if $w(y) < r$, i.e. in the domain of losses. These shapes of the value function generate risk-averse and risk-seeking attitudes 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 when transfers count as losses, but instead requires contracts that protect him from risk when transfers count as gains.

Second, the restriction $\lambda > 1$ implies that the agent is loss averse. That is, the disutility generated by a loss is, in absolute terms, larger that the utility generated by an equally sized gain. Loss aversion implies that contracts that expose the agent to losses are rejected unless they offer a substantial compensation for such exposure.

The agent's reference point, r, is assumed to be exogenous.[6] This assumption is consistent with the status-quo or current welfare position as reference point (Kahneman and Tversky, 1979). A limitation of assuming that r is exogenous is that the magnitude of transfers or the power of incentives do not affect what counts as a gain or as a loss to the agent. A stringent requirement. In subsequent sections, I will relax that assumption to gain robustness and generalizability.

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

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

The utility function in Eq. (2) consists of three components. The first one represents the agent's absolute evaluation of transfers. I refer to this expression as *expected consumption utility*. The parameter $\phi \in \{0,1\}$ determines whether the agent's preferences incorporate expected consumption utility. As it will be explained later, some theories of risk crucially require this component. The second component in Eq. (2) contains the value function described by Assumption 4. The parameter $\eta \geq 0$ captures the weight given to the (dis)utility resulting from contrasting potential transfers relative to the reference point. When $\eta = 0$ the agent does not evaluate outcomes relative to a reference point. As η becomes larger, more weight is given to the value function from Assumption 4. Relatedly, this component exhibits $\theta_{\mathbb{I}}$ which is an indicator function that takes value $\theta_{\mathbb{I}} = 1$ if the agent is in the domain of gains and $\theta_{\mathbb{I}} = 0$ otherwise. The last component in Eq. (2) captures the standard cost of effort from Assumption 1.

This general representation of reference-dependent preferences is based on the empirical study of Baillon et al. (2020). However, I consider a representation in which the contrasting of outcomes relative to the reference point is not linear, $w(y) - r$, but depends on comparisons of consumption

[6] 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, $u(w(y)) - u(r)$. This way of modeling reference dependence is central to obtaining previous findings in the literature (See for example Proposition 2 (ii)). Besides, it is essential to show that the main result of the paper, a contract with a bonus, is not a mere artifact of diminishing sensitivity.[7]

Due to the conjunction of Assumption 3 and Assumption 4, the utility $U(e, w(y), r)$ can take two shapes: concave or S-shaped. The former shape implies that the agent is globally risk averse, while the latter implies that he is risk seeking in the domain of losses. The formal definition of an S-shaped utility is given next.

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* $\tilde{y} \in [\underline{y}, \overline{y}]$ is an output realization.

It turns out that the shape of $U(e, w(y), r)$ is crucial to implementing incentives. Hence, Lemma 1 provides a sufficient and necessary condition, in terms of the curvatures of v and u, for $U(e, w(y), r)$ to be S-shaped. The proofs of the main results of the paper are relegated to Appendix A.

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

When the concavity of v is stronger than that of u, the agent's risk-seeking attitudes from diminishing sensitivity in losses outweigh the risk-averse attitudes from consumption utility. Hence, the agent's utility, $U(e, w(y), r)$, is convex in that domain. This is the intuition behind the condition in Lemma 1. Additionally, because diminishing sensitivity implies risk aversion in gains, the agent's utility, $U(e, w(y), r)$, is concave in that domain. Consequently, $U(e, w(y), r)$ is, in that case, S-shaped. Instead, when the condition in Lemma 1 does not hold, the convexity implied by $-v$ in the domain of losses is weaker than the concavity of u. The agent's utility is concave

[7] 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).

everywhere. This distinction made by Lemma 1 will be especially useful to implement a solution and to intuitively show how incentives change with the agent's risk attitudes.

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.

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.*

A bonus contract is second-best optimal as it allows the principal to exploit the agent's irrationalities to extract output. For performance levels in which the bonus is not awarded, $y < \hat{y}_s$, the contract specifies transfers that locate the agent in the domain of losses. A prospect that motivates him to exert high effort because that action minimizes the likelihood of obtaining these performance levels (See Assumption 2). Additionally, the risk exposure inflicted by such low payments is tolerated by virtue of the risk-seeking attitudes from diminishing sensitivity (Assumption 4). Motivation is thus achieved by the principal with excessively low monetary incentives.

The bonus is crucial for the contract to be accepted. Its magnitude is such that the agent is transitioned to the domain of gains if the critical threshold, $\hat{y}_s$, is exceeded. This transition is essential, as a contract that consisted exclusively of losses would be rejected by the loss-averse agent. Moreover, the location of the critical output threshold, $\hat{y}_s$, is such that, on expectation, the contract's exposure to losses is offset and the reservation utility is just met.

Theorem 1 (i) and (ii) show that a bonus contract is optimal regardless of whether U is concave or S-shaped. In fact, Theorem 1 (ii), when taken to the extreme case $v' = 1$, implies that diminishing sensitivity is not necessary for a bonus to emerge. An analogous conclusion, that loss aversion is not necessary for a bonus to emerge, is obtained using Theorem 1 (i) along with other results.[8] Therefore, the *qualitative* result that bonuses are optimal cannot be exclusively attributed to loss aversion, as done in previous studies, or to diminishing sensitivity, but emerges because of reference dependence, i.e. the conjunction of both irrationalities arising from the evaluation of outcomes relative to a reference point.

Moreover, a comparison of Theorem 1 (i) and (ii) suggests that the magnitude of the bonus becomes larger under stronger diminishing sensitivity. An agent with stronger risk-seeking attitudes in losses can be exposed to larger degrees of risk by obtaining lower payments in that domain. These lower payments given when $\hat{y}_s$ is approached from the left increase the magnitude of the bonus. The applied significance of this *quantitative* result is that omitting diminishing sensitivity, as it is common practice in theoretical and empirical studies, can lead to incorrect

[8] Specifically, Theorem (i) together with Theorem (iii) and Corollary 1 imply that a bonus would emerge as it provides insurance to the agent with strong diminishing sensitivity and generates strong punishments that incentivize high effort.

predictions. For example, bonus contracts that might not be powerful enough to motivate the agent, or that are not sufficiently profitable to the principal.

Theorem 1 also characterizes the optimal first-best contract. A lump-sum bonus provides full insurance when $U(e,w,r)$ is S-shaped. In that case, the agent is risk seeking in losses and thus willing to be fully exposed to risk in $y < \hat{y}_f$. Therefore, a strategy of offering excessively low transfers is feasible to the principal. Loss aversion in turn implies that, to ensure that the contract is accepted, transfers must eventually offset the losses included in the contract. This is achieved with the lump-sum bonus. Instead, when the agent is globally risk averse, i.e. when $U(e,w,r)$ is concave, the considerable risk exposure of the lump-sum bonus is not tolerated. Full insurance is then achieved with a constant transfer that protects the agent from any risk and locates him in the domain of gains.

To better understand the incentives imparted by the second-best optimal contract from Theorem 1, its shape is compared to that of the first-best optimal 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.[9] In the segment $y < \hat{y}_f$, both first- and second-best contracts pay the lowest possible transfer, $w_s^*(y) = w_f^*(y) = 0$, locating the agent in the domain of losses. In that segment, incentives are solely imparted by virtue of loss aversion. Furthermore, the bonus and the fact that

[9] 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 $\hat{y}_e > \hat{y}_s$ implies that w_s^* imparts punishments in the domain of gains.

$\hat{y}_s \geq \hat{y}_f$ generate sizeable punishments in $y \in (\hat{y}_{f,}\hat{y}_s)$. The second-best contract includes punishments for low performance when the agent's irrationalities become, on their own, insufficient for motivating the agent. Finally, that the transfers of the second-best contract increase in performance in $y > \hat{y}_s$, generates incentives in the standard way: rewards are given for high performance levels and punishments for intermediate performance levels.

The following corollary presents comparative statics that further elucidate the influence of reference dependence on incentive design. 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 $U\big(e, w_f^*(\tilde{y}), r\big)$ is S-shaped or at a lower $\hat{y}_s$ if $U\big(e, w_f^*(\tilde{y}), r\big)$ is concave.

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

A higher reference point implies that the bonus needs to be larger to ensure that the agent is transitioned from the domain of losses to the domain of gains. This larger bonus generates changes on the location of the critical threshold, $\hat{y}_s$, that crucially depend on the strength of diminishing sensitivity.[10] Moreover, a higher level of loss aversion implies that the agent experiences more disutility when exposed to losses. To keep the contract attractive, the bonus is awarded at lower output levels.

4. Old and New Results

In this section, I show that Theorem 1 is general enough to explain previous findings. Moreover, I demonstrate that Theorem 1 provides novel and important findings that fill prominent gaps in the literature.

[10] Specifically, the prospect of higher bonuses enables the principal to expose the agent to losses for a larger output segment if that exposure to risk is tolerated. Therefore, for sufficiently strong diminishing sensitivity, the threshold at which the bonus is awarded, $\hat{y}_s$, is higher. If the agent cannot tolerate that risk exposure, the bonus is given at a lower $\hat{y}_s$. This strategy is more cost-effective than further raising the size of the bonus.

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 a fixed transfer that fully insures the risk averse agent from risk.

4.2 Prospect Theory Preferences

Next, I consider reference-dependent preferences within the framework of prospect theory (Tversky and Kahneman, 1992).[11] According to that theory, the agent evaluates outcomes using only the value function from Assumption 4. Therefore, the preferences in Eq. (2) are modified with the restrictions $\phi = 0$, $\eta = 1$, and $u' = 1$.[12]

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

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

[12] 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 (2021).

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 from Theorem 1 captures the solution found by (Dittmann et al., 2010). The optimal contract exhibits a bonus that is awarded when a critical threshold output level $\hat{y}_{ps}$ is met. If output falls short of that critical output level, the lowest possible pay is given. For output levels higher than $\hat{y}_{ps}$ the contract's transfers increase with performance.

Corollary 4 extends the findings of Ditmann et al. (2010) in important ways. First, it shows that the first-best contract is a lump-sum bonus. This yields the relevant conclusion that the optimality of bonuses is not restricted to a setting of moral hazard but also emerges to provide full insurance to the agent. Second, the incentives of the bonus contract in Dittmann et al. (2010) are better understood when the first-best optimal contract, given in Corollary 4 (i), is also available.[13] Third, the characterization of the second-best contract given in Corollary 4 (ii), together with Corollary 5 and Corollary 6, generalize Dittmann et al. (2010). That is because in my setting, the bonus contract emerges without imposing a functional form of v, a non-trivial assumption (See footnote 2), and does not require an exogenous reference point r.

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

[13] Corollary 1 (i) and the explanation laid out thereafter still apply and provide an explanation of the incentives included in the second-best contract.

relative to a reference point applies only to losses. To the best of my knowledge no axiomatized theory of risk provides such 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 can be adapted to obtain the preference representation and the results of De Meza and Webb (2007).

4.3 Prospect Theory Preferences with Endogenous Reference Points

So far, r has been assumed to be exogenous. In the remainder of the paper, I show how the results from Theorem 1 can be adapted to account for endogenous reference points. Formally, the assumption $r > 0$ (Assumption 4) is, from here onward, replaced by $r: \left[\underline{y}, \overline{y}\right] \to \mathbb{R}_+$.

In this subsection, I focus on rules consistent with prospect theory and, for brevity of exposition, on the moral hazard case.

Salience Rules

I first consider salience-based reference point rules. These are rules resulting from unconscious comparisons made by the decision-maker between the outcomes or the probabilities included in the contract. One such rule is the *max-min* (Hershey and Schoemaker, 1985). There, the agent is cautious and takes as reference point the maximum value from a set consisting of the minimum outcome of each alternative. As an example, suppose that the agent chooses between two contracts, namely $w_1 := (0.5{:}\,200, 0.5{:}\,0)$ and $w_2 := 100$. Under the considered rule, he sets as reference point $r = \max\{\min\{w_1\}, \min\{w_2\}\} = \max\{0{,}100\}$. Baillon et al., (2020) found substantial empirical support for this way of forming reference points.

The following corollary presents the solution to the moral hazard problem under the max-min rule. The resulting contract is a bonus. More importantly, it turns out that the agent's reference point is the contract itself, the reference point rule assumed in Koszegi and Rabin (2006)'s choice acclimating equilibrium.

Corollary 5. *Under A1-A4,* $\phi = 0$, $\eta = 1$, $u' = 1$ and *the max-min rule, the agent's reference point is* $r = w_s^*(y)$. *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 rejecting the contract and exerting high or low effort. There exist thus four candidates for reference point, corresponding to the potential consequences implied by each choice. When the contract is rejected, the welfare level obtained by the agent is equal to $\overline{U}$. Since the participation constraint binds, this is also the expected utility implied by the optimal contract, $w_s^*(y)$. Because the incentive compatibility constraint also binds, this welfare level must be higher than that implied by $\min\{w_s^*(y)\}$, the minimum possible payment when the contract is accepted. Otherwise, the agent would not be incentivized to exert effort by means of punishments. Thus, the max-min rule gives that $r = w_s^*(y)$.

Corollary 5 provides an alternative explanation to behavior that has been previously attributed to stochastic and expectations-based reference points (Abeler et al., 2010, Crawford and Meng, 2011, Gill and Prowse, 2012). Accordingly, the seeming tendency of individuals to supply labor according to earnings' expectations is in fact guided by the salience of the outside option. The theoretical significance of Corollary 5 is also appealing. As it will be later shown, prospect theory is incompatible with expectations-based and the contract itself as reference points. I present a way to incorporate these reference points within the machinery of prospect theory without incurring in absurd implications.

Two additional salience-based reference points are studied: 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 $r = \min\{\max\{w_1\}, \max\{w_2\}\} = \min\{200,100\}$. On the other hand, the $w(y)$ *at max P* states that the output level realizing with the highest probability becomes the agent's reference point. In the considered example, the agent's reference point happens to be again 100.

The solutions to the principal's problem under these salience-based reference points are described by Corollary B.1. and Corollary B.2. in Appendix B. The results therein show that under the max-min rule the contract itself also emerges as reference point.

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, individuals exhibit higher performance in cognitive and/or physical tasks is when a high rather than a low goal is set (Heath et al., 1999). A result explained by an aversion to miss a goal, i.e. loss aversion, and a willingness to exert more effort as the goal is approached, i.e. diminishing sensitivity.[14]

In this section, I assume that the agent incorporates a goal chosen by the principal as reference point. This is consistent with abundant evidence showing that goals set by other party act as reference points (Heath et al., 1999, Locke and Latham 2002, Larrick et al., 2009), also in settings in which a principal sets the goal. (Corgnet et al. 2015, Corgnet et al., 2018, van Lent and Souverijn, 2020).

Let the goal be a production level, $g \in [\underline{y}, \overline{y}]$. Importantly, that production goal may not need to coincide with the principal's nor with the agent's expectations.[15] This difference makes goals consistent with prospect theory but not necessarily with other theories of risk with reference-dependence. A claim that will become evident in the next section.

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

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

[14] 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).

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

The goal divides the *output space* into gains and losses. Accordingly, transfers are contrasted to the payment given when the goal is just met, $w(g)$.

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 $(w(y), g)$ presented to the agent before he exerts effort. The steps followed in Theorem 1 along with additional assumptions and elaborations provide the following solution to this modified principal-agent 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 $\lim_{x \to 0} v'(x) = +\infty$, the optimal solution consists of the principal setting a challenging but, on expectation, attainable goal. Such production target is accompanied by a contract that awards a bonus if the goal is met or surpassed. The bonus has the property that it increases with the magnitude of the goal.

According to Worldatwork (2018) more than 75% of American firms award bonuses when an individual or organizational goal is achieved. The result in Proposition 1 provides a rationale to this widespread organizational practice.

Proposition 1 is, to the best of my knowledge, the first to fully characterize optimal contracts in a setting of moral hazard when goals are adopted as reference points. In related work, Corgnet et al.

(2018) also solve for an optimal contract under the considered reference point rule, but they restrict their analysis to linear contracts. Proposition 1 shows that this is not an innocuous assumption.[16]

4.4 Disappointment Aversion

A considerable bulk of the literature is sympathetic with the idea that the expected value of an alternative gains the status of reference point (Abeler et al., 2011, Terzi et al. 2015, Sprenger, 2015, Gneezy et al., 2016). However, imposing this rule is incompatible with prospect theory. To see how, consider an agent with preferences described by Eq. (2) along with the restrictions $\phi = 0$, $\eta = 1$ and $u' = 1$. When this agent is offered a fixed-wage contract $w_k := k$, his utility, under expectations-based reference points, becomes $(e, w_k, \mathbb{E}(w_k)) = 0$ for any $k > 0$. An absurd implication!

The merit of disappointment models is to include expectations-based reference points without incurring in this problem. They do so while maintaining the descriptively relevant phenomena of loss aversion and diminishing sensitivity.[17] This is mainly achieved by requiring expected consumption utility. In other words, the restriction $\phi = 1$ applies in those models.

Existing models of disappointment differ in the assumed reference point. In Bell (1985), the reference point is the expected value of a risky alternative, while in Loomes and Sugden (1986) is the expected consumption utility.[18] The following corollary shows how Theorem 1 can be adapted

[16] 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. That scale is expressed in monetary terms and, in contrast to previous work, does not combine the output and monetary domains.

[17] 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.

[18] Axiomatic foundations for disappointment models are given in Bell (1985) and Gul (1991).

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 $\overline{\overline{w}} := \int_{\underline{y}}^{\overline{y}} w(y) f(y|e) dy$. *Under A1-A4,* $\phi = 1$, *and* $r = \overline{\overline{w}}$ *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* $w_s^*(y) = \overline{\overline{w}} + f'\left(\frac{1}{\eta}\left(\frac{1}{\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right)} - 1\right)\right)$ *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* $w_f^*(y) = \overline{\overline{w}} + f'\left(\frac{1}{\eta}\left(\frac{1}{\mu} - 1\right)\right)$ *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 $v' = 1$, the model of Bell (1985) is relevant; the evaluation of transfers is performed relative to the expected vale of the contract. Instead, when $u' = 1$, the model of Loomes and Sugden (1986) becomes relevant; outcomes of the contract are evaluated relative to the agent's expected consumption utility. Corollary 6 shows that in both cases a contract with a bonus incentivizes the agent to exert high effort. The intuition of this result follows from the discussion of Theorem 1 (i) and (ii) given in Section 3.

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 implies that slight contract modifications generate changes in the magnitude and/or the location of the bonus. One such modification is implied by 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 magnitude and its location on the output space is distinct.

Remarkably, 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 preferences according to Gul (1991)'s disappointment model. There, the agent's reference point is his certainty equivalent. As in Corollary 6, I find that the optimal contract under that reference point rule exhibits a bonus. However, as it was patent in the preceding paragraphs, such bonus can have a different location and magnitude as compared to the contract in Corollary 6 due to the different reference point assumption. 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 approach to capture reference dependence. It starkly differs from previous representations because the reference point is stochastic.

To exemplify this reference point rule, suppose that the worker gets $w_3 := (0.5: 200, 0.5: 100)$. He adopts first $r = 200$. In that case, obtaining 100 feels like a loss that realizes with 25% probability, found from the product $0.5 * 0.5$, while obtaining 200 is a neutral outcome that also realizes with 25% probability. Subsequently, $r = 100$ is adopted. Then, obtaining 200 feels like a gain realizing with 25% chance and obtaining 100 is a neutral outcome. Thus, under the considered reference point rule, the contract w_3 is perceived as a 25% chance to win 100, a 25% chance to lose 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

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

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 and additional elaborations. The resulting optimal contract can acutely differ from those presented in Theorem 1 because it can be first-best and second-best optimal to offer a stochastic contract, i.e. a lottery.

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 and* $-1 \leq v'(0)(1 - \lambda)$*, 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* $L_d^* = (\rho{:}\, w_s^*, 1 - \rho{:}\, 0)$ *where* $\rho \in (0,1)$ *and* w_s^* *is the contract described in i) or ii), depending on the agent's utility shape, if* $-1 > v'(0)(1 - \lambda)$.

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

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

vi) *pays the lottery* $L_d^* = (\rho: w_s^*, 1 - \rho: 0)$ *where* $\rho \in (0,1)$ *and* w_f^* *is the contract described in v) or vi), depending on the agent's utility shape, if* $-1 > v'(0)(1-\lambda)$.

Stochastic contracts can be understood as the principal committing ex-ante to a probability that the agent's performance signals will be ignored (Herweg et al. 2010). Alternatively, they can be interpreted as the principal implementing incentive schemes that uniformly increase the agent's exposure to risk by some probability (Gonzalez-Jimenez, 2020).[19]

Participation can be ensured by stochastic contracts. Suppose that $w = 0$ is paid for all $y < \overline{y}$. Under the considered reference point rule, this strategy can generate larger disutility from losses as compared to other rules because the largest payment, $w = w(\overline{y})$ is also adopted as reference point. So, for r, representing a previously studied reference point, then $-\lambda\, v\left(u(w(\overline{y}))\,)\right) > -\lambda\, v(u(r) - u(w))$. To avoid this larger disutility, that could potentially lead to a rejection of the contract, a stochastic payment L_s^* including a non-zero probability of ending-up in gains is implemented.

Motivation under stochastic contracts is obtained at the expense of the agent's irrationalities. 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 sufficiently strong diminishing sensitivity

[19] For example, a fixed-wage can be transformed into a cost-equivalent bonus contract that pays a high amount with 50% chance. Similarly, this bonus contract can be modified into a bonus contract that pays a higher amount with lower chance.

implies that the agent is willing to accept the stochastic payment. This is the rationale behind Proposition 2 (i).[20]

Another reason behind the optimality of stochastic contracts is that, under the considered reference point rule, first-order stochastic dominance can be violated (Masatlioglu and Raymond, 2016). In that case, the principal can exploit the agent's irrationality by offering a stochastically dominated lottery. This is the result captured by Proposition 2 (iii). Instead, if the agent respects first-order stochastic dominance, he could be better off with a standard bonus contract as in Proposition 2 (ii).

Herweg et al. (2010) find that, under the considered preferences, the second-best contract is stochastic for high levels of loss aversion, $\lambda > 2$. More emphasis is however given to the result that the optimal contract is a lump-sum bonus for low loss aversion levels, i.e. $\lambda < 2$. These results can be captured by Proposition 2 (i)-(iii) under a slightly different value function specification and additional parametric restrictions.[21]

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* $b_s > 0$ *awarded if* $y = \overline{y}$

5. Extensions

This section extends the results presented in Section 3 to further gain generalizability and highlight the significance of Theorem 1. Here, I show that the results in that theorem are valid beyond a set of assumptions made on the principal's preferences and knowledge.

In this section, I return to the assumption that the agent's reference point is exogenous, i.e. $r > 0$, and lay focus on the moral hazard case.

[20] Specifically, stochastic contracts also ensure motivation in a standard way. The large monetary incentives given in $y = \overline{y}$ motivate the agent in a more traditional way.

[21] Specifically, let $x := w(y) - w(g)$. I 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). Moreover, it is assumed that losses enter the agent's utility in a different way. Instead of having $-u(-x)$ for $x < 0$, as assumed throughout this paper, such a loss enters the utility as $u(-x)$.

5.1 Principal with Reference Dependent Preferences

A discernible extension considers a setting in which the principal also exhibits reference-dependence. Formally, let the principal's preferences be characterized by

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

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

According to Eq. (7), the principal is loss averse. This irrationality applies to both her benefit and cost functions. An assumption consistent with the notion, also present in Assumption 4, that these biases apply to monetary outcomes.[22] Furthermore, she does not suffer from diminishing sensitivity. This simplifying 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 comparable to those presented in Theorem 1. The only difference appears for the case in which output attains intermediate levels. 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* $\hat{y}_p \in \left[\underline{y}, \overline{y}\right]$ *such that the second-best contract,* $w_s^*(y)$*:*

[22] Another possible representation of reference dependence is

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

This representation is 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 in that case the principal's problem is unchanged.

i) Is identical to the contract presented in Theorem 1 (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.*

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

For output levels ensuring $S(y) < r_p + w(y)$, the principal's loss aversion can affect optimal contracting. When output is sufficiently low so that $y < \hat{y}_s$ also holds, the loss-averse principal transfers most of the risk to the agent. If $U(e, w_s^*(\tilde{y}), r)$ is S-shaped, transfers are set as low as possible because the agent is risk seeking. Instead, if $U(e, w_s^*(\tilde{y}), r)$ is concave, transfers are non-zero but are low enough as to locate the agent in the domain of losses. This shape of the optimal contracts presented in Proposition 3 (ii) and (iii) are exactly like those in Theorem 1.

For higher output levels, so that $y \geq \hat{y}_s$ holds but $S(y) < r_p + w(y)$ is also true, the principal needs to offer some insurance to the risk-averse agent. However, the principal's loss aversion implies that more risk will be transferred to the agent as compared to the solution in Theorem 1. This is achieved by offering lower-powered incentives in the segment $y \in (\hat{y}_s, \hat{y}_p)$. This immediately implies the existence of a kink in the incentives scheme around $\hat{y}_p$, the point at which the principal herself transitions from losses to gains. After that point, she offers incentives that are as high-powered as in Theorem 1. This kink around $\hat{y}_p$ and the lower-powered incentives in $y \in (\hat{y}_s, \hat{y}_p)$ constitute a slight modification to Theorem 1. This modification is reflected in Proposition 3 (i) and (ii).

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, in the framework considered in this paper, this assumption becomes more demanding as she not only needs to know the agent's utility curvature but also his parameter of loss aversion. This extension relaxes the assumption that the principal exactly knows the agent's risk preferences.

Suppose that the principal is perfectly informed about the agent's utility but that she does not know his degree of loss aversion. For simplicity, assume that she can contract with agents with either high or low degrees of loss aversion. Formally, let $\lambda_i \in \{\lambda_L, \lambda_H\}$ where $\lambda_H > \lambda_L > 1$. Contracting with an agent with λ_H occurs with probability ω, while contracting with an agent with λ_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 λ_i, which is private information to the agent. Second, the principal offers a menu of contracts. 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 contract selected by the agent.

The principal's objective is to design contracts that enable her to screen agents to ensure participation and motivation. The following proposition shows that the optimal menu of contracts consists of two bonus contracts enhanced with informational rents.

Proposition 4. *Under A1-A4 and that λ_i is unknown to the principal, the optimal menu of contracts is the tuple $\{w_s^*(y)^H, w_s^*(y)^L\}$ such that*

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

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

Where $\hat{y}_H \in \left(\underline{y}, \overline{y}\right)$ *is the critical threshold specified in* $w_s^*(y)^H$.

According to Theorem 1, bonus contracts ensure that agents with reference-dependent preferences participate and exert high effort. Furthermore, Corollary 2 shows that the bonus feature of the contract appears regardless of the agent's degree of loss aversion. That is because the location of the bonus will be shifted to account for any level of loss aversion. Therefore, the optimal menu consists of bonus contracts.

Screening is guaranteed by complementing the contract targeting agents with lower loss aversion, $w_s^*(y)^L$, with an informational rent. The goal of the rent is to discourage these agents from mimicking high loss-averse agents. To achieve that, its magnitude is such that they become exactly indifferent between engaging in a strategy of mimicking or not doing so. On the other hand, agents with high loss aversion are not willing to engage in a strategy of mimicking. As suggested by Corollary 2, choosing $w_s^*(y)^L$ instead of $w_s^*(y)^H$ would considerably expose these agents to considerably losses and generate 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 the solution to the optimal 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*.

Crawford, V. P., & Meng, J. (2011). New York City cab drivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income. *American Economic Review*, *101*(5), 1912-32.

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. (2020) Incentive Contracts when Agents Distort Probabilities. Unpublished.

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

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

Gill, D., & Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, *102*(1), 469-503.

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.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, *92*(5), 1644-1655.

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.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist*, *57*(9), 705.

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.

Salanié, B. (2003). Testing contract theory. *CESifo Economic Studies*, *49*(3), 461-477.

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.

van Lent, M., & Souverijn, M. (2020). Goal setting and raising the bar: A field experiment. *Journal of Behavioral and Experimental Economics*, *87*, 101570.

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, whether $U(e, w(y), r)$ is S-shaped or concave is determined by the shape adopted by that function in the domain of losses. We consider realizations that locate the agent in that domain. Hence, for each $\tilde{y} \in [\underline{y}, \overline{y}]$ such that $w(\tilde{y}) < r$. Eq. (2) becomes:

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

The second derivative of (A1) with respect to $w(\tilde{y})$ is:

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

A sufficient and necessary condition for $U(e, w(\tilde{y}), r)$ to be convex in the domain of losses is $U''(e, w(\tilde{y}), r) \geq 0$ for each $\tilde{y} \in [\underline{y}, \overline{y}]$ such that $w(\tilde{y}) < r$. That condition can be rewritten using Eq. (A2) as:

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

■

Theorem 1

The proof follows four parts. In part i), the Lagrangian and first-order conditions are presented. Part ii) presents the existence and uniqueness of an improvement to the solution from the first-order conditions when $U(e_H, w(\tilde{y}), r)$ is S-shaped. This improvement is incorpoirated in the optimal contract. Part iii) states the properties of the second-best optimal contract for the two relevant cases, that is when $U(e_H, w(\tilde{y}), r)$ is concave and S-shaped. Finally, iv) presents the 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:

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

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

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

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 $(A5)$ with respect to y gives

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

if $\theta_{\mathbb{I}} = 1$. Since $u'' < 0$, $v'' < 0$, and $\frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right) \leq 0$ (Assumption 2), Eq. $(A8)$ shows that $\frac{dw_s^F(y)}{dy} \geq 0$. Thus, $w_s^F(y)$ is nondecreasing in performance in the domain of gains.

The derivative of $(A5)$ with respect to y gives

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

if $\theta_{\mathbb{I}} = 1$. The numerator of the right-hand side of $(A9)$ is negative due to Assumption 2, $\frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right) \leq 0$. Moreover, 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, Lemma 1 implies that $\frac{dw_s^F(y)}{dy} \leq 0$. An undesirable property, as it incentivizes the agent to destroy production. An alternative solution must be found.

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 each $\tilde{y} \in [\underline{y}, \overline{y}]$, the solutions $w_s^F(y)$ given by $(A6)$ and $(A7)$ are necessary and sufficient to solve the maximization problem in Eq. (3). However, these solutions are not necessarily optimal when $U(e_H, w(\tilde{y}), r)$ is S-shaped. Next, it is shown that in such case the principal is better off offering a lottery $L := (p: r, 1 - p: 0)$.

Since the expression $\phi u(w_s^F(y)) - \lambda v\left(u(r) - u(w_s^F(y))\right)$ increases in $w_s^F(y)$, there must exist a number $p \in (0,1)$ such that:

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

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. Eq. $(A10)$ and the concavity of v, imply:

$$\begin{aligned}\phi\left(pu(r) - u\left(w_s^F(y)\right)\right) &= (1 - p)\lambda v\left(u(r)\right) - \lambda\eta v\left(u(r) - u\left(w_s^F(y)\right)\right) \\ &\leq \lambda v\left((1 - p)u(r)\right) - \lambda\eta v\left(u(r) - u\left(w_s^F(y)\right)\right). \qquad (A11)\end{aligned}$$

Let first $\phi = 0$. The inequality in Eq. $(A11)$ becomes

$$\lambda v\left((1 - p)u(r)\right) \geq \lambda\eta v\left(u(r) - u\left(w_s^F(y)\right)\right). \qquad (A12)$$

Since $v' > 0$, then Eq. $(A12)$ gives $w_s^F(y) \geq pr$. Consequently, L is cost-effective for the principal.

To show that $w_s^F(y) \geq pr$ also holds under $\phi = 1$, I proceed by contradiction. Suppose that $w_s^F(y) < pr$. Multiplying both sides of Eq. $(A12)$ by minus one, and adding $u\left(w_{sb}^F(y)\right)$ to both sides we obtain:

$$u\big(w_s^F(y)\big) - \lambda v\big((1-p)u(r)\big) \le u\big(w_s^F(y)\big) - \lambda\eta v\left(u(r) - u\big(w_s^F(y)\big)\right). \qquad (A13)$$

Eq. $(A13)$ shows that Eq. $(A11)$ might not hold under $\phi = 1$ if $w_s^F(y) < pr \Rightarrow u\big(w_s^F(y)\big) < pu(r)$. This contradicts the concavity of v, the Assumption behind Eq. $(A11)$. Hence, it must be that $w_s^F(y) > pr$. It is straightforward to check that such condition is consistent with $(A11)$ and the assumptions of the model. Hence, L is more cost-effective for the principal than the solution implied by $(A7)$.

In the domain of gains, the principal cannot implement a similar strategy. Let $L_G := (q{:}\,G, 1 - q{:}\,r)$, where $G > w_s^F(y) \ge r$. Since $\phi u(w_s^F(y)) + \lambda v(u(w_s^F(y)) - u(r))$ increases in $w_s^F(y)$, there exists a number $q \in (0,1)$ such that:

$$\begin{aligned}\phi u(w_s^F(y)) &+ \eta v(u(w_s^F(y)) - u(r)) \\ &= \phi(qu(G) + (1-q)u(r)) + q\eta v\big(u(G) - u(r)\big). \qquad (A14)\end{aligned}$$

Therefore, replacing $w_s^F(y)$ from $(A3)$ by L_G with a probability q satisfying $(A14)$ leaves the agent's participation and incentive compatibility constraints unchanged.

Equation $(A13)$ and the concavity of v in the domain of gains, imply:

$$\begin{aligned}\phi\left(qu(G) + (1-q)u(r) - u\left(w_{sb}^F(y)\right)\right) &= \eta v\left(u\big(w_s^F(y)\big) - u(r)\right) - \eta q v\big(u(G) - u(r)\big) \qquad (A15) \\ &\ge \eta v\left(u\big(w_s^F(y)\big) - u(r)\right) - \eta v\big(qu(G) - qu(r)\big).\end{aligned}$$

Let first $\phi = 0$. Eq. $(A14)$ becomes

$$\eta v\big(qu(G) - qu(r)\big) \ge \eta v\left(u\big(w_s^F(y)\big) - u(r)\right), \qquad (A16)$$

leading to

$$\begin{aligned}u\left(w_{sb}^F(y)\right) &\le qu(G) + (1-q)u(r) \qquad (A17) \\ &< u(qG + (1-q)r).\end{aligned}$$

Therefore, L_G is not cost-effective for the principal as it generates higher expected costs: $qG + (1-q)r > w_s^F(y)$.

To show that L_G is not cost-effective for the principal under $\phi = 1$ we proceed by contradiction. Suppose that $w_s^F(y) \ge qG + (1-q)r$. Multiplying both sides of Eq. $(A16)$ by minus one, and adding $u(w_s^F(y))$ to both sides we obtain:

$$u\big(w_s^F(y)\big) - \eta v\big(qu(G) - qu(r)\big) \le u\big(w_s^F(y)\big) - \eta v\left(u\big(w_s^F(y)\big) - u(r)\right). \qquad (A18)$$

Eq. $(A18)$ shows that Eq. $(A15)$ cannot hold under $\phi = 1$ since $w_s^F(y) \ge qG + (1-q)r \Rightarrow u\big(w_s^F(y)\big) > qu(G) + (1-q)u(r)$. This contradicts the concavity of v, the assumption behind Eq. $(A15)$. Hence, it must be that $qG + (1-q)r > w_s^F(y)$. It is straightforward to check that such condition is consistent with $(A15)$ and the assumptions of the model. Therefore, L_G is less cost-effective for the principal than the solution implied by $(A6)$.

A similar rationale can be followed to show that $L_M := (q: G, 1-q: L)$, where $G > r > L > 0$, cannot improve upon the solution given in $(A6)$. When receiving that lottery, the agent derives utility

$$\phi(qu(G) - (1-q)\lambda u(r)) + qv\big(u(G) - u(r)\big) - (1-q)\eta\lambda\ v\big(u(G) - u(r)\big). \qquad (A19)$$

Comparison of Eqs. $(A14)$ and $(A19)$ shows that $U(e, L_M, r) < U(e, w_S^F(y), r)$. So, paying L_M instead of $w_S^F(y)$ in $(A6)$ does not keep the agent's incentive compatibility and/or the participation constrains unchanged. Also, $\mathbb{E}(L_M) > \mathbb{E}(L)$. So, not only L_M yields lower incentives, but it is also less cost-efficient to the principal.

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

$$U(e_H, L, r) = \left(\frac{\overline{L}}{r}\right)\phi u(r) - \left(1 - \frac{\overline{L}}{r}\right)\lambda\eta v\big(u(r)\big) - c, \qquad (A20)$$

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:

$$\frac{1}{\frac{\phi u(r) + \lambda\eta v\big(u(r)\big)}{r}} = \nu + \gamma\left(1 - \frac{f(\hat{y}_s|e_L)}{f(\hat{y}_s|e_H)}\right). \qquad (A21)$$

The existence and uniqueness of $\hat{y}_s$ is guaranteed by the fact that the left-hand side of $(A21)$ is positive and constant in y, while the right-hand side of that equation increases with y in $[0, \infty)$ (Assumption 2).

Eq. $(A21)$ implies that offering L with $p \in (0,1)$ either does not incentivize high effort or leads to non-binding constraints in $y \in [\underline{y}, \overline{y}] \backslash \{\hat{y}_s\}$. That is because if $y < \hat{y}_s$, the right-hand side of Eq. $(A21)$ 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. $(A20)$ shows that a reduction of $\overline{L}$ does not change the agent's marginal utility. Hence, is optimal to set $p = 0$ and $w_S^*(y) = 0$. 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 $\overline{L}$, it is optimal to set $p = 1$ which gives $w_S^*(y) = r$.

iii) Properties and definition of the optimal contract

Let $U(e_H, w(\tilde{y}), r)$ be S-shaped for each $\tilde{y} \in [\underline{y}, \overline{y}]$. Since the agent is loss averse, a contract consisting of paying $w_s^*(y) = 0$ everywhere, which is equivalent to $\hat{y}_s = \overline{y}$, cannot be a solution as it induces considerable disutility and would lead the agent to reject the contract. Also, the solution given by Eq. $(A6)$, which is equivalent to $\hat{y}_s = \underline{y}$, cannot be optimal on its own. In that case, it is profitable to the principal to deviate from that solution by offering $w_s^*(y) = 0$ at the low-end of the output space. Also, Eq. $(A21)$ shows that such deviation is not only more cost-effective, but also motivates the loss-averse agent to exert high effort to avoid the disutility from experiencing losses. Hence, it must be that $\hat{y}_s \in \left(\underline{y}, \overline{y}\right)$.

Thus, the optimal contract is given by:

$$w_s^*(y) = \begin{cases} 0 & if\ y < \hat{y}_s, \\ w_s^F(y) \text{ satisfying } (A6) & if\ y \geq \hat{y}_s. \end{cases} \qquad (A22)$$

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. The transfer $w_f^F(y)$ satisfying $(A7)$ cannot be a solution on its own, as leads to considerable disutility and would be rejected. 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 must combine the first-order conditions $(A6)$ and $(A7)$. The transition from one payment scheme to the other is given by the output level $\hat{y}_s \in (\underline{y}, \overline{y})$ satisfying:

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

The existence and uniqueness of $\hat{y}_s$ is guaranteed by the fact that the left-hand side of Eq. $(A23)$ can be negative if $\hat{y}_s = \overline{y}$ and sufficiently large η, increases as $\hat{y}_s$ decreases, and is positive if $\hat{y}_s = \underline{y}$.

In that case, the optimal incentive scheme is given by:

$$w_s^*(y) = \begin{cases} w_s^F(y) & satisfying\ (A6)\ \ if\ y \geq \hat{y}_s\,, \\ w_s^F(y) & satisfying\ (A7)\ \ if\ y < \hat{y}_s. \end{cases} \qquad (A24)$$

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

$$\frac{1}{u'\left(w_f^F(y)\right)\left(\phi + \eta v'\left(u\left(w_f^F(y)\right)\right) - u(r)\right)\right)} = \mu, \qquad (A25)$$

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

$$\frac{1}{u'\left(w_f^F(y)\right)\left(\phi + \lambda\eta v'\left(u(r) - u\left(w_f^F(y)\right)\right)\right)} = \mu, \qquad (A26)$$

if $\theta_{\mathbb{I}} = 0$. It is evident from $(A25)$ and $(A26)$ 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 each $\tilde{y} \in [\underline{y}, \overline{y}]$, the principal is better off offering a lottery $L = (p: r, 1 - p: 0)$ instead of $w_f^F(y)$ satisfying $(A26)$. That lottery can be offered to the agent with a $p \in (0,1)$ that satisfies:

$$\phi u\left(w_f^F(y)\right) - \lambda v\left(u(r) - u\left(w_f^F(y)\right)\right) = \phi p u(r) - (1-p)\lambda v\left(u(r)\right). \qquad (A27)$$

The existence of the $p \in (0,1)$ satisfying Eq. $(A27)$ 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 $\phi p u(r) - (1-p)\lambda v\left(u(r)\right)$ is constant. Therefore, replacing $w_f^F(y)$ from $(A26)$ by L leaves the agent's participation constraint unchanged.

Equation $(A27)$ and the concavity of v for any output realization $\tilde{y}$ such that $w(\tilde{y}) < r$, imply:

$$\begin{aligned}\phi\left(pu(r) - u\left(w_f^F(y)\right)\right) &= (1-p)\lambda v\left(u(r)\right) - \lambda\eta v\left(u(r) - u\left(w_f^F(y)\right)\right) \\ &\leq \lambda v\left((1-p)u(r)\right) - \lambda\eta v\left(u(r) - u\left(w_f^F(y)\right)\right) \qquad (A28)\end{aligned}$$

Under $\phi = 0$, the inequality above immediately implies $w_f^F(y) > pr$. Under $\phi = 1$, assuming that $w_f^F(y) < pr$ would be inconsistent with $(A28)$ and thus contradict the model's assumptions.

Hence, it must be that $w_f^F(y) > pr$. As such, L is more cost-effective for the principal than the candidate solution from Eq. $(A26)$.

When L is offered, the Lagrangian multiplier, μ, can be large enough to ensure:

$$\frac{1}{\frac{\phi u(r) + \lambda\eta v\big(u(r)\big)}{r}} = \mu. \qquad (A29)$$

However, if μ is smaller, so that the left-hand side of $(A29)$ 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. $(A20)$ shows that a reduction of $\overline{L}$ does not change the agent's marginal utility. Thus, is optimal to set $p = 0$ and $w_f^*(y) = 0$. In contrast, if μ is large enough, such that the left-hand side of $(A29)$ is smaller than the right-hand side of that equation, then $\overline{L}$, should be increased. Again, since $U(e_H, L, r)$ is linear in $\overline{L}$, as shown by Eq. $(A20)$, it is optimal to set $p = 1$ which gives $w_f^*(y) = r$.

Therefore, when $U(e_H, w(\tilde{y}), r)$ is S-shaped, 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 $(A25)$. 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 loss-averse agent would reject it. Next, suppose that only $w_f^*(y) = w_f^F(y)$ is given. 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, 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. $(A25)$ is defined next. Let $\hat{y}_f \in (\underline{y}, \overline{y})$ be the output level satisfying

$$\phi\int_{\hat{y}_f}^{\overline{y}} u\big(w_f^F(y)\big) f(y|e_H)\,\mathrm{d}y + \eta\int_{\hat{y}_f}^{\overline{y}} v\Big(u\big(w_f^F(y)\big) - u(r)\Big) f(y|e_H)\,\mathrm{d}y + \phi\int_{\underline{y}}^{\hat{y}_f} u(r) f(y|e_H)\mathrm{d}y$$
$$- \lambda\eta\int_{\underline{y}}^{\hat{y}_f} v\big(u(r)\big) f(y|e_H)\mathrm{d}y - c = \overline{U}. \qquad (A30)$$

The existence of $\hat{y}_f$ is guaranteed by the fact that the left-hand side of Eq. $(A30)$ can be negative if $\hat{y}_f = \overline{y}$ and for sufficiently large η, increases as $\hat{y}_f$ decreases, and is positive if $\hat{y}_f = \underline{y}$. Hence, the optimal incentive scheme is given by:

$$w_f^*(y) = \begin{cases} 0 & if\ y < \hat{y}_f, \\ w_f^F(y) & if\ y \geq \hat{y}_f. \end{cases} \qquad (A31)$$

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 $\tilde{y} \in [\underline{y}, \overline{y}]$, the solution also consists of two components: $w_f^F(y)$ satisfying $(A25)$, which implies $w_f^F(y) \geq r$, and $w_f^F(y)$ satisfying $(A26)$, which implies $w_f^F(y) < r$. Since the agent is loss averse, $\lambda > 1$, $w_f^F(y)$ satisfying $(A26)$ cannot be a solution on its own as it induces considerably disutility. This would lead the agent to reject the contract. Moreover, a combination of these two components would expose the agent to the risk of experiencing losses. Since $U(e_H, w(y), r)$ is concave, such exposure to losses cannot provide full insurance. Hence, it must be that the schedule $w_f^F(y)$ satisfying $(A18)$ is given everywhere. This proves the last part of the Theorem. ■

Corollary 1

It is first shown that μ is the same in the moral hazard and in the insurance problem. Multiply Eq. $(A6)$ by $f(y|e_H)$ and take expectations to obtain:

$$\int_{\underline{y}}^{\overline{y}} \frac{f(y|e_H)}{u'(w(y))\left(\phi + \eta v'\left(u(w(y))\right) - u(r)\right)} dy$$
$$= \int_{\underline{y}}^{\overline{y}} \mu f(y|e_H) dy + \int_{\underline{y}}^{\overline{y}} \gamma\left(f(y|e_H) - f(y|e_L)\right) dy. \qquad (A32)$$

Noting that $\int_{\underline{y}}^{\overline{y}} f(y|e) dy = 1$, the following expression can be obtained

$$\mathbb{E}\left(\frac{1}{u'(w(y))\left(\phi + \eta v'\left(u(w(y))\right) - u(r)\right)}\right) = \mu. \qquad (A33)$$

Similar steps lead to the following expression when Eq. $(A7)$ instead of Eq. $(A6)$ is used.

$$\mathbb{E}\left(\frac{1}{u'(w(y))\left(\phi + \lambda\eta v'\left(u(r) - u\left(w_f^F(y)\right)\right)\right)}\right) = \mu \qquad (A34)$$

Eqs. $(A33)$ and $(A34)$ show that μ does not depend on the value of γ, i.e. whether $\gamma = 0$ or $\gamma > 0$. Hence, a comparison of first- and second-best contracts can be performed assuming that μ takes the same value in both cases.

Part i). Let $U(e_H, w(\tilde{y}), r)$ be S-shaped for any $\tilde{y} \in [\underline{y}, \overline{y}]$. I proceed by contradiction by assuming $\hat{y}_f > \hat{y}_s$. Consider first the case in which the bonus of $w_s^*(y)$ is at least as large as that of $w_f^*(y)$. In that case, $w_s^*(y)$ does not impart punishments for low performance with respect to the first best.

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 incentivize high effort as it does not provide punishments for low performance. A contradiction since at the optimum the incentive compatibility constraint binds. Hence, it must be that $\hat{y}_f \geq \hat{y}_s$. A similar rationale applies when the bonus of $w_s^*(y)$ is assumed to be larger than that of $w_f^*(y)$.

Next, assume that the bonus in $w_s^*(y)$ is smaller than that in $w_f^*(y)$ and $\hat{y}_f \geq \hat{y}_s$. In that case, rewards are imparted non-monotonically. Denote by $\hat{y}_e$ the output level in $y > \hat{y}_f$ satisfying $w_s^*(\hat{y}_e) = w_f(\hat{y}_e)$. The existence and uniqueness of that output level is ensured by the properties: $w_s^*(\hat{y}_f) < w_f^*(\hat{y}_f)$, $w_s^*(y)$ increasing in performance in $y > \hat{y}_f$, $w_f^*(\hat{y}_f)$ being performance-insensitive in $y > \hat{y}_f$, and $w_s^*(y)$ implementing rewards at the highest-output levels, $y > \hat{y}_e$. Accordingly, $w_s^*(y)$ implements rewards with respect to $w_f^*(y)$ in $y \in (\hat{y}_s, \hat{y}_f)$, 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 might disincentivize high effort when sufficient probability mass in concentrated around $y \in (\hat{y}_s, \hat{y}_f)$. 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 $\tilde{y} \in [\underline{y}, \overline{y}]$. The optimal first-best contract $w_f^*(y)$ from Theorem 1, offers full insurance by bringing the agent to the domain of gains for all output realizations. The optimal second-best contract, $w_s^*(y)$, from Theorem 1, transitions 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 $\tilde{y} \in [\underline{y}, \overline{y}]$. Implicit differentiation of Eq. $(A6)$ gives

$$\frac{dw_s^*(y)}{dr} = \frac{u'(w_s^*(y))u'(r)\left(\eta v''\left(u(w_s^*(y)))\right) - u(r))\right)}{u''\left(w_s^*(y)\right)\left(\phi + \eta v'\left(u(w_s^*(y)))\right) - u(r))\right) + \left(u'(w_s^F(y))\right)^2\left(\eta v''\left(u(w_s^*(y)))\right) - u(r))\right)}. \quad (A35)$$

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

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. $(A21)$ to obtain:

$$\frac{d}{dr}\left(\frac{r}{\phi u(r)+\lambda\eta v\big(u(r)\big)}\right)=\frac{\Big(\phi u(r)-\phi u'(r)r+\lambda\eta v\big(u(r)\big)-\lambda\eta v'\big(u(r)\big)u'(r)r\Big)}{\Big(\phi u(r)+\lambda\eta v\big(u(r)\big)\Big)^2}. \qquad (A36)$$

Taylor's theorem around zero gives:

$$0=\phi u(r)+\lambda\eta v\big(u(r)\big)-\phi u'(r)r-\lambda\eta v'\big(u(r)\big)u'(r)r+\frac{\phi u''(r)r^2}{2}+\frac{\lambda\eta\big(v''\,(u(r))u'(r)+v'\,(u(r))u''(r)\big)r^2}{2}.$$

Since $u''<0$ and $v''<0$, Eq. $(A36)$ implies that $\frac{d}{dr}\left(\frac{r}{\phi u(r)+\lambda\eta v(u(r))}\right)>0$. Under a higher r, the equality in Eq. $(A21)$ is maintained by awarding the bonus at a higher threshold $\hat{y}_s$

Let $U(e_H,w(\tilde{y}),r)$ be concave. In that case, the result derived from Eq. $(A35)$ that $\frac{dw_s^*(y)}{dr}>0$ for the domain of gains continues to hold. Furthermore, implicit differentiation of Eq. $(A7)$ gives.

$$\frac{dw_s^*(y)}{dr}=\frac{-u'\big(w_s^*(y)\big)u'(r)\Big(\lambda\eta v''\,\Big(u(r)-u\big(w_s^*(y)\big)\Big)\Big)}{u''\,\big(w_s^*(y)\big)\Big(\phi+\lambda\eta v'\,\Big(u(r)-u\big(w_s^*(y)\big)\Big)\Big)-\Big(u'\big(w_s^*(y)\big)\Big)^2\Big(\lambda\eta v''\,\Big(u(r)-u\big(w_s^*(y)\big)\Big)\Big)}. \qquad (A37)$$

Lemma 1 shows that the denominator of Eq. $(A37)$ is negative for any $\tilde{y}\in\left[\underline{y},\overline{y}\right]$. Also, since $v''<0$, the numerator of that equation is positive and $\frac{dw_s^*(y)}{dr}<0$. The magnitude of the bonus increases as r increases.

To investigate the influence of changes in r on the location of the bonus, compute the derivative of Eq. $(A23)$ with respect to r to obtain:

$$-\eta\int_{\hat{y}_s}^{\overline{y}}u'(r)\,v'\,\Big(u\big(w_s^F(y)\big)-u(r)\Big)f(y|e_H)\,\mathrm{d}y-\lambda\eta\int_{\underline{y}}^{\hat{y}_s}u'(r)v'\,\Big(u(r)-u\big(w_s^F(y)\big)\Big)f(y|e_H)\mathrm{d}y<0 \qquad (A38)$$

Hence, for the equality in $(A23)$ to hold under higher values of r, $\hat{y}_s$ needs to become 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 each output realization $\tilde{y}\in\left[\underline{y},\overline{y}\right]$. The left-hand side of Eq. $(A21)$ decreases as λ increases. Therefore, to maintain that equality, and due to Assumption 2, $\hat{y}_s$ must decrease; $w_s^*(y)$ exhibits a smaller segment in which $w_s^*(y)=0$ is the solution. Finally, notice that λ 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

$$\frac{dw_s^*(y)}{d\lambda}=\frac{-u'\big(w_s^F(y)\big)\Big(\eta v'\,\Big(u(r)-u\big(w_s^F(y)\big)\Big)\Big)}{u''\big(w_s^F(y)\big)\Big(\phi+\lambda\eta v'\,\Big(u(r)-u\big(w_s^F(y)\big)\Big)\Big)-\Big(u'\big(w_s^F(y)\big)\Big)^2\Big(\lambda\eta v''\,\Big(u(r)-u\big(w_s^F(y)\big)\Big)\Big)}. \qquad (A39)$$

Lemma 1 shows that the denominator of Eq. $(A39)$ is negative. Since $u' > 0$, then $\frac{dw_S^*(y)}{d\lambda} > 0$. Moreover, note that λ does not enter in Eq.$(A6)$. The bonus shrinks.

To investigate the influence of changes in λ on the location of the bonus, note that the derivative of $(A24)$ with respect to λ gives $-\eta \int_{\underline{y}}^{\hat{y}_S} v\big(u(r) - u(w_S^F(y))\big) f(y|e_H) dy$, a negative expression. Hence, for the equality in $(A24)$ to hold under higher λ, $\hat{y}_S$ needs to be smaller. The exposition of the agent to losses is reduced as loss aversion increases. ■

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 $-\frac{u''(w(\tilde{y}))}{u'(w(\tilde{y}))} \geq 0$. Contradicting the necessary and sufficient condition given in Eq. $(A3)$. Consequently, the contracts satisfying Eq. $(A6)$ and Eq. $(A7)$ are sufficient to solve the principal's maximization problem.

Under the assumed restrictions, $(A6)$ and $(A7)$ become:

$$\frac{1}{u'(w_S^*(y))} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right). \qquad \text{(A40)}$$

Rearranging the above expression leads to:

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

Moreover, equation (A8) becomes:

$$\frac{dw_S^*(y)}{dy} = \frac{\gamma \frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right)\left(u'(w_S^F(y))\right)^2}{u''\left(w_S^F(y)\right)}. \qquad \text{(A42)}$$

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 $(A41)$, 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

$$w_f^*(y) = h'\left(\frac{1}{\mu}\right). \qquad \text{(A43)}$$

Equation $(A43)$ 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 $\tilde{y} \in \left[\underline{y}, \overline{y}\right]$. 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 solution from the first-order approach for the domain of gains is sufficient and necessary to solve the principal's problem. Under the assumed restrictions that solution, given by Eq. $(A6)$, becomes

$$\frac{1}{v'(w_S^*(y) - r)} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right). \qquad \text{(A44)}$$

Rearranging the above expression, the following closed-form expression is obtained:

$$w_S^*(y) \ = r + f'\left(\frac{1}{\mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right)}\right). \qquad \text{(A45)}$$

Moreover, Eq. (A8) becomes

$$\frac{dw_S^*(y)}{dy} = \frac{\gamma \frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right)\left(v'\ (w_S^F(y) - r)\right)^2}{v''(w_S^F(y) - r)}. \qquad (A46)$$

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. $(A21)$ becomes:

$$\frac{1}{\frac{\lambda v\left(u(r)\right)}{r}} = \mu + \gamma\left(1 - \frac{f\left(\hat{y}_{ps}\middle|e_L\right)}{f\left(\hat{y}_{ps}\middle|e_H\right)}\right). \qquad (A47)$$

So, the transition from losses to gains is given by the $\hat{y}_{ps}$ that satisfies $(A47)$. 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. $(A45)$ becomes

$$w_s^*(y) \ = r + f'\left(\frac{1}{\mu}\right). \qquad \text{(A48)}$$

Eq. $(A48)$ shows that $\frac{dw_f^*(y)}{dy} = 0$ if $\theta_{\mathbb{I}} = 1$. Moreover, under the considered restrictions, Eq. $(A30)$ becomes:

$$\int_{\hat{y}_{p1}}^{\overline{y}} v\big(w_f^*(y) - r\big) f(y|e_H)\, \mathrm{d}y - \lambda \int_{\underline{y}}^{\hat{y}_{pf}} v(r) f(y|e_H) \mathrm{d}y - c \ = \overline{U}. \qquad (A49)$$

The transition from losses to gains is given by the $\hat{y}_f$ that satisfies $(A49)$. The existence of $\hat{y}_{pf} \in (\underline{y}, \overline{y})$ is guaranteed by the fact that $\int_{\hat{y}_{pf}}^{\overline{y}} v\big(w_f^*(y) - r\big) f(y|e_H)\, \mathrm{d}y$ is positive, while $-\lambda \int_{\underline{y}}^{\hat{y}_{pf}} v(r) f(y|e_H) \mathrm{d}y$ is negative. The lower and upper limit of those integrals, $\hat{y}_{pf}$, can be adjusted to obtain a positive expression in the left-hand side of $(A49)$ equal to $\overline{U} \geq 0$. Also, that both $w_f^*(y)$ satisfying Eq. (A48) and the lottery payment L make the agent's participation constraint bind, as shown by Eqs. (A33), (A34), and (A27), guarantees the existence of $\hat{y}_{pf}$. 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 or rejecting the contract and choosing an effort level. These choices generate the set of candidates for max-min reference point $r = \max\{\min\{\overline{U}\}, \min\{w_s^*(y)\}\}$. Intuitively, rejecting the contract yields welfare $\overline{U}$ and the minimum that the contract pays, regardless of effort level, is $\min\{w_s^*(y)\}$.

Since the participation constraint binds at the optimum, then $\mathbb{E}\big(U(e, w_s^*(y), r)\big) = \overline{U}$. Also, because the incentive compatibility constraint binds, it must be that $\mathbb{E}\big(U(e, w_s^*(y), r)\big) > \mathbb{E}\big(U(r, \min\{w_s^*(y)\}, r)\big)$. Otherwise, the second-best contract would not implement punishments for low performance and thus would not incentivize high effort. Hence, $r = w_s^*(y)$

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 $r = w_s^*(y)$ 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:

$$\frac{1}{\frac{\lambda v\left(w_s^*(y)\right)}{w_s^*(y)}} = \mu + \gamma\left(1 - \frac{f(\hat{y}_u|e_L)}{f(\hat{y}_u|e_H)}\right). \qquad (A50)$$

The above condition is analog to that presented in Eq. $(A47)$ when $r = w_s^*(y)$. The existence and uniqueness of that output level is guaranteed by $\frac{1}{\frac{\lambda v\left(w_s^*(y)\right)}{w_s^*(y)}} > 0$, $\frac{d}{dy}\left(\frac{1}{\frac{\lambda v\left(w_s^*(y)\right)}{w_s^*(y)}}\right) = \frac{v(w_s^*(y)) - w_s^*(y)v'(w_s^*(y))}{\lambda\left(v(w_s^*(y))\right)^2} < 0$ due to Taylor's theorem around zero, and $-\frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right) \geq 0$ (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. (A44). Adapting that incentive schedule to account for the considered reference point rule gives:

$$\frac{1}{\int_{\underline{y}}^{\overline{y}} v'\left(w_s^*(y) - w_s^*(\tilde{\tilde{y}})\right) f(\tilde{\tilde{y}}|e) d\tilde{\tilde{y}}} = \mu + \gamma\left(1 - \frac{f(y|e_H)}{f(y|e_L)}\right), \qquad (A51)$$

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

$$w_s^*(y) = \begin{cases} 0 & if\ y < \hat{y}_u, \\ w_s^*(y) \text{ satisfying } (A51) & if\ y \geq \hat{y}_u. \end{cases}$$

■

To prove Proposition 1, a preliminary result is proved. The following lemma shows that, under certain conditions on the agent's incentive scheme, it is demonstrated that 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), *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:

$$\frac{\mathrm{d}U(e, w(y), g)}{\mathrm{d}g} = -\int_{g}^{\overline{y}} w'(g)v'\big(w(y) - w(g)\big)f(y|e)dy - \lambda\int_{\underline{y}}^{g} w'(g)v'\big(w(g) - w(y)\big)f(y|e)\mathrm{d}y. \qquad (A52)$$

Since $v' > 0$ and $\lambda > 1$, Eq. $(A52)$ is weakly negative if $w'(g) \geq 0$. In that case, higher goals induce disutility. This proves the first part of the lemma.

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

$$U(e, w(y), g) = v\big(w(\overline{y}) - w(g)\big) - \int_{g}^{\overline{y}} w'(y)v'\big(w(y) - w(g)\big)F(y|e)\mathrm{dy} - \lambda\int_{\underline{y}}^{g} w'(y)u'\big(w(g) - w(y)\big)\, F(y|e)dy - c(e). \qquad (A53)$$

The agent's incentive compatibility constraint can be rewritten using Eq.$(A53)$ as

$$-\int_{g}^{\overline{y}} w'(y)v'\big(w(y) - w(g)\big)\big(F(y|e_H) - F(y|e_L)\big)\mathrm{dy} - \lambda\int_{\underline{y}}^{g} w'(y)v'\big(w(g) - w(y)\big)\,\big(F(y|e_H) - F(y|e_L)\big)\mathrm{dy} \geq c. \qquad (A54)$$

To investigate whether higher goals incentivize high effort, derive $(A54)$ with respect to g to obtain:

$$-(\lambda - 1)w'(g)v'\big(w(g) - w(g)\big)\big(F(g|e_H) - F(g|e_L)\big) + \int_{g}^{\overline{y}} w'(y)w'(g)v''\big(w(y) - w(g)\big)\big(F(y|e_H) - F(y|e_L)\big)\mathrm{dy} - \lambda\int_{\underline{y}}^{g} w'(y)w'(g)v''\big(w(g) - w(y)\big)\,\big(F(y|e_H) - F(y|e_L)\big)\mathrm{dy}. \qquad (A55)$$

Since $v' > 0$, $v'' < 0$, and $F(y|e_L) \geq F(y|e_H)$, an implication of Assumption 2, Eq.$(A55)$ shows that higher goals generate higher effort only if $w'(g) > 0$, $w'(y) > 0$ in $y \geq g$ and $w'(y) \leq 0$ 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 $-\frac{v''\left(u(w(g))-u(w(\tilde{y}))\right)}{v'\left(u(w(g))-u(w(\tilde{y}))\right)} \geq 0$, corroborating the condition in Eq. $(A3)$. Therefore, the solution from the first-order condition is only sufficient and necessary for the domain of gains. That condition, given by Eq. $(A6)$, rewrites under the considered restrictions as:

$$\frac{1}{v'(w_s^*(y) - w_s^*(g))} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right). \qquad \text{(A56)}$$

Rearranging (A56) gives

$$w_s^*(y) = w_s^*(g) + f'\left(\frac{1}{\mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right)}\right). \qquad \text{(A57)}$$

Eq. $(A57)$ shows that the requirement $w'(g) > 0$ from Lemma 2, immediately implies $\frac{dw_s^*(y)}{dg} > 0$. Moreover, the derivative of Eq. (A56) 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, the second-best optimal contract must increase both in goals and performance in the domain of gains.

To study the location of the bonus, Eq. $(A21)$ is rewritten to account for the considered restrictions, yielding:

$$\frac{1}{\frac{\lambda v(w_s^*(g))}{w_s^*(g)}} = \nu + \gamma\left(1 - \frac{f(\hat{y}_g|e_L)}{f(\hat{y}_g|e_H)}\right). \qquad (A58)$$

The output level $\hat{y}_g \in \left[\underline{y}, \overline{y}\right]$ that satisfies $(A58)$ is the critical threshold that 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 agent with S-shaped utility function, paying $w_s^*(y) = 0$ in $y < \hat{y}_g$ is second-best optimal. Accordingly, the second-best contract is

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

This solution exhibits a discrete jump at $y = \hat{y}_g$, as $\lim_{y \to \hat{y}_g^+} w_s^*(y) = w_s^*(g)$ and $\lim_{y \to \hat{y}_g^-} w_s^*(y) = 0$. Moreover, a consequence of committing to an incentives scheme with $w_s^{*\prime}(g) > 0$, a condition that Lemma 2 shows necessary for goals to be motivating, is that the bonus must become larger as g increases.

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 her profits by setting $w_s^*(y) = 0$ and, due to the convexity of $U(e_H, w(\tilde{y}), r)$ in losses, the agent would be willing to accept that contract change. Now suppose $\hat{y}_g > g$. The agent is in that case overexposed to risk in $y \in [g, \hat{y}_g]$, where he is risk averse. This potentially leads the agent to reject the contract. Hence, to ensure that the contract is accepted, the principal offers $w_s^*(y)$ given by $(A57)$ for any $y \geq g$. It must be then that $\hat{y}_g = g$.

The result $\hat{y}_g = g$ has three relevant 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. Notice that this is consistent with the result of Corollary 2, obtained when the reference point was assumed to be exogenous. Thirdly, 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. Consequently, 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:

$$\mu\left(-w'(g)\theta_{\mathbb{I}}v'\big(w(y)-w(g)\big)f(y|e)-\lambda(1-\theta_{\mathbb{I}})w'(g)v'\left(w(g)-w(y)\right)f(y|e)\right)$$
$$+\gamma\left(-w'(g)\theta_{\mathbb{I}}v'\big(w(y)-w(g)\big)\big(f(y|e_H)-f(y|e_L)\big)\right.$$
$$\left.-\lambda w'(g)(1-\theta_{\mathbb{I}})v'\big(w(g)-w(y)\big)\big(f(y|e_H)-f(y|e_L)\big)\right)=0. \qquad (A59)$$

Denoting by g^* the goal level that satisfies $(A59)$, the following conditions are obtained:

$$-w'(g)\left(\mu v'\big(w(y)-w(g^*)\big)f(y|e)+\gamma u'\big(w(y)-w(g^*)\big)\big(f(y|e_H)-f(y|e_L)\big)\right)=0, \quad (A60)$$

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

$$-\lambda w'(g)\left(\mu v'\big(w(g^*)-w(y)\big)f(y|e)+\gamma v'\big(w(g^*)-w(y)\big)\big(f(y|e_H)-f(y|e_L)\big)\right)=0. \quad (A61)$$

if $\theta_{\mathbb{I}}=0$. Eqs. $(A60)$ and $(A61)$ imply:

$$\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right)=0. \qquad (A62)$$

Eqs. $(A56)$ and $(A62)$ are used to obtain:

$$v'(w_s^*(y)-w_s^*(g^*))=+\infty. \qquad (A63)$$

Since $\lim_{y\to g^+} w_s^*(y)=w_s^*(g)$, then, due to the Inada condition, $\lim_{y\to g^+} v'\big(w_s^*(y)-w_s^*(g^*)\big)=+\infty$. Because the realization of y is ex-ante unknown to the principal, she cannot set a goal that is just met to comply with $(A63)$. However, she can set a goal such that $(A63)$ holds on expectation. Using the fact that $v'(\cdot)$ is concave, then

$$\mathbb{E}\big(v'(w_s^*(y)-w_s^*(g)\big)\leq v'\left(\mathbb{E}\big(w_s^*(y)\big)-w_s^*(g)\right). \quad (A64)$$

Therefore, g^* can be set such that $\mathbb{E}\big(w_s^*(y)\big)-w_s^*(g^*)=\epsilon$ for arbitrarily small $\epsilon>0$. This gives $v'\left(\mathbb{E}\big(w_s^*(y)\big)-w_s^*(g^*)\right)=+\infty.$■

Corollary 6.

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

That solution, given in Eq. $(A6)$, becomes:

$$\frac{1}{\left(1+\eta v'\left(w_s^*(y)-\bar{\bar{w}}\right)\right)}=\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right). \qquad (A65)$$

Algebraic manipulations of the above equation yield $w_s^*(y)=\bar{\bar{w}}+f'\left(\frac{1}{\eta}\left(\frac{1}{\mu+\gamma\left(1-\frac{f(y|e_L)}{f(y|e_H)}\right)}-1\right)\right)$. Moreover, implicit derivation of Eq. $(A65)$ gives:

$$\frac{dw_s^*(y)}{dy}=\frac{\gamma\frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right)\left(1+\eta v'\left(w_s^*(y)-\bar{\bar{w}}\right)\right)^2}{\left(\eta v''(w_s^*(y)-\bar{\bar{w}})\right)}. \qquad (A66)$$

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 optimal increases in performance in the domain of gains.

The transition from losses to gains is given by Eq. $(A21)$, which, under the assumed restrictions, is equal to:

$$\frac{1}{\frac{1+\lambda v(\bar{\bar{w}})}{\bar{\bar{w}}}}=\mu+\gamma\left(1-\frac{f(\hat{y}_{ms}|e_L)}{f(\hat{y}_{ms}|e_H)}\right). \qquad (A67)$$

The bonus is given at $y=\hat{y}_{ms}$ for $\hat{y}_{ms}\in(\underline{y},\overline{y})$ satisfying Eq. $(A67)$. According to Theorem 1, it is second-best optimal to set $w_s^*(y)=0$ in $y<\hat{y}_{ms}$ and the solution from the first-order approach elsewhere. 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.

Next, I to study the first-best contract. Consider $\gamma=0$ in addition to $\phi=1$, $u'=1$, and $r=\bar{\bar{w}}$. Eq. $(A65)$ 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. $(A66)$, that transfer exhibits $\frac{dw_f^*(y)}{dy}=0$. Furthermore, Theorem 1, shows that is first-best optimal to provide this fixed-transfer for high output levels.

Theorem 1 also demonstrates that is first-best optimal to offer $w_f^*(y) = 0$ at the low-end of the output space. The transition from $w_f^*(y) = 0$ to $w_f^*(y) = \overline{\overline{w}} + f'\left(\frac{1}{\eta}\left(\frac{1}{\mu} - 1\right)\right)$ is given by the following equality, which adapts Eq. $(A30)$ to account for the assumed restrictions:

$$\int_{\hat{y}_{mf}}^{\overline{y}} w_f^*(y) f(y|e_H)\,\mathrm{d}y + \eta \int_{\hat{y}_{mf}}^{\overline{y}} v\left(w_f^*(y) - \overline{\overline{w}}\right) f(y|e_H)\,\mathrm{d}y - \lambda \int_{\underline{y}}^{\hat{y}_{mf}} v(\overline{\overline{w}}) f(y|e_H)\mathrm{d}y - c = \overline{U}. \quad (A68)$$

The output level $\hat{y}_{mf} \in [\underline{y}, \overline{y}]$ that satisfies Eq.$(A68)$ provides the transition from losses to gains. Theorem 1 shows that this output level is unique and interior. Hence, the first-best optimal contract is $w_f^*(y) = \begin{cases} 0 & if\ y < \hat{y}_{mf}, \\ \overline{\overline{w}} + f'\left(\frac{1}{\eta}\left(\frac{1}{\mu} - 1\right)\right) & if\ y \geq \hat{y}_{mf}. \end{cases}$ This proves the third part of the corollary.

Consider next $\phi = 1$, $v' = 1$, and $r = \overline{\overline{w}}$. Under these restrictions $U(e_H, w(\tilde{y}), \overline{\overline{w}})$ is concave for each $\tilde{y} \in \left[\underline{y}, \overline{y}\right]$ since $-\frac{u''(w(\tilde{y}))}{u'(w(\tilde{y}))} \geq 0$, contradicting the condition in Eq. $(A3)$. Hence, the solutions from the first-order approach are necessary and sufficient to solve the principal's maximization problem.

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

$$\frac{1}{u'(w_s^*(y))(1+\eta)} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right), \qquad (A69)$$

and

$$\frac{1}{u'(w_s^*(y))(1+\eta\lambda)} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right), \qquad (A70)$$

respectively. Eq. $(A69)$ 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 $\frac{dw_s^*(y)}{dy} \geq 0$, as shown by Eqs. $(A8)$ and $(A9)$.

The transition from $w_s^*(y)$ satisfying $(A69)$ to $w_s^*(y)$ satisfying $(A70)$ is given by the unique output level $\hat{y}_{ms} \in (\underline{y}, \overline{y})$ that satisfies:

$$\int_{\hat{y}_{ms}}^{\overline{y}} u(w_S^*(y))f(y|e_H)\,\mathrm{d}y + \eta\int_{\hat{y}_{ms}}^{\overline{y}} \left(u(w_S^*(y)) - u(\overline{\overline{w}})\right)f(y|e_H)\,\mathrm{d}y + \int_{\underline{y}}^{\hat{y}_{ms}} u(w_S^*(y)f(y|e_H)\mathrm{d}y$$
$$- \lambda\eta\int_{\underline{y}}^{\hat{y}_{ms}} \left(u(\overline{\overline{w}}) - u\left(w_s^*(y)\right)\right) f(y|e_H)\mathrm{d}y - c \ = \overline{U}. \qquad (A71)$$

The existence of $\hat{y}_{ms}$ is guaranteed by the fact that the solutions from Eqs. $(A69)$ and $(A70)$ make the participation constraint bind for gains and losses, respectively. This is evident from Eqs. (A33), (A34). As a result, the optimal incentive scheme is given by:

$$w_f^*(y) = \begin{cases} h'\left(\dfrac{1}{(1+\eta)\left(\mu + \gamma\left(1 - \dfrac{f(y|e_L)}{f(y|e_H)}\right)\right)}\right) & if\ y \geq \hat{y}_{ms}\,, \\ h'\left(\dfrac{1}{(1+\eta\lambda)\left(\mu + \gamma\left(1 - \dfrac{f(y|e_L)}{f(y|e_H)}\right)\right)}\right) & if\ y < \hat{y}_{ms}. \end{cases} \qquad (A72)$$

This proves the second part of the corollary.

To conclude, we analyze the first-best contract when utility is concave. Add $\gamma = 0$ to the considered set of restrictions: $\phi = 1$, $v' = 1$, and $r = \overline{\overline{w}}$. The first-order condition in $(A69)$ 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 paying $w_f^*(y) = \ h'\left(\frac{1}{(1+\eta)\mu}\right)$ everywhere 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, respectively. The Lagrangian of the principal's maximization problem writes as

$$\mathcal{L} = \big(S(y) - w(y)\big) f(y|e_H) + \mu\Bigg(u(w(y)) f(y|e_H) + \eta\theta_{\mathbb{I}} \int_{\underline{y}}^{\overline{y}} v\big(u(w(y) - u(w(\tilde{\tilde{y}}))\big) f(y|e_H) f(\tilde{\tilde{y}}|e_H)\mathrm{d}\tilde{\tilde{y}} - \eta\lambda(1-\theta_{\mathbb{I}}) \int_{\underline{y}}^{\overline{y}} v\left(u(w(\tilde{\tilde{y}})) - (u(w(y))\right) f(y|e_H) f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}} - c\Bigg) + \gamma\Bigg((u(w(y))\big(f(y|e_H) - f(y|e_L)\big) + \theta_{\mathbb{I}}\eta \int_{\underline{y}}^{\overline{y}} v\left((u(w(y)) - u(w(\tilde{\tilde{y}})\right)\big(f(y|e_H) - f(y|e_L)\big) f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}} - \lambda\eta(1-\theta_{\mathbb{I}}) \int_{\underline{y}}^{\overline{y}} v\left(u(w(\tilde{\tilde{y}}) - (u(w(y))\right)\big(f(y|e_H) - f(y|e_L)\big) f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}} - c\Bigg). \tag{A73}$$

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

$$-f(y|e_H) + \mu\Bigg(u'\big(w(y)\big) f(y|e_H) + \eta\theta_{\mathbb{I}} \int_{\underline{y}}^{\overline{y}} v'\big(u(w(y) - u(w(\tilde{\tilde{y}}))u'\,(w(y) f(y|e_H) f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}} + \eta\lambda(1-\theta_{\mathbb{I}}) \int_{\underline{y}}^{\overline{y}} v'\left(u\left(w(\tilde{\tilde{y}})\right) - (u(w(y))\right) u'\,(w(y) f(y|e_H) f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}\Bigg) + \gamma\Bigg((u'(w(y))\big(f(y|e_H) - f(y|e_L)\big) + \theta_{\mathbb{I}}\eta \int_{\underline{y}}^{\overline{y}} v'\left((u(w(y)) - u(w(\tilde{\tilde{y}})\right) u'(w(y)\big(f(y|e_H) - f(y|e_L)\big) f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}} + \lambda\eta(1-\theta_{\mathbb{I}}) \int_{\underline{y}}^{\overline{y}} v'\left(u(w(\tilde{\tilde{y}}) - (u(w(y))\right)\big(f(y|e_H) - f(y|e_L)\big) u'\big(w(y)\big) f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}\Bigg) = 0. \tag{A74}$$

Denoting by $w_s^F(y)$ the transfer satisfying $(A74)$, the following expressions are obtained after algebraic manipulations:

$$\frac{1}{u'\big(w_s^F(y)\big) + \eta \int_{\underline{y}}^{\overline{y}} v'\big(u(w_s^F(y) - u(w_s^F(\tilde{\tilde{y}}))u'\,(w_s^F(y)) f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right), \tag{A75}$$

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

$$\frac{1}{u'\big(w_s^F(y)\big) + \eta\lambda \int_{\underline{y}}^{\overline{y}} v'\left(u(w_s^F(\tilde{\tilde{y}}) - (u(w_s^F(y))\right) u'\,(w_s^F(y)) f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}} = \mu + \gamma\left(1 - \frac{f(y|e_L)}{f(y|e_H)}\right), \tag{A76}$$

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 $(A75)$ with respect to y exhibits $\frac{dw_s^F(y)}{dy} \geq 0$. Similarly, the derivative of $(A76)$ 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 decreasing, $\frac{dw_s^F(y)}{dy} \leq 0$, if $U\left(e_H, w(\tilde{y}), w(\tilde{\tilde{y}})\right)$ is S-shaped. As mentioned in Theorem 1, the latter is an undesirable property.

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 each $\tilde{y} \in [\underline{y}, \overline{y}]$. If $w_s^F(y)$ satisfying Eq. $(A76)$ 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

$$\mathbb{E}_y\left(u\left(w_s^F(y)\right)\right) - \eta\lambda \int_{\underline{y}}^{\tilde{y}} \int_{\underline{y}}^{\overline{y}} v\left(u\left(w_s^F(\tilde{\tilde{y}})\right) - u\left(w_s^F(y)\right)\right) f(\tilde{\tilde{y}}|e)\,\mathrm{d}\tilde{\tilde{y}} f(y|e)\mathrm{d}y$$
$$+ \eta \int_{\tilde{y}}^{\overline{y}} \int_{\underline{y}}^{\overline{y}} v\left(u\left(w_s^F(y)\right) - u\left(w_s^F(\tilde{\tilde{y}})\right)\right) f(\tilde{\tilde{y}}|e)\,\mathrm{d}\tilde{\tilde{y}} f(y|e)\mathrm{d}y,$$

increases in $w_s^F(y)$, there must exist a $p \in [0,1]$ such that:

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

Hence, replacing $w_s^F(y)$ from Eq. $(A76)$ by L_s leaves the agent's participation incentive compatibility constraints unchanged. Since $\mathbb{E}_y\left(u\left(w_s^F(y)\right)\right) - p\,\mathbb{E}_y\left(u\left(w_s^F(y)\right)\right) \geq 0$, Eq. $(A77)$,the concavity of v, and loss aversion, $\lambda > 1$, imply:

$$-\eta\int_{\tilde{y}}^{\overline{y}}\int_{\underline{y}}^{\overline{y}} v\left(u\left(w_s^F(y)\right)-u\left(w_s^F(\tilde{\tilde{y}})\right)\right) f(\tilde{\tilde{y}}|e)\,\mathrm{d}\tilde{\tilde{y}} f(y|e)\mathrm{d}y$$
$$+\eta\lambda\int_{\underline{y}}^{\tilde{y}}\int_{\underline{y}}^{\overline{y}} v\left(u\left(w_s^F(\tilde{\tilde{y}})\right)-u\left(w_s^F(y)\right)\right) f(\tilde{\tilde{y}}|e)\,\mathrm{d}\tilde{\tilde{y}} f(y|e)\mathrm{d}y$$
$$> -p\eta\left(\left(\left(\int_{\underline{y}}^{\overline{y}} v\left((1-p)u\left(w_s^F(y)\right)\right) f(y|e)\mathrm{d}y\right)\right.\right.$$
$$\left.\left.+\lambda\left(\int_{\underline{y}}^{\overline{y}} v\left((1-p)u\left(w_s^F(y)\right)\right) f(y|e)\mathrm{d}y\right)\right)\right) \tag{A78}$$

Some rearranging and using the fact that $\tilde{\tilde{y}} \in [\underline{y}, \overline{y}]$, gives

$$\lambda\int_{\underline{y}}^{\tilde{y}}\int_{\underline{y}}^{\overline{y}} v\left(u\left(w_s^F(\tilde{\tilde{y}})\right)-u\left(w_s^F(y)\right)\right)-pv\left((1-p)\,u\left(w_s^F(\tilde{\tilde{y}})\right)\right) f(\tilde{\tilde{y}}|e)\,\mathrm{d}\tilde{\tilde{y}} f(y|e)\mathrm{d}y >$$
$$\int_{\tilde{y}}^{\overline{y}}\int_{\underline{y}}^{\overline{y}} v\left(u\left(w_s^F(y)\right)-u\left(w_s^F(\tilde{\tilde{y}})\right)\right)-pv\left((1-p)u\left(w_s^F(\tilde{\tilde{y}})\right)\right) f(\tilde{\tilde{y}}|e)\,\mathrm{d}\tilde{\tilde{y}} f(y|e)\mathrm{d}y. \tag{A79}$$

The inequality in Eq. $(A79)$ is satisfied by

$$u(w_s^F(y)) > pu(w_s^F(y)). \tag{A80}$$

Thus, L_s is more cost-effective for the principal than the solution implied by $(A76)$.

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:

$$U\left(e_H, L, w_s^F(\tilde{\tilde{y}})\right)=\left(\frac{\overline{L}_s}{w_s^F(\tilde{\tilde{y}})}\right)u\left(w_s^F(\tilde{\tilde{y}})\right)+\frac{\overline{L}_s}{w_s^F(\tilde{\tilde{y}})}\left(1-\frac{\overline{L}_s}{w_s^F(\tilde{\tilde{y}})}\right)\eta\left(v\left(u\left(w_s^F(\tilde{\tilde{y}})\right)\right)-\lambda v\left(-u\left(w_s^F(\tilde{\tilde{y}})\right)\right)\right). \tag{A81}$$

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. $(A81)$ with respect to p is:

$$u\left(w_s^F(\tilde{\tilde{y}})\right)+(1-2p)\eta\left(v\left(u\left(w_s^F(\tilde{\tilde{y}})\right)\right)-\lambda v\left(-u\left(w_s^F(\tilde{\tilde{y}})\right)\right)\right)=0. \tag{A82}$$

The second derivative of Eq. $(A80)$ 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

$p_s^*(\tilde{\tilde{y}}) = \frac{1}{2} + \frac{u\left(w_s^F(\tilde{\tilde{y}})\right)}{2\eta\left(v\left(u\left(w_s^F(\tilde{\tilde{y}})\right)\right) - \lambda v\left(-u\left(w_s^F(\tilde{\tilde{y}})\right)\right)\right)}$. Hence, the principal implements the lottery $L_s^* = (p_s^*(\tilde{\tilde{y}}): w_s^F(\tilde{\tilde{y}}), 1 - p_s^*(\tilde{\tilde{y}}): 0)$ if the contract $w_s^F(y)$ satisfying Eq. $(A76)$ exhibits $0 < w_s^F(y) < w_s^F(\tilde{\tilde{y}})$.

It is next shown that L_s^* is implemented for all $y < \overline{y}$. I proceed by contradiction. If there was an interior threshold output, $\hat{y}_{ds}$b,elow which L_s^* is paid, and above which $w_s^F(y)$ from Eq. $(A75)$ is paid, that output level ought to satisfy:

$$\frac{1}{\frac{u(w_s^F(\hat{y}_{ds}))}{w_s^F(\hat{y}_{ds})} + \frac{1}{w_s^F(\hat{y}_{ds})}\left(1 - \frac{2\overline{L}}{w_s^F(\hat{y}_{ds})}\right)\eta\left(v\left(u(w_s^F(\hat{y}_{ds}))\right) - \lambda v\left(-u(w_s^F(\hat{y}_{ds}))\right)\right)} = \mu + \gamma\left(1 - \frac{f(\hat{y}_{ds}|e_L)}{f(\hat{y}_{ds}|e_H)}\right) \tag{A83}$$

Since p_s^*satisfies Eq. $(A82)$, the numerator of the left-hand side of Eq. $(A83)$ is zero. Due to $\frac{d}{dy}\left(\frac{f(y|e_L)}{f(y|e_H)}\right) \leq 0$ (Assumption 2), the solution to Eq. $(A82)$ is at the boundary of the output space, $\hat{y}_{ds} = \overline{y}$.

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 payment of $w_s^F(y)$ satisfying Eq. $(A75)$ evaluated at $y = \overline{y}$. Eq. $(A83)$ shows that 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 at $y = \overline{y}$. This proves 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. $(A75)$, which implies $w_s^F(y) \geq w_s^F(\tilde{\tilde{y}})$, and $w_s^F(y)$ satisfying Eq. $(A76)$, which implies $w_s^F(y) < w_s^F(\tilde{\tilde{y}})$. The optimal contract combines these two first-order conditions. To see how, notice that a contract only paying $w_s^F(y)$ satisfying Eq. $(A76)$ might lead to excess gains. The principal could deviate from this solution by implement higher outcomes that generate losses when incorporated as reference point. Moreover, a contract only paying Eq. $(A76)$ leads to excess losses and would be rejected. Therefore the optimal contract consists of both components.

The transition from $w_s^F(y)$ satisfying Eq. $(A75)$ to $w_s^F(y)$ satisfying Eq. $(A76)$ is given by $\hat{y}_{ds} \in [\underline{y}, \overline{y}]$ satisfying:

$$\int_{\underline{y}}^{\hat{y}_{ds}} u(w_s^F(y))f(y|e)dy + \int_{\hat{y}_{ds}}^{\overline{y}} u(w_s^F(y))f(y|e)dy + \eta\int_{\hat{y}_{ds}}^{\overline{y}}\int_{\underline{y}}^{\overline{y}} v\left(u(w_s^F(y)) - u\left(w_s^F(\tilde{\tilde{y}})\right)\right)f(\tilde{\tilde{y}}|e)f(y|e)d\tilde{\tilde{y}}dy$$
$$- \eta\lambda\int_{\underline{y}}^{\hat{y}_{ds}}\int_{\underline{y}}^{\overline{y}} v\left(u\left(w_s^F(\tilde{\tilde{y}})\right) - u(w_s^F(y))\right)f(y|e)\, f(\tilde{\tilde{y}}|e)\, d\tilde{\tilde{y}}dy - c(e) = \overline{U} \tag{A84}$$

The existence of $\hat{y}_{ds}$ is guaranteed inasmuch as the left-hand side of Eq. $(A84)$ can be negative when $\hat{y}_{ds} = \overline{y}$, it increases as $\hat{y}_{ds}$ increases and it becomes positive when $\hat{y}_{ds} = \underline{y}$.

Hence, a candidate for optimal contract is given by

$$w_s^*(y) = \begin{cases} w_s^F(y) & \text{satisfying } (A75) \quad \text{if } y \geq \hat{y}_{ds}, \\ w_s^F(y) & \text{satisfying } (A76) \quad \text{if } y < \hat{y}_{ds}. \end{cases} \qquad (A85)$$

That solution exhibits a discrete jump at $y = \hat{y}_{ds}$ since $\lambda > 1$ appears in the denominator of the right-hand side of $(A76)$ but this coefficient does not enter in $(A75)$. This proves the second part of the theorem.

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 stochastic contract $L_s^* = (p_s^*(\overline{y})\colon b_s, 1 - p_s^*(\overline{y})\colon 0)$ if $y < \overline{y}$ and b_s at $y = \overline{y}$, and the contract presented in Eq. $(A85)$ are dominated by the mixtures $L_d = (p\colon L_s^*, 1-p\colon 0)$ and $L_d = (p\colon w_s^*(y), 1-p\colon 0)$ for any $p \in (0,1)$. This proves the third 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. $(A75)$ collapses to

$$\frac{1}{u'(w(y)) + \eta \int_{\underline{y}}^{\overline{y}} v'(u(w(y) - u(w(\tilde{\tilde{y}}))u'\,(w(y)f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}} = \mu, \qquad (A86)$$

if $\theta_{\mathbb{I}} = 1$, and Eq. $(A76)$ collapses to

$$\frac{1}{u'(w(y)) + \eta\lambda \int_{\underline{y}}^{\overline{y}} v'\left(u(w(\tilde{\tilde{y}}) - (u(w(y))\right)u'\,(w(y)f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}} = \mu, \qquad (A87)$$

if $\theta_{\mathbb{I}} = 0$. Eqs. $(A86)$ and $(A87)$ 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 $U(e_H, w(\tilde{y}), w(\tilde{\tilde{y}}))$ is S-shaped for any $\tilde{y}$, the principal is better off paying a lottery $L_f = (p\colon w_f^F(\tilde{y}), 1-p\colon 0)$ 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:

$$\mathbb{E}_y\left(u\left(w_f^F(y)\right)\right) - \eta\lambda \int_{\underline{y}}^{\tilde{\tilde{y}}} \int_{\underline{y}}^{\overline{y}} v\left(u\left(w_f^F(\tilde{\tilde{y}})\right) - u\left(w_f^F(y)\right)\right) f(\tilde{\tilde{y}}|e)\, \mathrm{d}\tilde{\tilde{y}} f(y|e)\mathrm{d}y$$
$$+ \eta \int_{\tilde{\tilde{y}}}^{\overline{y}} \int_{\underline{y}}^{\overline{y}} v\left(u\left(w_f^F(y)\right) - u\left(w_f^F(\tilde{\tilde{y}})\right)\right) f(\tilde{y}|e)\, \mathrm{d}\tilde{\tilde{y}} f(y|e)\mathrm{d}y$$
$$= p\, \mathbb{E}_y\left(u\left(w_f^F(\tilde{\tilde{y}})\right)\right) - p(1-p)\eta\lambda\left(\int_{\underline{y}}^{\overline{y}} v\left(-u\left(w_f^F(\tilde{\tilde{y}})\right)\right) \mathrm{d}y f(y|e)\mathrm{d}y\right)$$
$$+ p(1-p)\eta\left(\int_{\underline{y}}^{\overline{y}} v\left(u\left(w_f^F(\tilde{\tilde{y}})\right)\right) f(y|e)\mathrm{d}y\right). \qquad (A88)$$

Therefore, replacing $w_f^F(y)$ from Eq. $(A76)$ by L_f leaves the agent's participation constraint unchanged. Eq. $(A88)$ 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:

$$\mathbb{E}_y\left(u\left(w_f^F(y)\right)\right) - p\, \mathbb{E}_y\left(u\left(w_f^F(\tilde{\tilde{y}})\right)\right)$$
$$\geq \eta\lambda \int_{\underline{y}}^{\tilde{\tilde{y}}} \int_{\underline{y}}^{\overline{y}} v\left(u\left(w_f^F(\tilde{\tilde{y}})\right) - u\left(w_f^F(y)\right)\right) f(\tilde{\tilde{y}}|e)\, \mathrm{d}\tilde{\tilde{y}} f(y|e)\mathrm{d}yy$$
$$- \eta \int_{\tilde{\tilde{y}}}^{\overline{\overline{y}}} \int_{\underline{y}}^{\overline{\overline{y}}} v\left(u\left(w_f^F(y)\right) - u\left(w_f^F(\tilde{\tilde{y}})\right)\right) f(\tilde{y}|e)\, \mathrm{d}\tilde{\tilde{y}} f(y|e)\mathrm{d}y$$
$$- p\eta\lambda\left(\int_{\underline{y}}^{\overline{y}} v\left(-(1-p)u\left(w_f^F(\tilde{\tilde{y}})\right)\right) \mathrm{d}y f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}\right)$$
$$+ p\eta\left(\int_{\underline{y}}^{\overline{\overline{y}}} v\left((1-p)u\left(w_f^F(\tilde{\tilde{y}})\right)\right) f(\tilde{\tilde{y}}|e)\mathrm{d}\tilde{\tilde{y}}\right). \qquad (A89)$$

Since $\mathbb{E}_y\left(u\left(w_f^F(y)\right)\right) - p\, \mathbb{E}_y\left(u\left(w_f^F(\tilde{\tilde{y}})\right)\right) > 0$ and $v' > 0$, the last inequality implies $w_f^F(y) > pw_f^F(\tilde{\tilde{y}})$. Hence, L_f is more cost-effective for the principal than the candidate solution given by Eq. $(A86)$.

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:

$$U\left(e, L, w_f^F(\tilde{\tilde{y}})\right) = \left(\frac{\overline{L_f}}{w_f^F(\tilde{\tilde{y}})}\right) u\left(w_s^F(\tilde{\tilde{y}})\right) + \frac{\overline{L_f}}{w_f^F(\tilde{\tilde{y}})}\left(1 - \frac{\overline{L_f}}{w_f^F(\tilde{\tilde{y}})}\right)\eta\left(v\left(u\left(w_f^F(\tilde{\tilde{y}})\right)\right) - \lambda v\left(-u\left(w_f^F(\tilde{\tilde{y}})\right)\right)\right), \qquad (A90)$$

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. $(A90)$ 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 each $\tilde{y}$, the first-best contract, $w_f^*(y)$, consists of L_f^*. Suppose instead that $w_f^F(y)$ satisfying Eq. $(A88)$ is given everywhere. Since $w_f^F(y) > pw_f^F(\tilde{\tilde{y}})$, the principal can profitably deviate from that solution by paying L_f at the lower end of the output space. Because L_f^* is evaluated as a sizeable loss when $w_f^F(y)$ satisfying $(A86)$ is taken as reference point, the principal increases the segment for which L_f^* is the solution until the boundary $\hat{y}_{df} = \overline{y}$ is reached. This strategy is cost-effective. Also, proceeding in such way would not fully locate the agent in losses; gains are experienced when the outcome of the lottery $w = 0$ is adopted as 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. $(A88)$, which implies $w_f^F(y) \geq w_s^F(\tilde{\tilde{y}})$, and $w_f^F(y)$ satisfying Eq. $(A89)$, 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 $(A89)$ 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. $(A88)$ is given. This proves the fifth part of the Proposition.

The last part of the proposition follows again from the condition in Masatlioglu and Raymond (2016). ■

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 $L_k := (1 - p: w_s^*(y), p: 0)$ generates the following utility:

$$KR(L) = (1-p)u\left(w_s^*(y)\right) + p(1-p)\eta\left(v\left(u\left(w_s^*(y)\right)\right) + \lambda v\left(-u\left(w_s^*(y)\right)\right)\right). \qquad (A91)$$

Since the second derivative of Eq. $(A91)$ is $-2\eta\left(v(u(w_s^*(y)\ ))+\lambda v\left(u\left(-w_s^*(y)\right)\right)\right)$, a negative expression, the interior probability that maximizes utility is given by the closed-form solution $p_k^*(y)=\frac{1}{2}-\frac{u\left(w_s^*(y)\right)}{2\eta\left(\lambda v(-u(w_s^*(y)))+v(u(w_s^*(y)))\right)}$.

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

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 set of candidates for min-max reference point are $r = \min\{\max\{\overline{U}\}, \max\{w_s^*(y)\}\}$. Intuitively, rejecting the contract yields welfare $\overline{U}$ and the maximum that the contract pays, regardless of effort level, is $\max\{w_s^*(y)\}$.

Since the participation constraint binds at the optimum, then $\mathbb{E}\big(U(e, w_s^*(y), r)\big) = \overline{U}$. Also, because the incentive compatibility constraint binds, it must be that $\mathbb{E}\big(U(e, w_s^*(y), r)\big) < \mathbb{E}\big(U(r, \max\{w_s^*(y)\}, r)\big)$. Otherwise, the second-best contract would not implement rewards for high performance and thus would not incentivize high effort. Hence, $r = w_s^*(y)$

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

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 $r = w_s^*\big(y_p\big)$. 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:

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

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

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

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,

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

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:

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

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:*

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

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

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

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:

$$\mathcal{L} = \left(-\lambda_p\left(r_p + w(y) - S(y)\right)\right) f(y|e_H)$$
$$+\mu\left[\phi u(w(y) + \theta_{\mathbb{I}}\eta v(u(w(y)) - u(r))f(y|e_H) - \lambda(1-\theta_{\mathbb{I}})\eta v(u(r) - u(w(y)))f(y|e_H) - c - \overline{U}\right] \qquad (C1)$$
$$+\gamma\left[\phi u(w(y)(f(y|e_H) - f(y|e_L)) + \eta\theta_{\mathbb{I}} v\left(u(w(y)) - u(r)\right)(f(y|e_H) - f(y|e_L)) - \lambda(1-\theta_{\mathbb{I}})\eta v(u(r) - u(w(y)))(f(y|e_H) - f(y|e_L)) - c\right].$$

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

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

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

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

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

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

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:

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

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

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

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 λ_H faces the following adverse selection constraint,

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

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 λ_L faces the following adverse selection constraint,

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

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

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

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

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

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 λ_H mimicking the agent with λ_L derives the following utility for a given effort level $\hat{e}$,

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

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:

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

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 λ_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

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

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