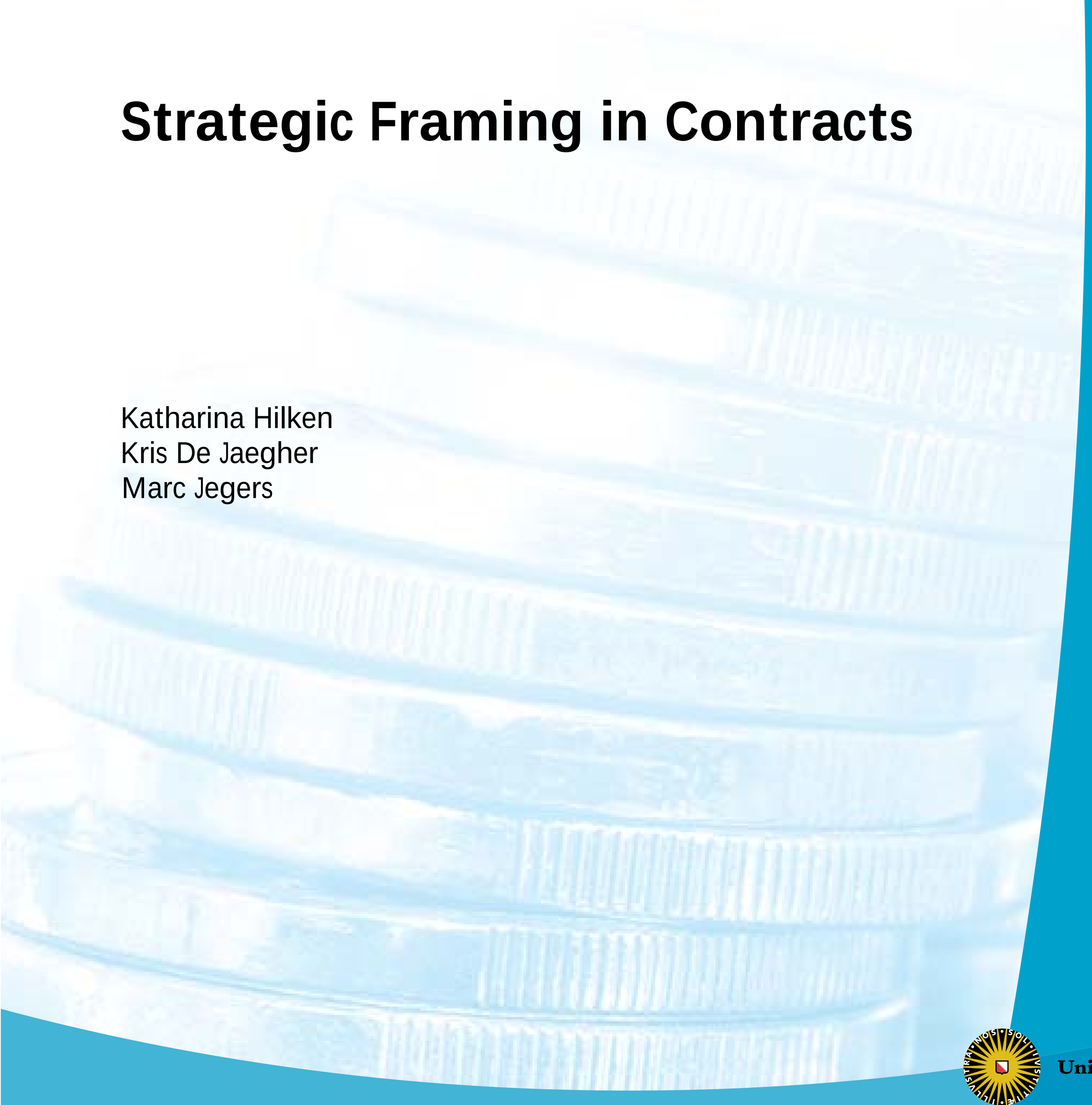
Tjalling C. Koopmans Research Institute

Tjalling C. Koopmans

Discussion Paper Series nr: 13-04



Strategic Framing in Contracts

Katharina Hilken
Kris De Jaegher
Marc Jegers

Universiteit Utrecht

Utrecht School of Economics

Tjalling C. Koopmans Research Institute
Utrecht School of Economics
Utrecht University

Kriekenpitplein 21-22
3584 EC Utrecht
The Netherlands
telephone +31 30 253 9800
fax +31 30 253 7373
website www.koopmansinstitute.uu.nl

The Tjalling C. Koopmans Institute is the research institute and research school of Utrecht School of Economics. It was founded in 2003, and named after Professor Tjalling C. Koopmans, Dutch-born Nobel Prize laureate in economics of 1975.

In the discussion papers series the Koopmans Institute publishes results of ongoing research for early dissemination of research results, and to enhance discussion with colleagues.

Please send any comments and suggestions on the Koopmans institute, or this series to J.M.vanDort@uu.nl

ontwerp voorblad: WRIK Utrecht

How to reach the authors

Please direct all correspondence to the first author.

Katharina Hilken
Vrije Universiteit Brussel
Department of Applied Economics
Pleinlaan 2
1050 Brussel
Belgium
E-mail: khilken@vub.ac.be

Kris De Jaegher
Utrecht University School of Economics
Utrecht University
Kriekenpitplein 21-22
3584 TC Utrecht
The Netherlands
E-mail: k.dejaegher@uu.nl

Marc Jegers
Vrije Universiteit Brussel
Department of Applied Economics
Pleinlaan 2
1050 Brussel
Belgium
E-mail: mjegers@vub.ac.be

This paper can be downloaded at: http:// www.uu.nl/rebo/economie/discussionpapers

Utrecht School of Economics
Tjalling C. Koopmans Research Institute
Discussion Paper Series 13-04

Strategic Framing in Contracts

Katharina Hilken[a]
Kris De Jaegher[b]
Marc Jegers[c]

[a]Department of Applied Economics
Vrije Universiteit Brussel

[b]Utrecht University School of Economics
Utrecht University

[c]Department of Applied Economics
Vrije Universiteit Brussel

March 2013

Abstract
We provide a hidden-action principal-agent model where the agent has reference-dependent preferences. The loss-averse agent considers the base wage as reference point, and bonuses and/or penalties as gains and losses, respectively. When choosing optimal payments, the principal strategically sets the base wage, knowing that this determines the agent's reference point. We consider two variants of the model. In a first variant, the agent's reservation utility is not reference-dependent. We show that it is always optimal in this case for the principal to employ bonuses. In a second variant, the reservation utility is reference-dependent and the principal may use penalties.

Keywords: Strategic Framing; Reference-Dependent Preferences; Principal-Agent Theory; Optimal Payment Schemes; Employment Contracts

JEL classification: D86, D03, J33, M52

Acknowledgements
We thank Bastian Westbrock, Stephanie Rosenkranz, Botond Köszegi, Ran Spiegler, Britta Hoyer, Steven De Schepper and Mitchell van Balen for useful comments and suggestions. Any remaining errors are our own.

1 Introduction

Consider a university that receives yearly subsidies based on its publication output of researchers. It is too costly for the university to monitor the number of hours each researcher effectively spends on writing papers, but to sustain the research staff it needs a certain amount of money. The more effort the individual researcher puts into writing papers, the higher the chances of publishing. If the university pays a fixed wage to the researcher, he will shirk, as effort is costly to him. Therefore, the university can base part of the remuneration of the researcher on the number of papers published, next to a base wage, invariant to performance. The goal of the design of payment schemes is to ultimately increase the efficiency and the academic performance of the university. How does the university determine the design of the variable pay for the researcher if his preferences are reference-dependent? How does the ability of the university to influence the reference point (strategic framing) of the loss averse researcher change the optimal payment scheme?

Rewarding faculty members for the number of publications is intensely discussed in the "publish-or-perish" literature (for a recent survey see Dalen & Henkens, 2012), especially the positive and negative effects of pressure on the quality of publications. This issue, as well as performance-based remuneration schemes of managers[1] and many other employment relations, falls into the scope of principal-agent theory[2]. Individual, team and executive reward systems, and especially payment plans, are the most fundamental forms of motivational strategies of firms (Griffin, 2011). In consequence, the design of employee compensation is a tool to achieve alignment of incentives with employers. Including reference-dependent preferences into these hidden-action problems gives other answers to the efficiency and design questions than standard theory.

A substantial amount of contradictions with economic theory have repeatedly been found in human decision making. In order to deal with some of these anomalies, Kahneman and Tversky (1979) have developed prospect theory as an alternative theory of decision making under risk. People form reference points and think in terms of gains and losses with respect to them. The losses experienced hurt the decision-maker more than equivalent gains give pleasure (loss aversion)[3], implying that the utility function is steeper in the loss region than in the gains region. As loss aversion even applies for small changes from the reference point, the utility function therefore has a kink at this reference point. The "reflection effect" means that an economic decision maker is only risk-averse in the domain of gains (with respect to the reference point); in the domain of losses though, he turns into a risk-seeking economic agent. While prospect theory does not have a formal theory of how the reference point is determined, the Asian Disease problem (Tversky & Kahneman, 1981) suggests that the mere wording of an otherwise identical choice problem (framing) can determine how agents form the reference point and perceive gains and losses[4].

De Meza and Webb (2007) have investigated the effect of prospect-theoretic preferences of the agent on the optimal payment scheme, but they assume a fixed reference point. We will

[1] During the financial crisis of 2007-2009 payment schemes of managers came under close scrutiny and criticism both with respect to their level as with respect to their sensitivity to firm performance (Edmans & Gabaix, 2009). Cuñat and Guadalupe (2009) find that, due to deregulation and increased competition in the banking and financial sector in the US, the fraction of performance-based pay in the total pay of executives increased significantly during the 1990s. Sample periods were 1993-1999 and 1995-2002, the periods after major deregulations in the financial and banking sectors in the USA.

[2] A general discussion of the standard principal-agent model can be found in Macho-Stadler and Pérez-Castrillo (2001), Laffont and Martimort (2002) and Mas-Colell, Whinston, and Green (1995). Bonner and Sprinkle (2002) summarize findings in the literature about the effects of monetary incentives on effort and task performance in general.

[3] A vast literature exists connected to loss aversion which will not be discussed in this paper. We will only give a concise selection of the most important papers in this area: the original definition is given by Kahneman and Tversky (1979); Kahneman, Knetsch, and Thaler (1991) find that loss aversion explains a large number of anomalies; experimental tests of loss aversion are presented by Schmidt and Traub (2002). Gächter, Johnson, and Herrmann (2007) look at differences in the loss aversion measure in risky and riskless choice experimentally.

[4] As noted by Kahneman and Tversky (1979, p.277-278), the prospect-theoretic value function developed on the basis of these anomalies, is a satisfactory approximation of the utility a decision-maker perceives, but it should generally be a function of two arguments: the wealth level (equalling the reference point) and size of the deviation from this reference point.

combine a concave utility function and a reference-dependent utility function that incorporates loss aversion. Our contribution is to consider the reference point as strategic as well. Accordingly, the choice of frame for the payment scheme will have an impact on the effort provision of the agent. In the context of the hidden action principal-agent problem, framing means expressing an otherwise completely identical net payment schedule in different ways, by splitting it up into a base wage combined with bonuses and penalties in several ways, where the base wage determines the reference point. Even if this does not affect the *net* payment obtained by the agent, and does not affect the principal's costs, framing an identical net payment as a low payment plus a large bonus, as an intermediate payment combined with a bonus or penalty, or as a high payment plus a large penalty, will matter for the agent's decision on whether or not to participate, and on which effort to provide when participating. Strategic framing is then the strategic setting of the base wage by the principal to influence the behavior of the agent with reference-dependent preferences.

We discuss two distinct cases, firstly reservation independence and secondly reservation dependence: consider again the researcher that decides between keeping the job at the university or quitting and staying at home. How the researcher perceives the gains and losses of staying at home might not be determined in monetary terms (but in a personal, fixed reservation utility), such that the university is not able to frame the perception of not participating. On the contrary, when the researcher contemplates changing to another university, the base wage comparison may determine how the researcher perceives gains and losses with respect to not participating. We find that when the agent is prone to reference point manipulation, using bonus contracts is optimal when the reservation utility is independent of the reference point. With reservation dependence, penalty contracts can become optimal. We therefore need to apply different variations of the model to match the current situation of the agent.

A common intuition is that setting penalties for bad performance in employment contracts has incentive effects on the agent (e.g. see De Meza & Webb, 2007). In reality though, it is observed that employers mainly implement contracts with bonuses for high performance rather than penalty contracts (Bebchuk & Grinstein, 2005). This is puzzling (Luft, 1994; Herweg, Müller, & Weinschenk, 2008) in the light of the common intuition, as one would expect that when payments are framed as losses, the agent would choose high effort levels in order to avoid losses as much as possible. With our model we can explain the fact that penalty contracts are not frequently used. This change of optimal contract, as explained in more detail at a later stage, is not due to incentive effects though.

In the subsequent section we introduce the related literature on reference-dependent principal-agent models, framing, reference point determinations as well as related experiments. We develop our model with an independent reservation utility in Section 3.1 and with a dependent reservation utility in Section 3.2. A simplified model with only two outcomes (Section 3.3), and a numerical example (Section 3.4), explain the intuition of our model. In Section 4, robustness checks are executed. Section 5 discusses possible extensions of the theoretical model and concludes.

2 Related literature

The nature of the reference point is highly discussed in the literature and is therefore also variously modelled. The assumption that the base wage serves as reference point in an employment contract is one strand: Brink (2008) recognizes that the base wage is "commonly interpreted as a guaranteed amount"(p.6). Similarly, Luft (1994) describes the agent's perception of the base wage as a "minimum guaranteed amount, almost an entitlement" (p.186). Armantier and Boly (2012) assume in their theoretical model for their experiment that the reference point is an increasing function of the base wage, so that the reference point is at least partially determined by the base wage which the principal sets.

The main objection to the base wage assumption is developed by Kőszegi and Rabin (2006) in a general model of reference-dependent preferences, where rational expectations of the distribution of outcomes (in this setting, the wages) of the agent serve as a stochastic reference point. They define the notion of a *personal equilibrium* to be a situation where the agent's decision given his current reference point, creates expectations consistent with this given reference point. Because of a self-fulfilling prophecy argument, there may be several personal equilibria, in which case the agent is assumed to choose his *preferred personal equilibrium*, i.e. the one that maximises his utility ex ante. An optimal employment contract will always include penalties *perceived* by the agent because of the stochastic nature of the reference point which is determined by expectations of the payments offered by the principal[5]. Daido and Itoh (2006) develop a basic principal-agent model based on Kőszegi and Rabin (2006) to explain the Pygmalion and Galatea effects. The former describes the effect that if the principal has higher expectations of an agent, this will result in higher performance of the agent. The Galatea effect refers to the internalization of these higher expectations, increasing what the agent expects of himself, which in turn increases performance.

Several other variables determining the reference point have been proposed in the literature, including the status-quo (usually interpreted as the agent's current wealth level, see Munro & Sugden, 2003), the agent's median income (De Meza & Webb, 2007), the agent's habitual level of past consumption (Wathieu, 2004), the goal set either by the principal or by the agent himself (Heath, Larrick, & Wu, 1999), the aspiration level of goals set for the future (Lopes & Oden, 1999), the outcomes achieved by peers (Falk & Ichino, 2003), or finally a combination of several such factors (Koop & Johnson, 2012). While we recognise that such variables might serve as reference point, in the setting of contracts though, the base wage is a natural reference point for our analysis.

Framing has been subject to experimental examination repeatedly in the literature: even if employment contracts have the same net payments (or are economically equivalent), the use of the word "bonus" gives a sensation of approval in contrast to the word "penalty"; the mere words convey value to the agent as non-monetary payoff (Luft, 1994). The results of the first of two experiments conducted by Luft (1994) show that there is a clear preference for bonus contracts and that framing is is beneficial for the principal: the expected payment of the bonus incentive contract, that is chosen over the flat payment, lies significantly below the expected payment that the penalty contract would have to offer. The meta-analysis by Levin, Schneider, and Gaeth (1998) finds a considerable amount of support for the framing effect in various domains, although specific characteristics diminish or eliminate this effect. Kühberger (1998) finds small to moderate effect sizes depending on the experimental design. On the one hand, Coursey, Hovis, and Schulze (1987) and Brookshire and Coursey (1987) show that framing diminishes in repeated principal-agent settings over time, with increased experience of participants. On the other hand, Luft (1994) explores the effectiveness of framing of bonus and penalty incentives in multiple periods and finds an even increasing effect over time. This finding is also supported by the field experiment of Hossain and List (2009), who also discover that framing is more effective in groups rather than for individuals.

After conducting an experiment on priming with outcome irrelevant information, Matthey (2008) suggests that "reference states are not fully determined 'within' the individual, but [that they] can relatively easily be manipulated from outside"(p.1). She finds evidence for reference-dependent preferences that are not solely dependent on outcomes, but also on environmental factors that have no influence on outcomes. We apply the concept of framing which, in contrast to priming, is connected to the actual decision task, but also does not have any influence on the outcomes. Priming and framing are closely related, priming though brings an issue to attention that has not been in focus and framing changes the way *how* an issue is thought about with different usage of wording (Scheufele & Tewksbury, 2007). The formation of a reference point happens in both cases and Matthey (2008) suggests that influencing it is simple. This gives a strong support for the effectiveness of framing.

Having thus treated experimental literature showing the importance of framing, we now treat related papers in the principal-agent literature. Models that allow for the fact that the agent may have reference-dependent preferences are rare. We utilise the model of De Meza and Webb (2007) in essence, allowing for the reference point to be set strategically. The closest theoretical formulation

[5] The principal may offer a pure bonus contract, for example with a low base wage and high bonuses, but the agent will always perceive the low outcome payment(s) as a loss because of the definition of the reference point.

of the principal-agent problem with reference-dependent preferences to our formulation is Just and Wu (2005)[6]. We point out the major differences with our model: as utility for the agent the authors use the pure gain-loss utility of prospect theory without an absolute utility part, where the agent is risk seeking in the loss region. In line with recent literature (see Kőszegi & Rabin, 2006), we set a standard "absolute" utility part with (linear) loss aversion below the reference point (and no additional gains). Hereby we ensure overall risk aversion of the agent (see Levy & Levy, 2002). This is consequential because as soon as one allows for risk lovingness, the principal can increase the willingness of the agent to participate by making the payment as risky as possible (De Meza & Webb, 2007). Because there is only a gain-loss utility and no absolute utility part in Just and Wu (2005), the marginal utility is maximised at the reference point. For this reason, a principal who wants to induce high effort by putting the marginal utility high, should always make sure that the effect of a marginal increase in effort is perceived to be around the reference point. From this it follows that the optimal payment is around the reference point when the reservation utility is reference-independent. In our model, where there is an absolute utility part, and where the overall utility function is always concave, marginal utility is not maximised around the reference point. This explains why it is optimal to set the base wage low with a reference-independent reservation utility. Furthermore, Just and Wu (2005) maximise with respect to the base wage holding the net payment fixed, rather than holding bonuses and penalties fixed. In this way, changing the base wage does not change net payments, and apart from the effort taken by the agent, does not change the principal's profits. As long as the participation constraint is slack, the principal should therefore increase the base wage, in order to induce maximal effort. In our model with two effort levels, we instead fix the effort level to be induced, and we maximise with respect to the base wage keeping bonuses and/or penalties fixed, so that the base wage does have a direct effect on the principal's profits. We show that with a reference-independent reservation utility the principal should set the base wage low. As this each time involves a binding participation constraint, loss framing is therefore never optimal in our model with a reference-independent reservation utility. In the model of Just and Wu, a bonus contract that is optimal with reservation independence, can include losses, as long as the *average* pay over all the states lies above the reference point. This is in contrast to our result, where *all* state-dependent payments lie above the strategically framed reference point.

A simple reference-dependent model, including loss aversion and diminishing sensitivity, but without uncertainty, is developed and tested experimentally in both the laboratory and the field by Armantier and Boly (2012). Performance of the participants is lowest without state-dependent incentives, followed by the bonus payment scheme, the penalty payment scheme and the highest performance is observed in the payment scheme that combines bonuses and penalties (contracts are economically equivalent). The mathematical as well as the experimental results show that using small penalties can enhance effort provision and large penalties may reduce it. Therefore, increasing the reference point (assumed to be increasing with the base wage) with strategic framing is only useful up to the extent to which the agent is able to achieve goals set by the principal. The experiment conducted is in line with their theoretical predictions. One major drawback of the model formulation is that the authors do not consider whether the agent (theoretically) wants to participate in the contract offered. As discussed earlier, the literature has shown that agents prefer bonus contracts. Employing penalties might be enhancing the effort provision of participants, but also might be counterproductive in that agents may no longer be willing to participate at all.

3 The model

3.1 Independent reservation utility

The reference-dependent principal-agent model of De Meza and Webb (2007) considers optimal incentive payments based on the assumption that the reference point, Y^R, is fixed. We introduce strategic framing, e.g. the principal's (her) ability to influence the agent's (his) reference point.

[6] They use the first-order approach for continuous effort (see Rogerson, 1985).

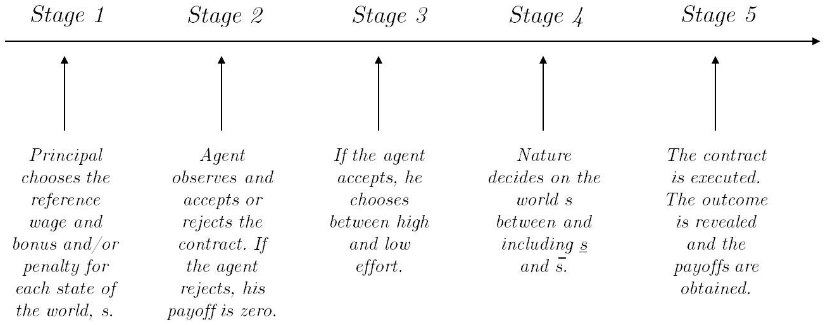

Stage 1
Stage 2
Stage 3
Stage 4
Stage 5
Principal chooses the reference wage and bonus and/or penalty for each state of the world, s.
Agent observes and accepts or rejects the contract. If the agent rejects, his payoff is zero.
If the agent accepts, he chooses between high and low effort.
Nature decides on the world s between and including $\underline{s}$ and $\bar{s}$.
The contract is executed. The outcome is revealed and the payoffs are obtained.

Figure 1: *Timeline of the principal-agent problem.*

The time schedule in Figure 1 depicts the different decision stages of the game. Each of the stages is explained in more detail in the following paragraphs.

At *stage 1*, the risk-neutral principal (she) designs a contract $Y(s)$ that she offers to the risk-averse agent (he). The design comprises the choice of a reference income Y^R, or interchangeably, the base wage, being the same for all states, and a piece rate $b(s)$, which is allowed to be positive or negative for each state of nature s. We assume hereafter that the principal wants to induce high effort $\bar{e}$. The agent decides to accept or reject the contract at *stage 2*. The agent accepts if the total (expected) utility is higher than his reservation utility, V^*, he rejects otherwise.

If the agent accepts the contract, at *stage 3* he chooses an effort $e \in \{\underline{e}, \bar{e}\}$, which is not verifiable by the principal. Depending on the effort e taken by the agent at *stage 3*, Nature determines the state of the world s at according to the conditional cumulative distribution function $F(s|e)$ and the resulting conditional probability density function $f(s|e)$ at *stage 4*. Because of the distribution function, effort and output are imperfectly correlated. The principal can only verify the state of the world $s \in [\underline{s}, \bar{s}]$. The uncertainty lies in the principal's inability to observe the exerted effort of the agent. Importantly, the state of the world s is imperfectly correlated to the effort level the agent provided through the conditional cumulative distribution function $F(s|e)$ and the resulting conditional probability density function $f(s|e)$, which we assume to be twice continuously differentiable. Each state s has strictly positive probability of occurrence for both effort levels, $f(s|e) > 0$. We assume that with increasing state realizations s, the probability conditional on low effort $\underline{e}$ relative to the probability conditional on high effort $\bar{e}$, $\frac{f(s|\underline{e})}{f(s|\bar{e})}$, is decreasing (monotone likelihood ratio property, MLRP). The optimal payment scheme is then monotonically increasing in the state realization s. The intuition behind this assumption is that the probability of occurrence of each state of nature depends positively, but not perfectly, on the effort of the agent. The higher state of nature s is observed, the higher the likelihood that this is because of high effort given by the agent rather than because of low effort. Additionally, we assume that the likelihood ratio is continuously decreasing.

At *stage 5*, the principal and the agent receive their payoffs according to the obtained outcome. We assume the simplest case, where the principal is risk neutral and therefore her utility is linear in s. She maximises her profits over all possible states of nature, $\int_{\underline{s}}^{\bar{s}} \big(s - Y(s)\big) f(s) ds$, with $Y(s) = Y^R + b(s)$ being the payment to the agent. The overall utility the agent derives from his payoff, is a combination of standard concave, twice continuously differentiable, Bernoulli utility, $U(Y(s))$, and reference-dependent utility with loss aversion, $Z(Y^R, b(s))$. Note that the principal will only offer contracts that are optimal, meaning that summed over all states of the world the offered payment scheme maximises the principal's expected profits and will give the required incentives to the agent. This reasoning is based on the fact that the principal has already considered the participation and incentive compatibility constraints of the agent and she chose the contract that is the best to be offered. The maximisation problem with respect to profits is equivalent to minimising the cost of employing the agent for a pre-specified effort level (Mas-Colell et al., 1995,

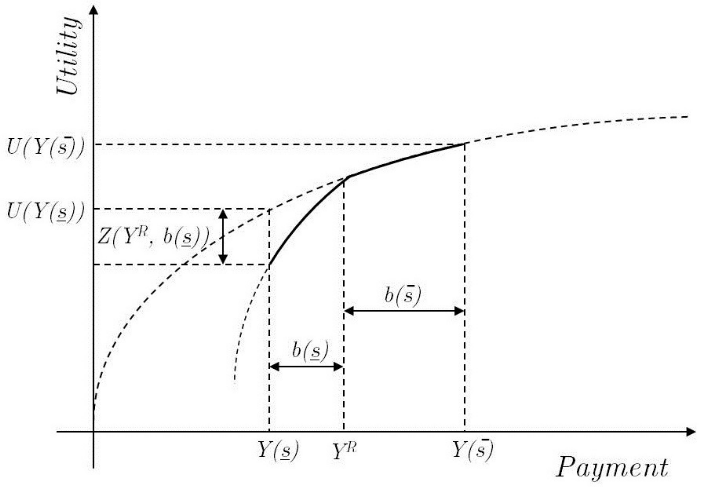


Figure 2: *Bernoulli utility function of the agent net of the cost of e.*

p.480).

$$\int_{\underline{s}}^{\bar{s}} Y(s) f(s|\bar{e}) ds \tag{1}$$

The agent's overall utility, incurring the cost of effort, which is denoted as $c(e)$ and assumed to be strictly increasing in effort, is then equal to:

$$U\big(Y(s)\big) - Z\big(Y^R, b(s)\big) - c(e) \tag{2}$$

with

$$Z\big(Y^R, b(s)\big) = \begin{cases} 0 & \text{if } \; Y(s) > Y^R \\ g\Big[U(Y^R) - U\big(Y^R + b(s)\big)\Big] & \text{if } \; Y(s) \leq Y^R \end{cases} \tag{3}$$

For a reference-dependent utility with a linear loss aversion coefficient, l, this becomes:

$$g\Big[U(Y^R) - U\big(Y^R + b(s)\big)\Big] = l\Big[U(Y^R) - U\big(Y^R + b(s)\big)\Big] \tag{4}$$

where $l > 0$ if $Y^R + b(s) = Y(s) \leq Y^R$.

With linear loss aversion, the overall utility function of the agent is concave over the whole range of possible payments with the assumptions $U'' < 0$ and $U' > 0$. Figure 2 shows that combining these two utility functions creates a kink at the reference income, but preserves the concavity of the overall utility of the agent. The agent is therefore risk-averse over the whole range of possible payments, by assumption. Through the kink in the reference-dependent utility function (and subsequently also in the overall utility) of the agent at the reference income, as we will show, the optimal incentive scheme has a region where the payment is independent of performance.

The Inada condition ($lim_{Y \to 0} U'(Y) = \infty$) ensures that the principal will never offer the agent a wage lower than or equal to zero in one of the states. This is because the marginal increase in income to the agent has such a great utility effect in such a case that the principal can reduce the income in another state drastically without making the agent worse off. Therefore, no payment to the agent is ever profit-maximizing for the principal. The Inada condition is necessary for an interior solution.

The mathematical derivation of the stages 1 to 5, as described above, follows. The principal minimises his costs of employing the agent subject to the participation and incentive compatibility constraints (PC and IC, respectively) of the agent, given that high effort $\bar{e}$ is to be elicited (with

$\Delta c(e) = c(\bar{e}) - c(\underline{e})$ and $\Delta f(s|e) = f(s|\bar{e}) - f(s|\underline{e})$):

$$\min_{b(s),Y^R} Y^R + \int_{\underline{s}}^{\bar{s}} b(s) f(s|\bar{e}) ds \tag{5}$$

s.t.

$$[PC] \qquad \int_{\underline{s}}^{\bar{s}} \Big[U\big(Y^R+b(s)\big) - \theta l\Big(U(Y^R) - U\big(Y^R+b(s)\big)\Big)\Big] f(s|\bar{e})ds - c(\bar{e}) \geq V^* \tag{6}$$

$$[IC] \qquad \int_{\underline{s}}^{\bar{s}} \Big[U\big(Y^R+b(s)\big) - \theta l\Big(U(Y^R) - U\big(Y^R+b(s)\big)\Big)\Big] \Delta f(s|e)ds \geq \Delta c(e) \tag{7}$$

θ is an indicator function used in the first-order conditions that ensures losses enter only if the outcome dependent on the state s, lies below the reference income:

$$\theta = \begin{cases} 0 & \text{if } Y(s) \geq Y^R \\ 1 & \text{if } Y(s) < Y^R \end{cases} \tag{8}$$

This indicator ensures that the utility that the agent derives from the payment he receives is equal to the standard expected utility above the reference point, and equal to a concave reference-dependent utility function (which is the combination of standard utility and a function including loss aversion) below the reference point. We call this the principal-agent model with a loss averse agent with reference-dependent preferences with a reference-independent reservation utility.

Lagrangian optimisation gives first-order conditions (FOCs) (9) - (12), with γ and λ being the multipliers of the participation constraint and the incentive compatibility constraint, respectively.

$$\int_{\underline{s}}^{\bar{s}} \Big[U\big(Y^R + b(s)\big) - \theta l\Big(U(Y^R) - U\big(Y^R + b(s)\big)\Big)\Big] f(s|\bar{e})ds - c(\bar{e}) \geq V^* \tag{9}$$

$$\int_{\underline{s}}^{\bar{s}} \Big[U\big(Y^R + b(s)\big) - \theta l\Big(U(Y^R) - U\big(Y^R + b(s)\big)\Big)\Big] \Delta f(s|e)ds \geq \Delta c(e) \tag{10}$$

$$f(s|\bar{e}) - \gamma U'\big(Y^R + b(s)\big)(1 + \theta l) f(s|\bar{e}) - \lambda U'\big(Y^R + b(s)\big)(1 + \theta l)\Delta f(s|\bar{e}) \geq 0 \tag{11}$$

$$\begin{aligned} 1 - \gamma \int_{\underline{s}}^{\bar{s}} (1 + \theta l) U'\big(Y^R + b(s)\big) f(s|\bar{e})ds - \lambda \int_{\underline{s}}^{\bar{s}} (1 + \theta l) U'\big(Y^R + b(s)\big)\Delta f(s|e)ds \\ + \gamma U'(Y^R) \int_{\underline{s}}^{\bar{s}} \theta l f(s|\bar{e})ds + \lambda U'(Y^R) \int_{\underline{s}}^{\bar{s}} \theta l \Delta f(s|e)ds \geq 0 \end{aligned} \tag{12}$$

The intuition behind FOC (12) is that increasing the reference income of the agent in the loss region, from the perspective of the participation constraint, causes a cost to the principal as she needs to offer higher payments to make the agent participate. While the LHS of first-order condition (12) is increasing in the reference income in the loss region (for a risk averse agent), the LHS of first-order condition (9) decreases with the reference income. Inequality (11) leads to an extended form of the standard solution for incentive payments:

$$\frac{1}{U'\big(Y^R + b(s)\big)} \geq (1 + \theta l)\left[\gamma + \lambda \frac{\Delta f(s|e)}{f(s|\bar{e})}\right] \tag{13}$$

Expression (13) should be met with equality for all performance outcomes s for the incentive payments to be optimal. Through the kink in the overall utility of the agent though, which is due to loss aversion below the reference income, expression (13) is only met with equality for performance outcomes s below a certain threshold level and above a certain threshold level[7]. The incentive payment at each of the two threshold levels is exactly the reference income. Between these threshold levels, expression (13) is only met with inequality and we have a flat part in the payment scheme for which the reference income is paid. Proposition 1 looks at the possible incentive schemes with a fixed reference income, $\bar{Y}^R$, in a more detailed way than in De Meza and Webb (2007) to allow for an immediate and clear proof of Proposition 2, which allows for strategic framing.

Proposition 1. *Consider the principal-agent model with a loss averse agent with reference-dependent preferences whose reservation utility is independent. Five shapes of the optimal incentive schemes are possible as a function of each level of the fixed reference income,* $\bar{Y}^R$*.*

1. *Bonus contract:*
   If $\frac{1}{U'(\bar{Y}^R)} \leq \gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})}$*, the payment scheme is continuous, strictly increasing in performance and only bonuses are paid; all* $b(s) \geq 0$*.*

2. *Penalty contract:*
   If $\frac{1}{U'(\bar{Y}^R)} \geq (1+l)\Big[\gamma + \lambda \frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})}\Big]$*, the payment scheme is continuous, strictly increasing in performance and only penalties are paid; all* $b(s) \leq 0$*.*

3. *Bonus contract with a flat segment at* $\bar{Y}^R$*:*
   If $\gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})} \leq \frac{1}{U'(\bar{Y}^R)} < (1+l)\Big[\gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})}\Big]$*, the scheme pays the reference income* $\bar{Y}^R$ *up to some threshold beyond which it is continuous, strictly increasing in performance.*

4. *Penalty contract with a flat segment at* $\bar{Y}^R$*:*
   If $\gamma + \lambda \frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})} < \frac{1}{U'(\bar{Y}^R)} \leq \gamma + \lambda \frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})}$*, the payment scheme is continuous, strictly increasing up to some threshold, beyond which the reference income* $\bar{Y}^R$ *is paid.*

5. *Bonus and penalty contract:*
   If $(1+l)\Big[\gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})}\Big] < \frac{1}{U'(\bar{Y}^R)} < \gamma + \lambda \frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})}$*, the reference income* $\bar{Y}^R$ *is paid for the performance interval between and including the threshold levels* s' *and* $\hat{s}$*. Below and above these threshold levels respectively, the payment scheme is continuous, strictly increasing.*

Proof.
Step 1 shows that the high effort is not induced if the payments to the agent are insensitive to performance over the whole range of s, as this would be incentive incompatible. Step 2 looks at the case where the first-order condition (13) is exactly zero at $\bar{Y}^R$. Step 3 derives the several possible minima, and in part applies Step 2.

Step 1. If the following inequality is satisfied for a given $\bar{Y}^R$, the payment to the agent is not incentive incompatible:

$$\gamma + \lambda \frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})} \leq \frac{1}{U'(\bar{Y}^R)} \leq (1+l)\left[\gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})}\right] \tag{14}$$

By definition, a variable payment $b(\underline{s})$ is non-positive if framed as a loss, and $b(\bar{s})$ is non-negative if framed as a gain. For a payment scheme to be incentive compatible, (at least) the lowest outcome payment needs to be smaller than the highest outcome payment, e.g. $b(\underline{s}) < b(\bar{s})$, or we have a monotonically increasing payment scheme which is ensured by MLRP. In parallel, expression (13) has to be satisfied for these payments. For inequality (14) to apply, it would have to be true that $b(\underline{s}) = 0$ and $b(\bar{s}) = 0$. In this case, $\bar{Y}^R$ would be paid in all states s and the payment scheme would be independent of performance, or a flat, fixed payment to the agent. This is not incentive compatible and can therefore never be an optimal solution to the minimisation problem.

[7] At the threshold levels expression (13) is still met with equality. We define these threshold levels in more detail below.

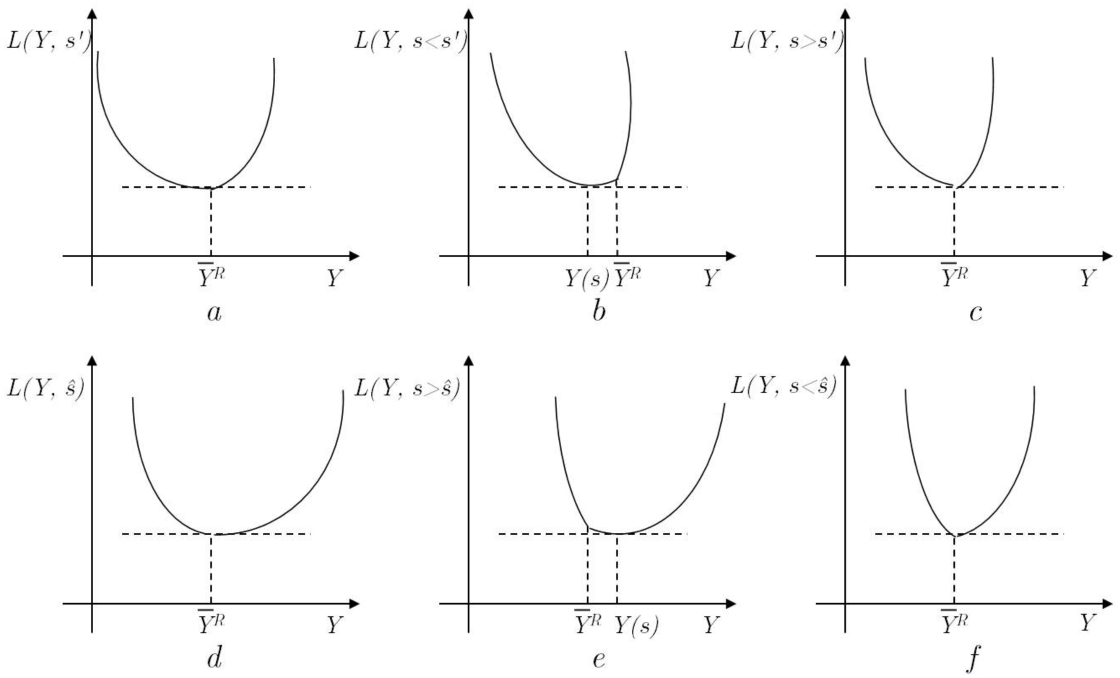


Figure 3: *Proposition 1.*

Step 2. Description of the behaviour of the Lagrangian around the threshold levels s' and $\hat{s}$, if they exist. We are considering infinitely small changes from each state to determine the optimal payments, first with a fixed reference point, $\bar{Y}^R$. Reformulating expression (13) gives the following:

$$\frac{1}{U'\big(Y^R + b(s)\big)} - (1 + \theta l)\left[\gamma + \lambda \frac{\Delta f(s|e)}{f(s|\bar{e})}\right] \geq 0 \tag{15}$$

Let there be an s' with $b(s') = 0$ such that:

$$\frac{1}{U'(\bar{Y}^R)} - (1 + l)\left[\gamma + \lambda \frac{\Delta f(s'|e)}{f(s'|\bar{e})}\right] = 0 \tag{16}$$

For the state s', we are at the minimum because for any deviation from the payment of $\bar{Y}^R$ equation (16) is different from zero: by making $b(s)$ negative, the derivative turns negative and the costs can be decreased by increasing the variable payment. By setting $b(s) > 0$, the second term of equation (16) decreases discretely, as moving to the gains region lets the loss aversion measure disappear. The derivative is positive now and decreasing the bonus equivalently decreases the costs of the principal. Figure 3 depicts the Lagrangian minimand $L(Y, s)$ as a function of Y for fixed s. The above described case is represented in Figure 3a.

Now for $s < s'$, the second term of equation (16) decreases by the MLRP and the FOC (16) is positive at $\bar{Y}^R$, approaching from the left. Decreasing $b(s)$ then ensures that the derivative continues to be zero, the minimum of the Lagrangian lies below the reference income (see Figure 3b). Therefore, for states below the threshold level s', penalties are paid, e.g. $b(s) < b(s') = 0$, and the payment scheme is continuous, strictly increasing up to s'.

Also for a state s just above s', and still below $\hat{s}$, e.g. $\hat{s} > s > s'$ (depicted in Figure 3c), paying $Y(s') < \bar{Y}^R$, turns the FOC (16) negative by the MLRP, approaching $\bar{Y}^R$ from the left. Increasing the payment is cost-minimising. On the contrary, paying $Y(s') > Y^R$ turns the FOC (16) positive, approaching $\bar{Y}^R$ from the right. As the payment is in the gains region, the loss aversion measure suddenly disappears and the drop of the second part is not compensated for by the increase by the MLRP. This happens exactly at $\hat{s}$.

Let there be an $\hat{s}$ with $b(\hat{s}) = 0$ such that:

$$\frac{1}{U'(\bar{Y}^R)} - \left[\gamma + \lambda \frac{\Delta f(\hat{s}|e)}{f(\hat{s}|\bar{e})}\right] = 0 \tag{17}$$

For the state $\hat{s}$, we are at the minimum because for any deviation from the payment of Y^R equation (17) is different from zero: by making $b(s)$ positive, the derivative turns positive and the costs can be decreased by decreasing the variable payment. By setting $b(s) < 0$, the second term of equation (17) increases discretely, as moving to the loss region lets the loss aversion measure appear. The derivative is negative now and increasing the variable payment is equivalently decreasing the costs of the principal. This case is presented in Figure 3d.

Now, for cases where $s > \hat{s}$, depicted in Figure 3e, by the MLRP the first derivative in equation (17) is negative if $Y(s) = \bar{Y}^R$. Increasing the payment for s is cost-minimising as the FOC (17) continues to be equal to zero. Therefore, $b(s) > b(\hat{s}) = 0$ and the payment scheme is continuous, strictly increasing for any state $s > \hat{s}$.

For states just below $\hat{s}$, but still above s', the second part of equation (17) decreases by MLRP, implying that for $\bar{Y}^R$, approaching from the right, the FOC (17) is positive (see Figure 3). Consequently, the payment should be decreased. For payments just below the reference income, the second part of the equation discretely increases through the appearance of loss aversion and turns the FOC (17) negative. The optimal payment is therefore the reference income for states s for which it is true that $s' < s < \hat{s}$. From this it follows that there is a flat region in the payment scheme.

Step 3. The payment schemes with different reference incomes, $\bar{Y}^R$:

1. Bonus contract:
   If $\frac{1}{U'(\bar{Y}^R)} - \left[\gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})}\right] \leq 0$, each payment $Y(s)$ must be larger than $\bar{Y}^R$, therefore all $b(s) > 0$ and we have the situation as depicted in Figure 3e for all s. The states s' and $\hat{s}$ are not in $s \in [\underline{s}, \bar{s}]$.

2. Penalty contract:
   If $\frac{1}{U'(\bar{Y}^R)} - (1+l)\left[\gamma + \lambda \frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})}\right] \leq 0$, each payment $Y(s)$ must be smaller than $\bar{Y}^R$ for all s, therefore all $b(s) < 0$ and we have the situation as depicted in Figure 3b. The states s' and $\hat{s}$ are not in $s \in [\underline{s}, \bar{s}]$.

3. Bonus contract with a flat segment at $\bar{Y}^R$:
   If $\left[\gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})}\right] < \frac{1}{U'(\bar{Y}^R)} \leq (1+l)\left[\gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})}\right]$ the situation can be described as in Figures 3c and f for states s above and below the threshold levels. The reference income is paid for all states from $\underline{s}$ to $\hat{s}$ and the payment scheme is increasing above $\hat{s}$.

4. Penalty contract with a flat segment at $\bar{Y}^R$:
   If $\left[\gamma + \lambda \frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})}\right] \leq \frac{1}{U'(\bar{Y}^R)} < (1+l)\left[\gamma + \lambda \frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})}\right]$, the situation can be described as in Figures 3c and f. The reference income is paid for all states from s' to $\bar{s}$ and the payment scheme is increasing below s'.

5. Bonus and penalty contract:
   If $(1+l)\left[\gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})}\right] < \frac{1}{U'(\bar{Y}^R)} < \left[\gamma + \lambda \frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})}\right]$, the reference income is paid for all states between s' and $\hat{s}$ and the payment scheme is continuous, strictly increasing for all states $s \in [\underline{s}; s']$ and $s \in [\hat{s}; \bar{s}]$.

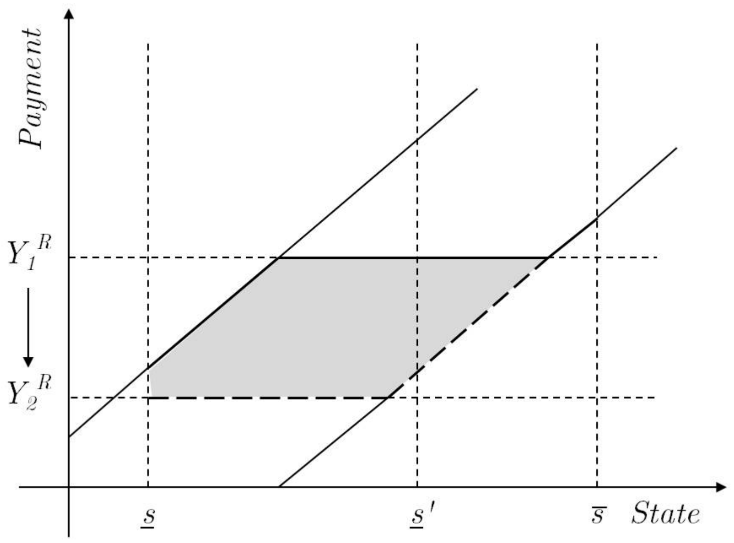


Figure 4: *Proposition 2.*

□

An intuition for the cases treated in Proposition 1 can be found in Figure 4. Ignore the reference point for the moment, and consider the net payment schedule for which expression (13) is valid with equality for gains ($\theta = 0$), and the net payment schedule for which expression (13) is met with equality for losses ($\theta = 1$), without putting a lower or higher bound on s. Whatever the reference point, and whatever $\underline{s}$ and $\bar{s}$, these net payment schedules do not change. By adding any horizontal line at the height of the given base wage, we can now graphically derive s' and $\hat{s}$. The shape of the optimal net payment schedule for a given base wage is finally determined by the position of $\underline{s}$ and $\bar{s}$ with respect to s' and $\hat{s}$, thus making only particular parts of the drawn net payment schedule relevant. Consider the change from $\underline{s}$ to an alternative low outcome $\underline{s}'$: the optimal payment scheme changes from an option-like scheme to a strictly increasing one without a flat region (as the reference income is not included in the payment scheme).

We describe the intuition of strategically framing a base wage that is optimal, anticipating on Proposition 2, in the following paragraphs. Assuming a variable reference income, Y^R, and the possibility of strategic framing, the optimal payment scheme is a low reference income and bonuses only. With the strictly increasing parts of the payment scheme being fixed, the principal is minimising her costs by decreasing the reference income as far as possible. Consider a low reference income (Y_2^R) and a high reference income (Y_1^R) in Figure 4. The payments to the agent with a low reference income are either lower or equal to the ones with a high reference income. Depending on how low the reference income can be set, and the definition of the possible states, the optimal payment scheme is given. If there is no institutional constraint on the minimum base wage, then putting Y^R such that all payments are seen as bonuses is always optimal. If the base wage is set accordingly, it's level is no longer of importance.

Suppose the principal considers setting the base wage at Y_1^R. With the possible states in the range from $\underline{s}$ to $\bar{s}$, it is self-evident that decreasing the reference income to Y_2^R, means lower costs for the principal with still an optimal incentive scheme. The grey area in Figure 4 measures how much the individual payments over all states decrease with decreasing the reference point. Weighing the vertical distance corresponding to each s by the density matching this s, one can calculate the principal's cost savings of lowering the base wage to Y_2^R. In each state s the principal pays the same or less to the agent (in at least one of the states she pays less). This intuition is expressed formally in the following proposition.

Proposition 2. *In the principal-agent model with reservation independence, the principal never frames payments as losses. All payments take the form of a reference income plus a bonus (such* *that the condition in Proposition 1 is satisfied). Once this is true, the level of the reference income does not matter.*

Proof.

Given that we know the form of the optimal incentive scheme for each possible Y^R, we can rewrite the participation constraint and the incentive compatibility constraint, allowing us to take the derivative with respect to Y^R. We separate this constraint that has been derived already for Proposition 1 for three sets of states: from $\underline{s}$ to s' (loss region), from s' to $\hat{s}$ (states at the reference income) and from $\hat{s}$ to $\bar{s}$ (gains region). We show that the FOC with respect to Y^R is positive everywhere, meaning that the principal's costs always increase for higher Y^R. How the threshold levels s' and $\hat{s}$ are affected by Y^R, and how they are determined is implicitly given by equations (16) and (17). To accommodate the existence of these threshold levels we apply the Leibniz integral rule in FOC (21). The first-order conditions for $b(s)$ are already derived in the proof of Proposition 1.

$$min_{b(s),Y^R} Y^R + \int_{\underline{s}}^{\bar{s}} b(s) f(s|\bar{e}) ds \tag{18}$$

s.t.

$$\begin{aligned}[PC] \qquad & \int_{\underline{s}}^{s'} \Big[U\big(Y^R + b(s)\big) - l\Big(U\big(Y^R\big) - U\big(Y^R + b(s)\big)\Big)\Big] f(s|\bar{e}) ds + U(Y^R) \int_{s'}^{\hat{s}} f(s|\bar{e}) ds \\ & + \int_{\hat{s}}^{\bar{s}} \Big[U\big(Y^R + b(s)\big)\Big] f(s|\bar{e}) ds - c(\bar{e}) \geq V^* \end{aligned} \tag{19}$$

$$\begin{aligned}[IC] \qquad & \int_{\underline{s}}^{s'} \Big[U\big(Y^R + b(s)\big) - l\Big(U\big(Y^R\big) - U\big(Y^R + b(s)\big)\Big)\Big] \Delta f(s|e) ds + U(Y^R) \int_{s'}^{\hat{s}} \Delta f(s|e) ds \\ & + \int_{\hat{s}}^{\bar{s}} \Big[U\big(Y^R + b(s)\big)\Big] \Delta f(s|e) ds \geq \Delta c(e) \end{aligned} \tag{20}$$

$$\begin{aligned} 1 - \gamma \Bigg[& \int_{\underline{s}}^{s'} \Big[U'\big(Y^R + b(s)\big) - l\Big(U'\big(Y^R\big) - U'\big(Y^R + b(s)\big)\Big)\Big] f(s|\bar{e}) ds \\ & + \Big[U\big(Y^R + b(s')\big) - l\Big(U\big(Y^R\big) - U\big(Y^R + b(s')\big)\Big)\Big] f(s'|\bar{e}) \frac{\partial s'}{\partial Y^R} \\ & + U'(Y^R) \int_{s'}^{\hat{s}} f(s|\bar{e}) ds - U(Y^R) f(s'|\bar{e}) \frac{\partial s'}{\partial Y^R} + U(Y^R) f(\hat{s}|\bar{e}) \frac{\partial \hat{s}}{\partial Y^R} \\ & + \int_{\hat{s}}^{\bar{s}} \Big[U'\big(Y^R + b(s)\big)\Big] f(s|\bar{e}) ds - U\big(Y^R + b(\hat{s})\big) f(\hat{s}|\bar{e}) \frac{\partial \hat{s}}{\partial Y^R} \Bigg] \\ - \lambda \Bigg[& \int_{\underline{s}}^{s'} \Big[U'\big(Y^R + b(s)\big) - l\Big(U'\big(Y^R\big) - U'\big(Y^R + b(s)\big)\Big)\Big] \Delta f(s|e) ds \\ & + \Big[U\big(Y^R + b(s')\big) - l\Big(U\big(Y^R\big) - U\big(Y^R + b(s')\big)\Big)\Big] \Delta f(s'|e) \frac{\partial s'}{\partial Y^R} \\ & + U'(Y^R) \int_{s'}^{\hat{s}} \Delta f(s|e) ds - U(Y^R) \Delta f(s'|e) \frac{\partial s'}{\partial Y^R} + U(Y^R) \Delta f(\hat{s}|e) \frac{\partial \hat{s}}{\partial Y^R} \\ & + \int_{\hat{s}}^{\bar{s}} \Big[U'\big(Y^R + b(s)\big)\Big] \Delta f(s|e) ds - U\big(Y^R + b(\hat{s})\big) \Delta f(\hat{s}|e) \frac{\partial \hat{s}}{\partial Y^R} \Bigg] \geq 0 \end{aligned} \tag{21}$$

For all s in the strictly increasing parts of the payment schedule, the first-order condition with respect to $b(s)$, inequality (11), equals zero:

$$\begin{aligned} &f(s|\bar{e}) - \gamma\Big[U'\big(Y^R + b(s)\big) - \theta l\Big(-U'\big(Y^R + b(s)\big)\Big)\Big]f(s|\bar{e}) \\ &-\lambda\Big[U'\big(Y^R + b(s)\big) - \theta l\Big(-U'\big(Y^R + b(s)\big)\Big)\Big]\Delta f(s|e) = 0 \end{aligned} \tag{22}$$

Expressing equation (22) differently, taking the integral over the range of s in the loss region and in the gains region, respectively:

$$\begin{aligned} &\int_{\underline{s}}^{s'} f(s|\bar{e})ds - \gamma\int_{\underline{s}}^{s'}\Big[U'\big(Y^R + b(s)\big) - l\Big(-U'\big(Y^R + b(s)\big)\Big)\Big]f(s|\bar{e})ds \\ &-\lambda\int_{\underline{s}}^{s'}\Big[U'\big(Y^R + b(s)\big) - l\Big(-U'\big(Y^R + b(s)\big)\Big)\Big]\Delta f(s|e)ds = 0 \end{aligned} \tag{23}$$

$$\int_{\hat{s}}^{\bar{s}} f(s|\bar{e})ds - \gamma\int_{\hat{s}}^{\bar{s}}\Big[U'\big(Y^R + b(s)\big)\Big]f(s|\bar{e})ds - \lambda\int_{\hat{s}}^{\bar{s}}\Big[U'\big(Y^R + b(s)\big)\Big]\Delta f(s|e)ds = 0 \tag{24}$$

Substituting equations (23) and (24) into FOC (21) and given that $b(s') = 0$ and $b(\hat{s}) = 0$ gives the following inequality:

$$\begin{aligned} &1 - \int_{\underline{s}}^{s'} f(s|\bar{e})ds - \int_{\hat{s}}^{\bar{s}} f(s|\bar{e})ds \\ &-\gamma\Big[-lU'(Y^R)\int_{\underline{s}}^{s'} f(s|\bar{e})ds + U'(Y^R)\int_{s'}^{\hat{s}} f(s|\bar{e})ds\Big] \\ &-\lambda\Big[-lU'(Y^R)\int_{\underline{s}}^{s'} \Delta f(s|e)ds + U'(Y^R)\int_{s'}^{\hat{s}} \Delta f(s|e)ds\Big] \geq 0 \end{aligned} \tag{25}$$

From FOC (11) we know that expression (25) is true as for all s at Y^R the following inequality holds[8], which is also true if integrated over the states s' to $\hat{s}$:

$$f(s|\bar{e}) - \gamma\Big[U'(Y^R)f(s|\bar{e})\Big] - \lambda\Big[U'(Y^R)\Delta f(s|e)\Big] > 0 \tag{26}$$

This proves that Y^R should be put as low as possible, meaning that optimally, the base wage should be framed as low as possible and the state dependent payments are all framed in the form of bonuses. If one assumes additionally that the base wage cannot be negative, then setting Y^R equal to zero is optimal.

□

3.2 Dependent reservation utility

We now look at the case where the reservation utility of the agent is dependent on the reference income given by the principal. For a fixed outside option of value P, it is the reference income Y^R that determines how the reservation utility V^* is perceived. The principal needs to take this reference-dependent reservation utility into account when offering a contract to the agent. We will therefore define the reservation utility as follows, keeping the assumption of linear loss aversion:

$$V^* = U(P) - \eta l\big(U(Y^R) - U(P)\big) \tag{27}$$

[8] Note that $1 - \int_{\underline{s}}^{s'} f(s|\bar{e})ds - \int_{\hat{s}}^{\bar{s}} f(s|\bar{e})ds = \int_{s'}^{\hat{s}} f(s|\bar{e})ds$

With:

$$\eta = \begin{cases} 0 & \text{if } \ P \geq Y^R \\ 1 & \text{if } \ P < Y^R \end{cases} \quad (28)$$

Equation (27) can be seen in the light of an agent confronted with two contracts, one with a fixed payment of P and one with a base wage Y^R and bonuses and/or penalties. The agent compares the fixed payment to the base wage. If the fixed payment is larger than the base wage ($P \geq Y^R$), the agent perceives a *gain by not participating.* If the fixed payment is smaller than the base wage ($P < Y^R$), the agent feels a *loss by not participating*: the reservation utility is smaller compared to the reservation independent case. Equivalently, it is easier to satisfy the participation constraint. The principal, being able to influence the base wage (the reference point), now faces the additional choice of setting the base wage above or below the outside option P. The advantage of setting the base wage above the outside option is that there is an immediate, discrete decline in the reservation utility. The disadvantage is clearly that, with a higher base wage and penalties in the contract, for the agent to perceive the same utility, the principal has to give higher net payments. With the following minimisation problem we consider whether setting the base wage higher (and therefore facilitating participation) can permit penalty contracts to be optimal.

The minimisation problem of the principal with a reference-dependent outside option, substituting V^* into participation constraint (6), only changes FOCs (9) and (12) to FOCs (29) and (30), respectively:

$$\int_{\underline{s}}^{\bar{s}} \left(U\big(Y^R + b(s)\big) - \theta l \Big(U(Y^R) - U\big(Y^R + b(s)\big) \Big) \right) f(s|\bar{e}) ds - c(\bar{e}) \geq U(P) - \eta l \Big(U(Y^R) - U(P) \Big) \quad (29)$$

$$\begin{aligned} 1 - \gamma \int_{\underline{s}}^{\bar{s}} (1 + \theta l) U'\big(Y^R + b(s)\big) f(s|\bar{e}) ds - \lambda \int_{\underline{s}}^{\bar{s}} (1 + \theta l) U'\big(Y^R + b(s)\big) \Delta f(s|e) ds \\ + \gamma U'(Y^R) \int_{\underline{s}}^{\bar{s}} \theta l f(s|\bar{e}) ds + \lambda U'(Y^R) \int_{\underline{s}}^{\bar{s}} \theta l \Delta f(s|e) ds - \gamma \eta l U'(Y^R) \geq 0 \end{aligned} \quad (30)$$

Allowing the outside option to be dependent on the reference income set by the principal, inequality (30) might be met with equality by increasing the reference income Y^R above the outside option P. The equivalent condition for reservation independence, inequality (25), is larger than zero everywhere, being the reason to keep the reference income as low as possible. The additional negative term that enters with reservation dependence makes it possible to derive an interior solution for the optimal reference income[9]. By increasing the base wage, the principal then faces a trade-off between decreasing the impact of the outside option on the costs of the contract and increasing these costs by (possibly) shifting payments to the loss region.

Proposition 3. *In the principal-agent model with reservation dependence, it cannot be excluded that the principal should frame at least one payment as a loss.*

Proof.

The first-order condition for the minimisation problem is similar to the case with reservation independence. An additional term, $-\gamma \eta l U'(Y^R)$, adds to the LHS of first-order condition (21). The FOC with respect to $b(s)$, inequality (11), does not change when including reservation dependence. Inequality (25) changes to the following, which is not necessarily non-negative as before because of the additional term that enters:

$$1 - \int_{\underline{s}}^{s'} f(s|\bar{e}) ds - \int_{\hat{s}}^{\bar{s}} f(s|\bar{e}) ds$$

[9] We then assume that the second-order condition is met.

$$-\gamma\Big[-lU'(Y^R)\int_{\underline{s}}^{s'} f(s|\bar{e})ds + U'(Y^R)\int_{s'}^{\hat{s}} f(s|\bar{e})ds + \eta l U'(Y^R)\Big]$$

$$-\lambda\Big[-lU'(Y^R)\int_{\underline{s}}^{s'} \Delta f(s|e)ds + U'(Y^R)\int_{s'}^{\hat{s}} \Delta f(s|e)ds\Big] \tag{31}$$

This result shows that there might be an interior solution to the minimisation problem, namely when equation (31) is equal to zero. This possibly means framing one payment as a loss, but not necessarily. □

A more intuitive explanation of this result follows. Suppose the outside option of the agent is relatively low such that increasing the reference income above it does not involve shifting state-dependent payments to the loss region. One might then find a reference income for which expression (31) is equal to zero, an interior solution to the minimisation problem which involves framing the outside option as a loss, but still framing all the state-dependent payments in the contract as gains. Suppose now that increasing the reference income above the relatively high outside option involves shifting some state-dependent payments to the loss region. The principal then faces the trade-off between the benefits of the outside option being seen as a loss by the agent and the (additional) costs of framing the payments within the contract as a loss. Whether then increasing the reference income above the outside option is beneficial for the principal then depends on which of these effects described above outweighs the other.

3.3 Discrete Example

In this section we show the results we have obtained in the propositions on the basis of a discrete example. We compare (partial) loss framing to gain framing with two outcomes and two effort levels, for simplicity. To be able to display the three possible cases of framing in one expression, we utilize the indicator $\theta_i = 0, 1$; with $i \in 1, 2$; which can be translated as the number of payments in the loss region. If gain framing applies, $\theta_i = 0$ for all i. If the low outcome payment is in the loss region, $\theta_1 = 1$ and $\theta_2 = 0$. Similarly, if both payments are in the loss region, $\theta_i = 1$ for all i. We define $p(s|e)$ to be the discrete, two outcome probability equivalent to the continuous outcome probability $f(s|e)$. We start with *reservation independence* and adapt inequalities (6) and (7) the the situation described here:

$$[PC] \qquad (1+\theta_2 l)p(\bar{s}|\bar{e})U\big(Y^R + b(\bar{s})\big) + (1+\theta_1 l)p(\underline{s}|\bar{e})U\big(Y^R + b(\underline{s})\big)$$

$$-\big(p(\underline{s}|\bar{e})\theta_1 + p(\bar{s}|\bar{e})\theta_2\big)lU(Y^R) - c(\bar{e}) \geq V^* \tag{32}$$

$$[IC] \qquad (1+\theta_2 l)p(\bar{s}|\bar{e})U\big(Y^R + b(\bar{s})\big) + (1+\theta_1 l)p(\underline{s}|\bar{e})U\big(Y^R + b(\underline{s})\big)$$

$$-\big(p(\underline{s}|\bar{e})\theta_1 + p(\bar{s}|\bar{e})\theta_2\big)lU(Y^R) - c(\bar{e}) \geq$$

$$(1+\theta_2 l)p(\bar{s}|\underline{e})U\big(Y^R + b(\bar{s})\big) + (1+\theta_1 l)p(\underline{s}|\underline{e})U\big(Y^R + b(\underline{s})\big)$$

$$-\big(p(\underline{s}|\underline{e})\theta_1 + p(\bar{s}|\underline{e})\theta_2\big)lU(Y^R) - c(\underline{e}) \tag{33}$$

We solve for the low and high outcomes from the participation and incentive compatibility constraint. In this specific case of two outcomes the cost-minimising solution lies at the intersection of the two curves, (32) and (33). Therefore, we subsequently calculate the intersection points for all cases of framing, once for the high outcome payment, once for the low outcome payment:

$$U\big(Y^R + b(\bar{s})\big) = \frac{V^* + c(\bar{e})}{1+\theta_2 l} + \frac{p(\underline{s}|\bar{e})\Delta c(e)}{(1+\theta_2 l)\Delta p(\bar{s}|e)} + \frac{\theta_2 l}{1+\theta_2 l}U(Y^R) \tag{34}$$

$$U\big(Y^R + b(\underline{s})\big) = \frac{V^* + c(\bar{e})}{1+\theta_1 l} - \frac{p(\bar{s}|\bar{e})\Delta c(e)}{(1+\theta_1 l)\Delta p(\bar{s}|e)} + \frac{\theta_1 l}{1+\theta_1 l}U(Y^R) \tag{35}$$

These are the equilibrium solutions for the high and the low outcome payments. At these intersections, the high outcome payment is independent of whether the low outcome payment is in the loss region or not. Likewise, the low outcome payment is independent of whether the high outcome payment is in the loss region or not. We will first compare the gain framing and partial loss framing: equation (34) shows that the high outcome payment is the same in the case when gain framing is applied and when only the low outcome payment is framed as loss as $\theta_2 = 0$. We can therefore see this payment as fixed and rearrange equation (35) to look at the changes partial loss framing brings to the low outcome payment.

$$U\big(Y^R + b(\underline{s})\big) + \theta_1 l\Big(U\big(Y^R + b(\underline{s})\big) - U(Y^R)\Big) = V^* + c(\bar{e}) - \frac{p(\bar{s}|\bar{e})\Delta c(e)}{\Delta p(\bar{s}|e)} \tag{36}$$

At the intersection of the incentive compatibility constraint and the participation constraint, the equilibrium, the utility the principal makes the agent perceive (the RHS in equation (36)) is always the same. Specifically, the RHS of equation (36) is equal to the utility of the low outcome payment when gain framing is applied (which we label $U\big(Y_G^R + b(\underline{s})_G\big)$). On the LHS we have the relationship between the utility of the low outcome payment, $U\big(Y_{PL}^R + b(\underline{s})_{PL}\big)$, and the reference point utility when (partial) loss framing is applied. We label variables referring to gain framing with the subscript G, to partial loss framing with subscript PL and to loss framing with subscript L:

$$U\big(Y_{PL}^R + b(\underline{s})_{PL}\big) + l\Big(U\big(Y_{PL}^R + b(\underline{s})_{PL}\big) - U(Y_{PL}^R)\Big) = U\big(Y_G^R + b(\underline{s})_G\big) \tag{37}$$

Further rearranging gives an indication of how the optimal low outcome payments with and without loss framing relate:

$$l\Big(U\big(Y_{PL}^R + b(\underline{s})_{PL}\big) - U(Y_{PL}^R)\Big) = U\big(Y_G^R + b(\underline{s})_G\big) - U\big(Y_{PL}^R + b(\underline{s})_{PL}\big) \tag{38}$$

As $U(Y_{PL}^R) > U\big(Y_{PL}^R + b(\underline{s})_{PL}\big)$ for any loss frame, the LHS is negative. For the equality to be met, the RHS needs to be negative as well, which means that $U\big(Y_G^R + b(\underline{s})_G\big)$ needs to be smaller than $U\big(Y_{PL}^R + b(\underline{s})_{PL}\big)$. By the increasing nature of the utility function this can only be met when $Y_G^R + b(\underline{s})_G < Y_{PL}^R + b(\underline{s})_{PL}$. This clearly shows that framing the lower payment as a loss is not beneficial for the principal: the optimal high outcome payments do not differ between the gain frame and the partial loss frame. The optimal low outcome payment for loss framing is larger than the one for gain framing as shown above. This is also confirmed by the low outcome payment, under the premise that the principal stays in the partial loss frame:

$$U\big(Y_{PL}^R + b(\bar{s})_{PL}\big) = \frac{c(\bar{e}) + V^*}{1+l} - \frac{p(\bar{s}|\bar{e})\Delta c(e)}{(1+l)\Delta p(\bar{s}|e)} + \frac{l}{1+l}U(Y_{PL}^R) \tag{39}$$

The higher the reference point in the partial loss frame, the higher the net payment for the low outcome. Decreasing the reference point is the best the principal can do. Therefore, with a higher reference point for partial loss framing and an overall higher net payment for the state-dependent payments, loss framing is not beneficial for the principal.

Secondly, we compare the equilibrium values for partial loss framing and loss framing in the same manner. In both cases, the low outcome payment in equation (35) is the same as it does not depend on whether the high outcome payment is in the loss region or not. We will rearrange equation (34), denoting the payments in the case of loss framing with the subscript L:

$$U\big(Y_L^R + b(\bar{s})_L\big) + l\bigg(U\big(Y_L^R + b(\bar{s})_L\big) - U(Y_L^R)\bigg) = V^* + c(\bar{e}) + \frac{p(\underline{s}|\bar{e})\Delta c(e)}{\Delta p(\bar{s}|e)} \tag{40}$$

The RHS of equation (40) is equal to the high outcome payment for partial loss framing:

$$U\big(Y_L^R + b(\bar{s})_L\big) + l\Big(U\big(Y_L^R + b(\bar{s})_L\big) - U(Y_L^R)\Big) = U\big(Y_{PL}^R + b(\bar{s})_{PL}\big) \tag{41}$$

Again rearranging shows the relation between the partial loss framing high outcome payment and the one for loss framing:

$$l\Big(U\big(Y_L^R + b(\bar{s})_L\big) - U(Y_L^R)\Big) = U\big(Y_{PL}^R + b(\bar{s})_{PL}\big) - U\big(Y_L^R + b(\bar{s})_L\big) \tag{42}$$

Similarly to the comparison between partial loss framing and gain framing, the LHS is negative because $U\big(Y_L^R + b(\bar{s})_L\big) < U(Y_L^R)$. This in turn means that the RHS needs to be negative as well to meet the equality, which can only be true if $U\big(Y_{PL}^R + b(\bar{s})_{PL}\big) < U\big(Y_L^R + b(\bar{s})_L\big)$. Due to the fact that the utility function is increasing, this means that $Y_{PL}^R + b(\bar{s})_{PL} < Y_L^R + b(\bar{s})_L$. Likewise, within the loss frame, the principal should decrease the reference point as much as possible which is reflected in the net payment for the high outcome:

$$U\big(Y_L^R + b(\bar{s})_L\big) = \frac{c(\bar{e}) + V^*}{1+l} - \frac{p(\underline{s}|\bar{e})\Delta c(e)}{(1+l)\Delta p(\bar{s}|e)} + \frac{l}{1+l}U(Y_L^R) \tag{43}$$

From these results we can conclude that loss framing is more costly to the principal than partial loss framing, and the best option the principal has, is to express all payments as gains. The intuition for this result is straightforward: for any reference point the principal sets, the utilities perceived by the agent, for the low and the high outcome payments, are the same for economically equivalent contracts. Increasing the reference point, means that, as soon as the agent starts to perceive losses, the principal will have to pay the agent more to make him perceive the same low and high outcome utilities. Because of this, the principal wants to keep the reference point as low as possible. Subsequently, gain framing is optimal, and the reference point should be set at:

$$U(Y^R) \leq c(\bar{e}) + V^* - \frac{p(\bar{s}|\bar{e})\Delta c(e)}{\Delta p(\bar{s}|e)} \tag{44}$$

If you now consider *reservation dependence*, there are several changes to the reservation independence case. We will illustrate that with reservation dependence, framing at least one payment as loss, and increasing the reference point such that not participating feels like a loss, is optimal for the principal in the two outcome case. First of all, the dependence of the reservation utility on the reference income requires to substitute $U(P) - \eta l\big(U(Y^R) - U(P)\big)$ for V^* into equations (34) and (35) to get the equilibrium solutions for the general case:

$$U\big(Y^R + b(\bar{s})\big) = \frac{(1+\eta l)U(P) + c(\bar{e})}{1+\theta_2 l} + \frac{p(\underline{s}|\bar{e})\Delta c(e)}{(1+\theta_2 l)\Delta p(\bar{s}|e)} + \frac{\theta_2 l - \eta l}{1+\theta_2 l}U(Y^R) \tag{45}$$

$$U\big(Y^R + b(\underline{s})\big) = \frac{(1+\eta l)U(P) + c(\bar{e})}{1+\theta_1 l} - \frac{p(\bar{s}|\bar{e})\Delta c(e)}{(1+\theta_1 l)\Delta p(\bar{s}|e)} + \frac{\theta_1 l - \eta l}{1+\theta_1 l}U(Y^R) \tag{46}$$

Depending on the level of $U(P)$ (which is fixed), the point where η switches from being zero to one, may be in the gain frame region, in the partial loss frame region or in the loss frame region. This is because the optimal payments for each of these regions are functions of the outside option, P, and the reference point is bound to a specific region by each frame. Therefore, we have two cases according to the switching point of η (we provide the proof for the two cases in Appendix 6.2, where we also exclude the loss region under the first case):

Case 1 The switching point of η lies in the partial loss frame region and losses with respect to the outside option are not felt in the gains region, iff

$$c(\bar{e})p(\bar{s}|\underline{e}) \geq c(\underline{e})p(\bar{s}|\bar{e}) \tag{47}$$

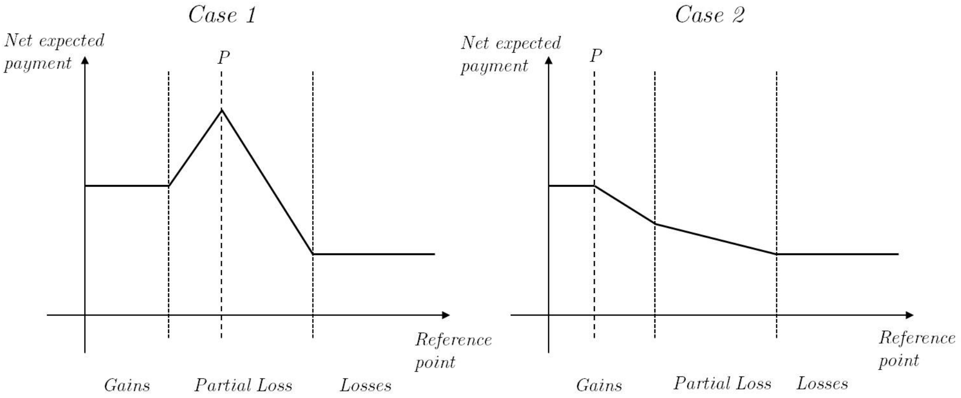


Figure 5: *Case 1 depicts the effect of increasing the reference point on the net expected payments iff condition (47) is true. Respectively, Case 2 schematically exhibits the net payments iff condition (48) applies.*

Case 2 The switching point of η lies in the gains region and losses with respect to the outside option are felt in all three frames, iff

$$c(\bar{e})p(\bar{s}|\underline{e}) < c(\underline{e})p(\bar{s}|\bar{e}) \tag{48}$$

It follows that we only need to consider whether it is beneficial to increase the reference point in these two cases. For this purpose, we need to inspect the net payments (see equations (45) and (46)) for low and high outcomes when changing from the gain frame to the partial loss frame, and from the partial loss frame to the loss frame. The net expected payments as a function of the reference point are depicted in Figure 5 for both cases. From the graph it is clear that there is no discontinuity between the frames, which can be proved by letting $b(\bar{s})$ and $b(\underline{s})$ decrease to zero. If the principal frames as gains, and Case 1 applies, the reference point has no impact on the payments and therefore the level of it is irrelevant. On the contrary, if Case 2 applies and therefore the agent already feels a loss by not participating in the gains region, increasing the reference point decreases both net payments. For both cases it is then optimal, under the premise that we stay in the gains region, that the reference point is increased as much as possible, namely to $U(Y_G^R) = U\big(Y_G^R + b(\underline{s})_G\big)$. It is in the partial loss region, where the two cases differ: in Case 1, increasing the reference point has no effect on the net high payment when not participating is still seen as a gain ($\eta = 0$), but the principal has to increase the low net payment as well to compensate. If the reference point is increased further, such that it is increased above the outside option, the low net payment does not change, but the high net payment decreases with the reference point. The net expected payment for partial loss framing with the highest possible reference point ($\eta = 1, U(Y_{PL}^R) = U(Y_{PL}^R + b(\bar{s})_{PL})$) is lower than the one for gain framing with $\eta = 0$. In Case 2, the agent immediately feels losses by not participating and therefore the net payments directly decrease when increasing the reference point. Increasing the reference point further, such that only losses are seen within the contract (and η is always equal to one), the net payments are again independent of the reference point. From this cost-benefit analysis with reservation dependence we can see that framing as (partial) loss is optimal in contrast to the reservation independence case: framing as partial loss with $b(\bar{s})_{PL} = 0$, or analogously $U(Y_{PL}^R) = U\big(Y_{PL}^R + b(\bar{s})_{PL}\big)$, is equivalent to framing as losses with any $U(Y_L^R) > U\big(Y_L^R + b(\bar{s})_L\big)$ with respect to the principal's costs.

3.4 Numerical Example

Consider the relationship of the university (principal) and a researcher (agent) again. The university plans to switch from a fixed payment to a variable payment scheme. Suppose that the only task of the researcher is to publish one paper per year in his domain and that the payment the researcher receives per year is a base salary and bonuses or penalties that are determined by the number of papers he published: zero or one (assuming for the ease of exemplification that publishing more than one paper a year is impossible). In this first scenario, the researcher does not compare her outside option to the reference point. Like in the theoretical model above, we assume that if the university chooses to offer a contract to the researcher and the latter participates, he will give high effort. This means that the researcher works hard by writing one paper per year, but he might get unlucky and might not publish it. In 30% of the years he works at the university this is the case. In 70% of the cases he can publish one paper. If the researcher only gives low effort, the chances of not publishing are 80%. He might get lucky, in 20% of the cases, he will be able to publish one paper. The university considers to offer one of three economically equivalent contracts, which all have an expected cost of 26,410 Euros per year:

Contract 1. We will pay you a base wage of 20,000 Euros per year. If you don't publish, we do not pay you any bonus. If you publish one paper, we will pay you a bonus of 9,157 Euros.

Contract 2. We will pay you a base wage of 23,000 Euros per year. If you don't publish, we will deduct 3,000 Euros from your base wage. If you publish one paper, we will pay you a bonus of 6,157 Euros.

Contract 3. We will pay you a base wage of 30,000 Euros per year. If you don't publish, we will deduct 10,000 Euros from your base wage. If you publish one paper, we will deduct 843 Euros from your base wage.

The first contract is a pure bonus contract, the second is a mixture of bonus and penalty, the third is a pure penalty contract. In standard theory these three contracts should result in the same effort provision. But as preferences are reference-dependent (and framing is effective), this is not the case any longer. For this example, we use the utility function of the form $U(x) = 150\sqrt{x}$; a loss aversion value of $l = 2$; a reservation utility of $V^* = 18,293$ Euro; cost of high effort $c(\bar{e}) = 6,000$ Euro and the cost of low effort $c(\underline{e}) = 3,800$ Euro. With the choice of probabilities the MLRP is satisfied: $f(\bar{s}|\bar{e}) = 0.7$; $f(\bar{s}|\underline{e}) = 0.2$.

The researcher declines the second and third contract as a result of loss aversion although the net payment to him would be the same in all three contracts. Keeping the base wage and the high outcome payment in the second contract, the penalty would have to be decreased in absolute terms to 1,024 Euros for the researcher to participate. This is a more expensive alternative for the university with average costs of 27,003 Euros. Although using penalties in contracts has incentive effects, this does not play a role: the incentive compatibility constraint is easier satisfied with penalties included in the contract, but the participation constraint is less easily satisfied. For the loss contract to be accepted by the researcher, both payments would have to be increased, more specifically the low outcome penalty would have to be equal to -4,172 Euros with the high outcome penalty being -1 Euro[10].

Now, we examine the second scenario, where the researcher considers moving to another university. The researcher compares his actual base wage to the base wage that another university is offering him[11]. Again, the university considers offering one of the three contracts as described above. For our choice of probabilities and cost levels, the outside option is smaller than the reference income, $P < Y^R$ (see switching point of η in condition (48)). With reservation dependence (and $\eta = 1$), the participation constraint is satisfied more easily and is not binding in all contracts, and it is the incentive compatibility constraint that becomes binding in all contracts.

We consider each of the contracts in turn: the bonus contract (Contract 1) just meets the incentive compatibility constraint, therefore none of the variable payments can be reduced and

[10] If the high payment would be kept at -843 Euros, there is no possible contract to be offered. If both the payments are changed, there are several other possibilities to formulate the contract with losses. We have chosen the high outcome payment such that we are still just in the loss region and then set the low outcome payment such that the agent still participates. The incentive compatibility constraint is non-binding in this case, the participation constraint stays binding. This is not the optimal solution as we are not at the intersection of the participation and incentive compatibility constraint (which is impossible to achieve with these variables).

[11] The base wage of the current contract is equivalent to the reference income, Y^R, the base wage of the alternative contract is equivalent to the outside option, P. We consider an outside option that satisfies $V^* = U(P)$, such that the cases are comparable to reservation independence. This is true for a P of 14,873 Euros.

the expected costs remain to be 26,410 Euros for the university. In the bonus and penalty contract (Contract 2), both constraints are met with inequality, therefore the high outcome payment can be reduced to 0 Euro to still be in the partial loss region. Then, the low outcome payment can additionally be decreased to -13,574 Euros, making the participation constraint binding and the incentive constraint non-binding. The expected costs for the university reduce to 18,928 Euros and is subsequently cheaper than the bonus contract. The participation contract of the penalty contract (Contract 3) is also only met with inequality, the high outcome payment can be reduced accordingly to -10,000 Euros (note that this is one possibility only) and the low outcome payment to -15,367 Euros. Calculating the expected costs for the university (18,390 Euros) reveals that the university should employ the penalty contract.

4 Robustness

There are several specifications of the theoretical model that could be considered as extensions. We discuss the four most important ones in this section and evaluate the possible changes it would bring to the existing analysis, though we do not consider fundamental changes to the model here.

We define loss aversion as linear, which keeps the loss utility concave. Prospect theory though considers the utility in the loss region as convex, the agent turns risk seeking in this region. With this change to the model, the marginal utility is highest just below the reference point and the principal does not need to pay a risk premium to the agent. The agent takes the risk voluntarily. Hence, it might be beneficial for the principal to increase the reference point to a level just above the low(er) outcome payment(s). Accordingly then, partial loss frames, the combination of gains and losses, might be optimal with reservation independence. On the other hand though, setting the reference point such that one payment is in the loss region makes the decision maker *more* risk averse. This would mean that the principal would have to pay an even larger risk premium and it would not be worth paying penalties. With reservation dependence, the agent might take more risk with respect to the outside option as well, therefore the chance of not participating becomes larger when employing penalties in the contract. For that reason, it is ambiguous how the principal should frame the contract.

In this analysis we have also assumed that only losses enter the utility of the agent. One could imagine that if the agent feels a gain, that his utility increases in a similar way as losses decreases utility. This would make the propositions on reservation independence discussed even stronger, as the agent feels additional gains if the principal sets bonuses which allows him to decrease the reference point even further. If we assume that the lowest possible reference point is zero, the principal should set it exactly at that point. With reservation dependence, the additional gains in the utility of the agent put downward pressure on the optimal reference point. The payments though, should still lie in the partial loss region and with not participating feeling as a loss.

Another restriction of our model is the discreteness of the effort the agent provides. Extending it to continuous effort does not lead to considerably different results: the incentive compatibility constraint is replaced by the first-order condition that maximises the incentive constraint with respect to e (the formal requirements and necessary assumptions of the first-order approach are defined by Rogerson, 1985[12]). If you suppose that e_1 is the best solution to the maximisation and e_2 the next-best effort level the agent can choose from a set of many discrete effort levels, the principal faces the same minimisation problem as before: redefining $\Delta f(s|e) = f(s|e_1) - f(s|e_2)$ and $\Delta c(e) = c(e_1) - c(e_2)$ results in the same first-order conditions as before. Letting the difference in effort levels decrease to zero in the limit, therefore having a continuum of effort levels, does not change the proposition that, for any required effort level[13], the principal should frame as a gain in the case of reference independence.

[12] The requirement concerns the linearity of the distribution function, e.g. $f(s|e) = ef(s|\bar{e}) + (1-e)f(s|\underline{e})$, with $e \in [0,1]$. It means that the higher the effort level the agent exerts, the higher the probability that a high outcome is reached. The maximisation problem is then convex.

[13] Except for the lowest possible effort level, when the principal pays a fixed wage, as incentivising is not necessary. We do not consider that the principal might find inducing the lowest possible effort level, $\underline{e}$, most profitable as this lies outside the scope of the paper.

Another anomaly embedded in prospect theory is the weighting function: a perception of probabilities that is different from the objective one could have an impact on the design of optimal incentive schemes, although it appears that building this into the model will only have a minor effect and is more suited for fine-tuning the contract rather than changing the general picture of the optimal payments.

5 Discussion and concluding remarks

"An employer would have to offer a higher set of payoffs to get employees to accept an incentive contract if it is described as a penalty than if it is described as a bonus; and if employers do not gain some greater benefit by using penalty language, there is no reason for them to offer the penalty contracts and incur the extra expense." (Luft, 1994, p.199). Our paper was set out to develop a theory of contracts with reference-dependent preferences including loss aversion and allowing for the reference income to be set directly by the principal. With the base wage as reference point, bonus contracts are optimal if the agent does not evaluate reservation opportunities against the reference point from within the contract. Penalties are optimal only under specific circumstances, namely when the agent compares his outside option to the reference point strategically set by the principal. The theoretical results explain the experimental findings that, amongst others, Luft (1994) describes.

The nature of the reference income is matter of much debate in both the theoretical and experimental literature. Assuming that the principal can influence the reference income by setting a base payment is the key point of this paper. Although we name it the "base wage", the results apply for any *directly* set reference point. Armantier and Boly (2012) assume that the reference point is increasing in the base salary, which should be analysed according to our theory as well. Additionally, this assumption of a reference point that is susceptible to direct framing, allows the principal to employ pure bonus and pure penalty contracts, next to the combination of these incentive payments (although we show the latter two not to be always optimal). As framing payments as gains for the agent is optimal with reservation independence, the question arises in this case whether the agent would still see the low base wage as reference income. Consider an employee who receives a low base wage and for each outcome he receives a bonus on top, increasing in size with increasing performance levels. He will then *expect* to receive more than the base wage. This would mean that the employee could form the reference point from his expectations of the payment in the contract rather than taking the base wage as reference point. The main alternative definition of the reference point are expectations à la Kőszegi and Rabin (2006). With this endogenous formulation, the reference point could only be modified indirectly, by anticipating on the on reference point that a net payment schedule offered by the principal would induce. Importantly, this stochastic nature of the reference point does not permit for pure penalty and pure bonus contracts. Individual decision-makers are subject to some psychology that makes them form the reference point automatically. Although researching this mechanism is important, it falls outside the scope of this paper.

Specific to employment contracts is that the diversity of formulation opportunities is large: it might be that the optimal payment scheme changes when the agent is working in a team and is evaluated and paid relative to his co-workers. Here, the various options regarding the underlying performance variable also alter the payment scheme. The reference income is a relative variable then. Framing of payments in peer comparison situations has other, peculiar, advantages and disadvantages. It can have positive effects on motivation of groups or individuals, but can also discourage agents that end up on the lower end of the performance comparison. Up until now, we have also assumed that the agent is employed over one period only - changing this to a multi-period principal-agent setting will probably also reveal that different payment schemes are required to align incentives, especially when shifting reference points would be taken into consideration. Although bonus payments are seen as the cause of the financial crisis by many, it is important to note that specifying the time span over which the contract is executed is crucial to the effectiveness of the incentives. Only giving short-term incentives, like in the banking sector in the period leading up to the financial crisis, is not optimal. A combination of short- and long-term pre-specified target outcomes (that are translated into the payments) and effective control mechanisms can hopefully prevent bonus payments from being the trigger for future (financial) crises. Furthermore, the outcomes of this paper are not only applicable to the financial and banking sector.

Obviously, the optimal contract set out in this paper is not suited for all principal-agent relationships, although it encompasses a vast range of possibilities. Practical problems of, for example, the observability of output created by the agent makes the correct determination of payments difficult. The incentive effect of the choice of the contract can be muted. Another question that theoretical research cannot answer is whether implementing performance-based payment schemes influences the quality of the output or not.

Connected to the supposition of a reference income for the agent is the question of the possibility of strategic framing. As already set out in the literature, the reference point of the agent can be influenced, but the main question is how and to what extent. As reasoned before, the base wage definitely has at least some influence on the formation of the reference point of the agent. Consistent experimental testing of potential (combinations of) reference points is missing, but they are necessary to build meaningful models for optimal compensation schemes. Incorporating experimental outcomes then into this existing model will give more powerful tools for designing efficient contracts. Our further research focuses on testing the five possible payment schemes in practice. Experiments are expected to reveal the difference in incentive effects of the alternative payment schemes. It will also show whether assuming that the base wage induces a reference income can explain the behaviour of agents (and therefore the share in the reference income). The question of further psychological factors influencing the formation of the reference point however remains.

6 Appendix

6.1 Multipliers

First-order stochastic dominance (FOSD) refers to the assumption that the distribution of the state realization s conditional on $\bar{e}$ stochastically dominates its distribution conditional on $\underline{e}$, $F(s|\bar{e}) \leq F(s|\underline{e})$, for all s, with strict inequality on some open set of s. This in turn means that the probability of reaching at least some s is higher under high effort, $\bar{e}$, than under low effort, $\underline{e}$ (see Mas-Colell et al., 1995, p.194ff). Note that FOSD is implied by MLRP and is therefore valid throughout the model.

Proposition 4. *The multipliers are not zero for all s that satisfy $s \leq s'$ or $s \geq \hat{s}$.*

Proof.

Suppose that $\gamma = 0$. By FOSD, there must be a range of low states s where $f(s|\underline{e}) > f(s|\bar{e})$ and it follows then that $\Delta f(s|e) < 0$. The right-hand side of expression (13), which is met with equality for all s satisfying $s \leq s'$ or $s \geq \hat{s}$, would be negative which in turn implies that $U'(Y(s)) < 0$ for some low states s, which is excluded by the definition of the form of the utility function. Therefore, $\gamma > 0$.

Suppose now that $\lambda = 0$. At first sight it seems that it is still possible to have two payments, for high and low performance, without violating the first-order condition: one payment, $Y(\underline{s})$ below Y^R for low effort and one payment $Y(\bar{s})$ above Y^R for high effort. By expression (13) these would be the following:

$$\frac{1}{U'\big(Y(\underline{s})\big)} = \gamma(1+l) \tag{49}$$

as there is loss aversion and

$$\frac{1}{U'\big(Y(\bar{s})\big)} = \gamma \tag{50}$$

without loss aversion. This in turn would mean that $U'\big(Y(\bar{s})\big) > U'\big(Y(\underline{s})\big)$, which is impossible given the assumption that $U'' < 0$. Therefore, $\lambda > 0$. □

6.2 Proof for cost-benefit analysis in Subsection 3.3

Case 1. *The switching point of η lies in the partial loss frame region.*

This is the case iff:

$$c(\bar{e})p(\bar{s}|\underline{e}) \geq c(\underline{e})p(\bar{s}|\bar{e}) \tag{51}$$

We set the reference point at the highest possible value such that the payments are still in the gains region, and no losses are perceived with respect to the outside option, P:

$$U(Y_G^R) = U\big(Y_G^R + b(\underline{s})_G\big) = c(\bar{e}) + U(P) - \frac{p(\bar{s}|\bar{e})\Delta c(e)}{\Delta p(\bar{s}|e)} \tag{52}$$

If and only if condition (51) is met, $U(Y_G^R) \leq U(P)$ and subsequently η is equal to zero. Increasing $U(Y_G^R)$ above $U\big(Y_G^R + b(\underline{s})_G\big)$ just a little means that we are entering the partial loss frame and that η can still be equal to zero and no losses are seen with respect to the outside option. Setting the reference point at the highest possible payment within the partial loss frame, shows that η needs to switch to the value of one in this region:

$$U(Y_{PL}^R) = U\big(Y_{PL}^R + b(\bar{s})_{PL}\big) = c(\bar{e}) + U(P) - \frac{p(\underline{s}|\bar{e})\Delta c(e)}{\Delta p(\bar{s}|e)} \tag{53}$$

For $U(Y_{PL}^R) \leq U(P)$, it would have to be true that $c(\bar{e}) + \frac{p(\underline{s}|\bar{e})\Delta c(e)}{\Delta p(\bar{s}|e)} < 0$, which is impossible. It follows, that increasing the reference point further, such that we are in the loss frame with the optimal payments, *always* includes framing the outside option as loss as well.

Case 2. *The switching point of η lies in the gain region.*

This is the case iff:

$$c(\bar{e})p(\bar{s}|\underline{e}) < c(\underline{e})p(\bar{s}|\bar{e}) \tag{54}$$

We set the reference point at the highest possible value such that the payments are still in the gains region, and losses are perceived with respect to the outside option, P:

$$U(Y_G^R) = U\big(Y_G^R + b(\underline{s})_G\big) = c(\bar{e}) + (1+l)U(P) - lU(Y_G^R) - \frac{p(\bar{s}|\bar{e})\Delta c(e)}{\Delta p(\bar{s}|e)} \tag{55}$$

If and only if condition (54) is met, $U(Y_G^R) > U(P)$ and therefore η is equal to one. Increasing the reference point above $U(Y_G^R)$, such that we are either in the partial loss or the loss frame, means that the outside option is seen as a loss in any of the cases.

References

Armantier, O., & Boly, A. (2012, April). Framing of Incentives and Effort Provision. *Mimeo.*

Bebchuk, L., & Grinstein, Y. (2005). The Growth of Executive Pay. *Oxford Review of Economic Policy*, *21*(2), 283-303.

Bonner, S. E., & Sprinkle, G. B. (2002). The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Accounting, Organizations and Society*, *27*(4-5), 303-345.

Brink, A. G. (2008). *The Effects of Risk Preference and Loss Aversion on Individual Behavior under Bonus, Penalty, and Combined Contract Frames.* Dissertation, Doctor in Philosophy, Florida State University; College of Business.

Brookshire, D. S., & Coursey, D. L. (1987, September). Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures. *The American Economic Review*, *77*(4), 554-566.

Coursey, D. L., Hovis, J. L., & Schulze, W. D. (1987, August). The Disparity Between Willingness to Accept and Willingness to Pay Measures of Value. *The Quarterly Journal of Economics*, *102*(3), 679-690.

Cuñat, V., & Guadalupe, M. (2009, March). Executive compensation and competition in the banking and financial sectors. *Journal of Banking and Finance*, *33*(3), 495-504.

Daido, K., & Itoh, H. (2006, July). The Pygmalion Effect: An Agency Model with Reference-Dependent Preferences. *CESifo Area Conference on Applied Microeconomics*, 1-27.

Dalen, H. P. v., & Henkens, K. (2012). Intended and Unintended Consequences of a Publish-or-Perish Culture: A Worldwide Survey. *Jounal of the American Society for Information Science and Technology*, *63*(7), 1282-1293.

De Meza, D., & Webb, D. (2007, March). Incentive Design under Loss Aversion. *Journal of the European Economic Association*, *5*(1), 66-92.

Edmans, A., & Gabaix, X. (2009). Is CEO Pay Really Inefficient? A Survey of New Optimal Contracting Theories. *European Financial Management*, *15*(3), 486-496.

Falk, A., & Ichino, A. (2003, March). Clean Evidence on Peer Pressure. *Institute for the Study of Labor (IZA) Discussion Paper Series No. 732*.

Gächter, S., Johnson, E. J., & Herrmann, A. (2007, July). Individual-Level Loss Aversion in Riskless and Risky Choices. In *Centre for Decision Research and Experimental Economics (CeDEx), Discussion Paper Series.* The University of Nottingham. (ISSN 1749-3293)

Griffin, W. (2011). *Management. Principles and Practices* (10th ed.). South-Western, Cengage Learning.

Heath, C., Larrick, R. P., & Wu, G. (1999, February). Goals as Reference Points. *Cognitive Psychology*, *38*(1), 79-109.

Herweg, F., Müller, D., & Weinschenk, P. (2008, October). The Optimality of Simple Contracts: Moral Hazard and Loss Aversion. *Bonn Econ Discussion Papers, Bonn Graduate School of Economics (BGSE), University Bonn*, *2008*(17).

Hossain, T., & List, J. A. (2009, October). The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations. *NBER Working Paper No. 1562*.

Just, D., & Wu, S. (2005, March/April). Loss Aversion and Reference Points in Contracts. *Paper prepared for presentation at the SCC-76 Meeting*.

Kahneman, D., Knetsch, J., & Thaler, R. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives*, *5*(1), 193-206.

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

Kőszegi, B., & Rabin, M. (2006, November). A Model of Reference-Dependent Preferences. *The Quarterly Journal of Economics*, *121*(4), 1133-1165.

Koop, G. J., & Johnson, J. G. (2012, January). The Use of Multiple Reference Points in Risky Decision Making. *Journal of Behavioral Decision Making*, *25*(1), 49-62.

Kühberger, A. (1998, July). The Influence of Framing on Risky Decisions: A Meta-Analysis. *Organizational Behavior and Human Decision Processes*, *75*(1), 23-55.

Laffont, J.-J., & Martimort, D. (2002). *The Theory of Incentives. The Principal-Agent Model.* Princeton and Oxford: Princeton University Press.

Levin, I., Schneider, S., & Gaeth, G. (1998, November). All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects. *Organizational Behavior and Human Decision Processes*, *76*(2), 149-188.

Levy, H., & Levy, M. (2002). Experimental Test of the Prospect Theory Value Function: A Stochastic Dominance Approach. *Organizational Behavior and Human Decision Processes*, *89*(2), 1058-1081.

Lopes, L. L., & Oden, G. C. (1999). The Role of Aspiration Level in Risky Choice: A Comparison of Cumulative Prospect Theory and SP/A theory. *Journal of Mathematical Psychology*, *43*(2), 286-313.

Luft, J. (1994). Bonus and penalty incentives. Contract choice by employees. *Journal of Accounting and Economics*, *18*(2), 181-206.

Macho-Stadler, I., & Pérez-Castrillo, J. (2001). *An Introduction to the Economics of Information. Incentives and Contracts* (2nd ed.). Oxford, New York: Oxford University Press.

Mas-Colell, A., Whinston, M., & Green, J. (1995). *Microeconomic Theory.* New York, Oxford: Oxford University Press.

Matthey, A. (2008, October). *On the Formation and Manipulation of Reference States.* (from www.econ.mpg.de on 13/04/2012)

Munro, A., & Sugden, R. (2003). On the theory of reference-dependent preferences. *Journal of Economic Behavior and Organization*, *50*(4), 407-428.

Rogerson, W. P. (1985, November). The First-Order Approach to Principal-Agent Problems. *Econometrica*, *53*(6), 1357-1367.

Scheufele, D. A., & Tewksbury, D. (2007, March). Framing, Agenda Setting and Priming: The Evolution of Three Media Effects Models. *Journal of Communication*, *57*(1), 9-20.

Schmidt, U., & Traub, S. (2002). An Experimental Test of Loss Aversion. *The Journal of Risk and Uncertainty*, *25*(3), 233-249.

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, *211*(4481), 453-458.

Wathieu, L. (2004, May). Consumer Habituation. *Management Science*, *50*(5), 587-596.