Strategic Framing in Contracts

Contracts under hidden action

Katharina Hilken
Fonds Wetenschappelijk Onderzoek (FWO), Vlaanderen
Vrije Universiteit Brussel (VUB), Brussels, Belgium

Kris de Jaegher
Utrecht School of Economics (USE), Universiteit Utrecht, Utrecht, The Netherlands

Marc Jegers
Vrije Universiteit Brussel (VUB), Brussels, Belgium

Working Paper Version: March 9, 2012

JEL-codes: D86; D03; J33; M52

Abstract

We provide a model of the principal-agent relationship with hidden action where the agent thinks in terms of gains and losses with respect to a reference point. A loss averse agent's reference point is the fixed payment that he receives, the gains and losses are respectively any bonuses or penalties. When choosing the net payment for each outcome produced by the agent, the principal takes into account that the base wage chosen determines the agent's reference point, and therefore his behaviour. We consider two variants of the model. In a first variant, the agent's reservation utility is not reference-dependent. We show that the principal always employs bonuses in this case. In a second variant, the reservation utility is reference-dependent. In this case, the principal may also use penalties.

1 Introduction

During the financial crisis of 2007-2009 payment schemes of managers became under close scrutiny and criticism both with respect to their level as 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 USA, the fraction of performance-based pay in the total pay of executives increased significantly during the 1990s [1]. 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. Observation of current remuneration schemes shows that the predominant and increasingly important incentive system for executives consists of bonuses, e.g. stocks or options (Bebchuk & Grinstein, 2005), rather than penalties or a combination of the two. The question arises whether equity-based compensation is optimal to achieve the alignment of incentives (Harris, 2009).

[1] Sample periods were 1993-1999 and 1995-2002, the periods after major deregulations in the financial and banking sectors in the USA.

The optimal payment plans are (theoretically) derived as the solution to the standard principal-agent problem[2] as there is asymmetric information between owners and executives as to effort provision. Standard principal-agent theory is based on expected utility theory; the axioms though have been found to be inconsistent with individual decision-making under uncertainty in experiments. This led to the development of prospect theory (Kahneman & Tversky, 1979) which formalizes experimental phenomena. One of them is framing which is interpreted in various manners in the existing literature. Levin, Schneider, and Gaeth (1998) compiled experimental studies of the framing effect from the years 1985-1996 and found a considerable amount of support in various domains, although specific characteristics diminish or eliminate this effect[3].

The Asian Disease Experiment[4] (ADE) is the most famous example of framing and preference reversal: the decision taken by the participant is crucially dependent on the reference point induced by the wording of the choice problem. This also implies that psychic valuation does not depend on absolute levels, but on whether the agent gains or loses with respect to a reference point. Loss aversion, the observation that losses hurt more than the same amount of gain gives pleasure, is another feature of human decision-making that is reflected in the ADE. It can explain a respectable quantity of anomalies, such as [5]: the equity premium puzzle (Barberis N., 2008, Benartzi & Thaler, 1995) and the endowment effect (e.g. Genesove & Mayer, 2001). A loss aversion coefficient of about two has been estimated by Tversky and Kahneman (1991), the magnitude depending on the dimension which is looked at. The academic discussion of whether loss aversion exists rather moved to the causes of it and the angle from which to look at it. However, the ADE can only be explained completely with the so-called reflection effect (Kahneman & Tversky, 1979): People are risk-averse in the domain of gains and risk-seeking in the domain of losses.

Incorporating framing and loss aversion into principal-agent theory in contracts is the focus of this paper. We are taking the broad, but most compelling, definition of framing by Hallahan (1999, p.208):

> Framing operates by biasing the cognitive processing of information by individuals.

Strategically framing can then be defined to be the explicit use of wording and actions to influence the behaviour of an agent in the interest of the principal. As the wording of a choice problem influences the preferences of the agent, we suspect that shaping the payment scheme will have an impact on the effort provision of the agent. The division of the total (expected) salary into base payment (which we will assume to be the reference point), bonuses and/or penalties, is therefore an effective tool of influencing the perception of the agent. We propose that giving the agent a low base salary and bonuses will have a different effect on the effort provision than a contract with a high base salary and penalties as well as one with a base wage and a combination of bonuses and penalties (all with the same net payment to the agent for any given outcome). As noted by Kahneman and Tversky (1979, p.277-278) their value function 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. We will consequently combine a concave utility function and a reference-dependent utility function that incorporates loss aversion. The goal of salary design is to ultimately increase the efficiency and the profitability of the contractual relationship for the principal. Adding prospect-theoretic preferences to the utility derived from the level of wealth in the general principal-agent model and making the reference point variable leads to the optimal incentive scheme for the agent when

[2]A general discussion of the standard principal-agent models can be found in Macho-Stadler and Pérez-Castrillo (2001), Laffont and Martimort (2002) and Mas-Colell, Whinston, and Green (1995).

[3]Participants make more risky choices when human lives are concerned rather than property and with other peoples' money. The framing effect is eliminated in studies with only experts, when participants need to provide a rationale for their decision and when provided with full information. Ambiguous results are shown with respect to the time the subjects get for their decision. Gender and the extremeness of probabilities also seem to have an effect on the strength of the framing effect. Furthermore, it is important to note that the strength of the framing effect depends on the topic and the amount at stake.

[4]Introduced by Kahneman and Tversky (1984) and described in the appendix.

[5]Only the more recent findings and experiments are referred to. A collection of the outcomes of older research can be found in for example Camerer (2000) and/or Kahneman, Knetsch, and Thaler (1991).

the principal is exploiting the possibility of direct strategic framing. 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. Our contribution is to consider the reference point as strategic as well, where the reference point is assumed to be equal to the base wage.

Various difficulties arise by modelling reference-dependent preferences in the principal-agent setting: Through the kink in the utility function of the agent at the reference income, the incentive scheme has a region where the payment is independent of performance. Additional assumptions on the feasible ranges of states for which the agent will still participate are necessary to ensure a performance pay. Furthermore, modelling the reference income is crucial to the suggested outcomes. With this also comes the effectiveness of strategic framing with respect to direct or indirect manipulation of the agent. This theoretical research shows that, assuming a directly influencable reference income of the agent by setting the base wage, bonus contracts are optimal. Depending on the definition of performance outcomes and the whether the outside option is dependent or independent of the reference income, other payment schemes might be applied.

The paper is organized as follows. The model is described in section 2. We discuss two different cases: One where the reservation utility of the agent is independent of the reference income and one where it is dependent. The crucial assumption of this model is that the reference point is equal to the base wage and can therefore be determined by the principal. Section 3 discusses possible extensions of the theoretical model and the hypotheses that need to be tested experimentally in the future. Furthermore, criticism of the effectiveness of performance-based pay coming from the behavioural theories is discussed. Section 4 concludes.

2 The Model

2.1 Optimal incentive scheme with linear loss aversion

The reference-dependent principal-agent model considers optimal incentive payments based on the assumption that the reference point, Y^R, is fixed. We extend their model by introducing strategic framing, e.g. the principal′s (her) ability to influence the agent′s (his) reference point. 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.

In *stage 1*, the risk-neutral principal designs the contract $Y(s)$[6] that she offers to the risk-averse agent. The agent derives utility $U(.)$ from each payment. The design comprises the choice of the base wage, which is assumed to be the reference payment, Y^R, 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. The principal will only offer contracts that are optimal, meaning that summed over all states of the world the offered payment scheme maximizes the principal′s profits and will give the required incentives to the agent (formalized in the incentive compatibility constraint, IC). We assume hereafter that the principal wants to induce high effort ē. The agent decides to accept or reject the contract in *stage 2*. The agent accepts if the total (expected) utility is higher than his reservation utility, V^*, he rejects otherwise. This intuition is stated in the participation constraint (PC).

If the agent accepts the contract, in *stage 3* he chooses an effort $e \in \{\underline{e}, \bar{e}\}$. The cost of effort for the agent is denoted as $c(e)$ which is assumed to be linear and strictly increasing in effort. The principal can only verify the state of the world $s \in [\underline{s}, \bar{s}]$, on which nature decides in *stage 4*. For any fixed payment scheme, the higher the outcome, the better off the principal. Importantly, the state of the world is imperfectly correlated to the effort level the agent provided through the conditional distribution function $F(s|e)$ and the resulting conditional density function $f(s|e)$. The uncertainty lies in the principal′s inability of observing the exerted effort of the agent. Each state s has positive probability of occurrence under each effort level, $\{f(\underline{s}|\underline{e}); f(\underline{s}|\bar{e})\} > 0$. Assuming the monotone likelihood ratio property (MLRP)[7] $\frac{\Delta f(s|e)}{f(s|\bar{e})}$ is "increasing in the state realization" (De Meza & Webb, 2007, p.71), meaning that likelihood of reaching a higher state s with low

[6] How s is defined and determined is explained below.

[7] $\Delta f(s|e) = f(s|\bar{e}) - f(s|\underline{e})$

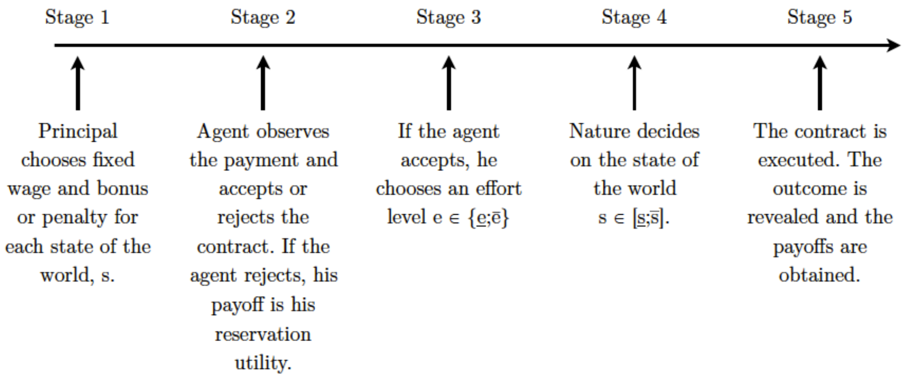

Figure 1: Timeline of the principal-agent problem

effort $\underline{e}$ relative to the likelihood with high effort $\bar{e}$ is decreasing with increasing state realizations. The intuition behind this assumption is that the probability of occurrence of each state of nature positively, but not perfectly, depends on the effort of the agent. Higher effort of the agent should lead to a higher probability of reaching a given state s and also to higher payments in each state. The optimal payment scheme is then monotonically increasing in the state realization s. The MLRP also implies that the principal who is better off with higher states will also prefer higher effort levels.

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 never profit-maximizing for the principal. With the Inada condition, we will find an internal solution.

In *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, $W(s) = s$. She maximizes her profits, $Y(s) = Y^R + b(s) : \int_{\underline{s}}^{\bar{s}} (s - Y(s)) f(s) ds$. The maximization problem with respect to profits is equivalent to minimizing the cost of employing the agent for a pre-specified effort level (Mas-Colell et al., 1995, p.483 and p.487):

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

The overall utility the agent derives from his payoff is a combination of standard expected utility, $U(Y(s))$, and reference-dependent utility with loss aversion, $Z(Y^R, b(s))$, defined as :

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

with

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

For a linear loss aversion coefficient, l:

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

for $l > 0$ 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.

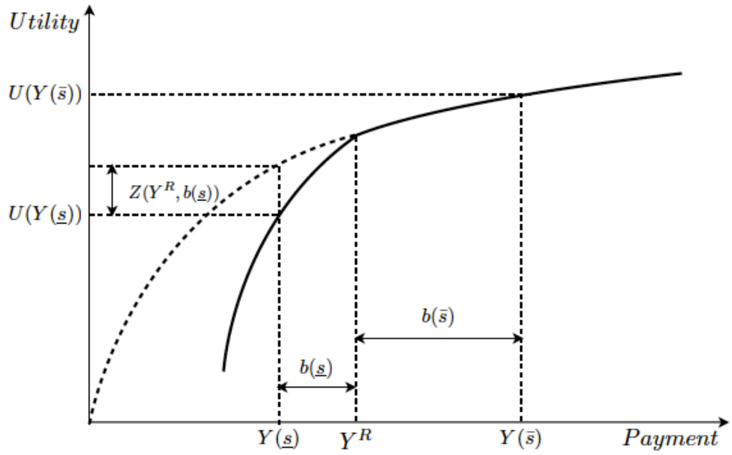

Figure 2: Overall utility function of the agent net of e

2.1.1 Reservation independence

The mathematical derivation of the stages above follows. The principal minimizes his costs of employing the agent subject to the participation and incentive compatibility constraint of the agent, given that the effort level $\bar{e}$ is to be elicited (with $\Delta e = \bar{e} - \underline{e}$):

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

s.t.

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

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

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

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

This variable 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 prospect-theoretic utility function below the reference point to include loss aversion.

Lagrangian optimization gives first-order conditions (FOCs) (9) - (12), with γ and λ corresponding to the multipliers of the participation constraint and the incentive compatibility constraint, respectively[8].

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

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

[8]The derivatives with respect to $b(s)$ and Y^R only differ if the bonus/penalty payment is evaluated differently than the reference income. This would result in two utility functions. We are assuming that the agent evaluates the bonus/penalty in the same fashion as the reference income, subscripts are therefore left out.

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

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

The intuition behind (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 (12) is increasing in the reference income in the loss region for a risk averse agent, (9) decreases with the reference income. Equation (11) is an extended form of the standard solution for incentive payments:

$$\frac{1}{U'(Y^R + b(s))} = (1+\theta l)(\gamma + \lambda \frac{\Delta f(s|e)}{f(s|\bar{e})}) \tag{13}$$

For the incentive payment to be optimal, (13) needs to be satisfied, given the reference income. This in turn means that equation (13) needs to be met with equality for all reference incomes, for optimal and non-optimal ones. Obviously, this then also holds for the cases of strategic framing. The reasoning as to why the multipliers are not equal to zero is provided in the appendix.

Proposition 1

Given MLRP and loss aversion, five shapes of the optimal incentive schemes are possible as a function of each level of the reference income, $\bar{Y}^R$.

We are assuming that

$$(1+l)(\gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})}) < \gamma + \lambda \frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})} \tag{14}$$

as otherwise it would be possible that we have a flat payment scheme, which is not incentive compatible.

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 smoothly increasing in performance and only bonuses are paid; all $b(s) \geq 0$.

2. Penalty contract:

   If $\frac{1}{U'(\bar{Y}^R)} \geq (1+l)(\gamma + \lambda \frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})})$, the payment scheme is smoothly 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 $\gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})} \leq \frac{1}{U'(\bar{Y}^R)} < (1+l)(\gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})})$, the scheme pays the reference income $\bar{Y}^R$ up to some threshold beyond which it is smoothly 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 increasing smoothly up to some threshold, beyond which the reference income $\bar{Y}^R$ is paid.

5. Bonus and penalty contract:

   If $(1+l)(\gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})}) < \frac{1}{U'(\bar{Y}^R)} < \gamma + \lambda \frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})}$, the reference income $\bar{Y}^R$ is paid for some finite, compact performance interval, above and below some threshold level though, the payment scheme is smoothly increasing.

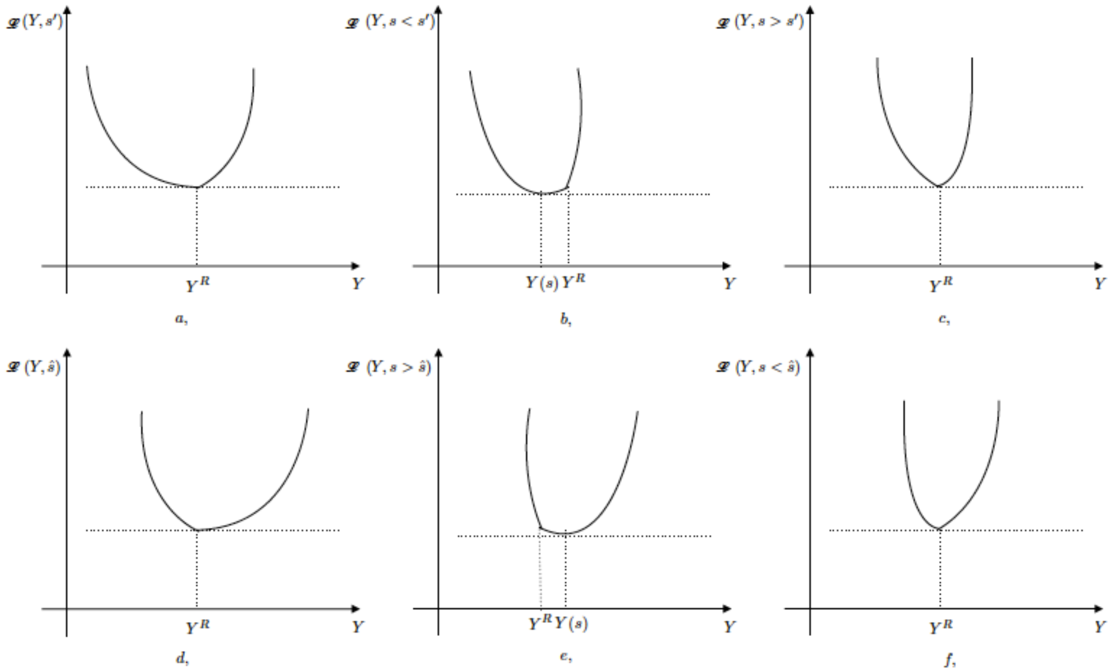

Figure 3: Proposition 1.

Proof. **Step 1.** Suppose that the assumption (14) is not satisfied and the reference income conforms to:

$$\gamma + \lambda\frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})} \leq \frac{1}{U'(\bar{Y}^R)} \leq (1+l)(\gamma + \lambda\frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})}) \tag{15}$$

With $b(\underline{s}) \leq 0$ and $b(\bar{s}) \geq 0$ by definiton. By MLRP, $b(\underline{s}) = 0$ and $b(\bar{s}) = 0$. Therefore, $\bar{Y}^R$ is paid in all states s and the payment scheme is independent of performance. This is not incentive compatible and can therefore never be an optimal solution to the minimization problem.

Step 2. Description of the behaviour of the Lagrangian around the threshold levels, 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 equation (11) gives the following:

$$\frac{1}{U'(Y^R + b(s))} - (1+\theta l)(\gamma + \lambda\frac{\Delta f(s|e)}{f(s|\bar{e})}) \geq 0 \tag{16}$$

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

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

For the state s', we are at the minimum because for any deviation from the payment of Y^R equation (17) is different from zero: By making $b(s)$ 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 (17) decreases discretely, as moving to the gains region lets the loss aversion measure disappear. The derivative is positive now and decreasing the bonus is equivalently decreases the costs of the principal. Figure 3 depicts the Lagrangian minimand $\mathcal{L}(Y, s)$ as a function of Y for fixed s. The above described case is represented in a.

Then for $s < s'$, the second term of equation (17) decreases by the MLRP and the FOC is positive. 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, penalties are paid, e.g. $b(s) < b(s') = 0$, and the payment scheme is smoothly increasing up to s'.

Then for $s > s'$[9] which is depicted in Figure 3c, paying $Y(s') < Y^R$, turns the FOC negative by the MLRP. Increasing the payment is cost-minimizing. On the contrary, paying $Y(s') > Y^R$ turns the FOC positive. 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:

$$\frac{1}{U'(\bar{Y}^R)} - (\gamma + \lambda \frac{\Delta f(\hat{s}|e)}{f(\hat{s}|\bar{e})}) = 0 \tag{18}$$

For the state $\hat{s}$, we are at the minimum because for any deviation from the payment of Y^R equation (18) 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 (18) 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.

Then for cases where $s > \hat{s}$, depicted in Figure 3e, by the MLRP the first derivative in equation (18) is negative if $Y(s) = Y^R$. Increasing the payment for s is cost-minimizing as the FOC continues to be equal to zero. Therefore, $b(s) > b(\hat{s}) = 0$ and the payment scheme is smoothly increasing for any state $s > \hat{s}$.

For states just below $\hat{s}$, but still above s', the second part of equation (18) decreases by MLRP, implying that for $\bar{Y}^R$, the FOC is positive. 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 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 $\bar{Y}^R$.

1. Bonus contract:

   If $\frac{1}{U'(\bar{Y}^R)} - (\gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})}) \leq 0$, 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. The states s' and $\hat{s}$ are not in $s \in [\underline{s}, \bar{s}]$.

2. Penalty contract:

   If $\frac{1}{U'(\bar{Y}^R)} - (1+l)(\gamma + \lambda \frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})}) \leq 0$, each payment $Y(s)$ must be smaller than $\bar{Y}^R$, 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 $(\gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})}) < \frac{1}{U'(\bar{Y}^R)} \leq (1+l)(\gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})})$ the situation can be described as in Figures 3c and f. 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 $(\gamma + \lambda \frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})}) \leq \frac{1}{U'(\bar{Y}^R)} < (1+l)(\gamma + \lambda \frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})})$, 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:

[9] It is important to note here, that the state s is just above s' and below $\hat{s}$, which is defined below.

If $(1+l)(\gamma+\lambda\frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})}) < \frac{1}{U'(\bar{Y}^R)} < (\gamma+\lambda\frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})})$, the situation is a mix between cases 1 and 4. The reference income is paid for all states between s' and $\hat{s}$ and the payment scheme is smoothly increasing for all states $s \in [\underline{s}; s']$ and $s \in [\hat{s}; \bar{s}]$.

QED

□

Some important insights can be drawn from the previous proof: With loss aversion, there will always be a flat region at the reference income in the payment scheme of the agent. The strictly increasing parts of the incentive scheme are always situated at the same position. For all states s for which the payments are in the gains region, it needs to be true that $\frac{1}{U'(Y(s))} = \gamma + \lambda\frac{\Delta f(s|e)}{f(s|\bar{e})}$. As γ, λ and the distribution function are fixed, the payments in the gains region are consequently fixed as well. Likewise, for all states s for which the payments are in the loss region, it needs to be true that $\frac{1}{U'(Y(s))} = (1+l)(\gamma+\lambda\frac{\Delta f(s|e)}{f(s|\bar{e})})$. As γ, λ, the loss aversion measure l and the distribution function are fixed for each case, the payments in the loss region are consequently fixed as well. With a fixed reference income, $\bar{Y}^R$, and possible states $s \in [\underline{s}, \bar{s}]$ defined, the optimal incentive scheme is predetermined. The optimal payment scheme is determined by the minimum possible reference income and the location of the lowest and highest possible state. From these properties of the (possible) optimal incentive schemes it can be derived that framing payments as gains is optimal for the principal.

Assuming a variable reference income, Y^R, and the possibility of strategic framing by setting the base wage of the agent, 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 minimizing her costs by decreasing the reference income as far as possible. As in the Figure 4 it can be seen that the payments to the agent with a low reference income (Y_2^R) are always below or equal to the ones with a higher reference income (Y_1^R). 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 constraint on the base wage, like a minimum wage, then the putting Y^R as low as possible is optimal (we assume that the principal is not able to require the agent to pay for the contract, therefore the reference income needs to be non-negative). As long as the contract only includes bonus payments to the agent, the level of the reference income is irrelevant.

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. In each state s the principal pays the same or less to the agent (in at least one of the states she pays less). Suppose now, that Y_2^R is the lowest possible reference income for which the agent is still willing to participate in the contract. Dependent on where the states are situated, the optimal payment scheme is given. Consider the change from $\underline{s}$ to $\underline{s}'$: 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). This intuition is expressed formally in the following propositions.

Proposition 2

In the model with a value of the outside option that is not affected by the reference point, the principal never frames payments as losses. All payments take the form of a base wage plus a bonus (such that the condition in proposition 1.1. is satisfied). Once this is true, the size of the reference point 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 to take the derivative with respect to Y^R. We divide the derivative for three sets of states: from $\underline{s}$ to s', from s' to $\hat{s}$ and from $\hat{s}$ to $\bar{s}$. As we are considering infinitely small changes, the integral's borders are not changed.

$$min_{b(s),Y^R} Y^R + \int_{\underline{s}}^{\bar{s}} b(s)f(s|\bar{e})ds \tag{19}$$

s.t.

$$[PC]\int_{\underline{s}}^{s'}[U(Y^R+b(s))-l(U(Y^R)-U(Y^R+b(s)))]f(s|\bar{e})ds+U(Y^R)\int_{s'}^{\hat{s}}f(s|\bar{e})ds$$
$$+\int_{\hat{s}}^{\bar{s}}[U(Y^R+b(s))]f(s|\bar{e})ds-\bar{e}\geq V^* \tag{20}$$

$$[IC]\int_{\underline{s}}^{s'}[U(Y^R+b(s))-l(U(Y^R)-U(Y^R+b(s)))]\Delta f(s|e)ds+U(Y^R)\int_{s'}^{\hat{s}}\Delta f(s|e)ds$$
$$+\int_{\hat{s}}^{\bar{s}}[U(Y^R+b(s))]\Delta f(s|e)ds\geq\Delta e \tag{21}$$

$$1-\gamma[\int_{\underline{s}}^{s'}[U'(Y^R+b(s))-l(U'(Y^R)-U'(Y^R+b(s)))]f(s|\bar{e})ds$$
$$+(U'(Y^R+b(s'))-l(U'(Y^R)-U'(Y^R+b(s'))))f(s'|\bar{e})+U'(Y^R)\int_{s'}^{\hat{s}}f(s|\bar{e})ds-U'(Y^R)f(s'|\bar{e})$$
$$+U'(Y^R)f(\hat{s}|\bar{e})+\int_{\hat{s}}^{\bar{s}}[U'(Y^R+b(s))]f(s|\bar{e})ds-U'(Y^R+b(\hat{s}))f(\hat{s}|\bar{e})]$$
$$-\lambda[\int_{\underline{s}}^{s'}[U'(Y^R+b(s))-l(U'(Y^R)-U'(Y^R+b(s)))]\Delta f(s|e)ds$$
$$+(U'(Y^R+b(s'))-l(U'(Y^R)-U'(Y^R+b(s'))))\Delta f(s'|e)+U'(Y^R)\int_{s'}^{\hat{s}}\Delta f(s|e)ds$$
$$-U'(Y^R)\Delta f(s'|e)+U'(Y^R)\Delta f(\hat{s}|e)$$
$$+\int_{\hat{s}}^{\bar{s}}[U'(Y^R+b(s))]\Delta f(s|e)ds-U'(Y^R+b(\hat{s}))\Delta f(\hat{s}|e)]\geq 0 \tag{22}$$

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

$$f(s|\bar{e})-\gamma[U'(Y^R+b(s))-\theta l(-U'(Y^R+b(s))]f(s|\bar{e})-\lambda[U'(Y^R+b(s))-\theta l(-U'(Y^R+b(s))]\Delta f(s|e)=0 \tag{23}$$

Or expressed differently, first for losses and subsequently for gains:

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

$$\int_{\hat{s}}^{\bar{s}}f(s|\bar{e})ds-\gamma\int_{\hat{s}}^{\bar{s}}[U'(Y^R+b(s))]f(s|\bar{e})ds-\lambda\int_{\hat{s}}^{\bar{s}}[U'(Y^R+b(s))]\Delta f(s|e)ds=0 \tag{25}$$

Substituting this into equation (22) and given that $b(s')=0$ and $b(\hat{s})=0$ gives the following inequality:

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

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

$$-\lambda[-lU'(Y^R)\int_{\underline{s}}^{s'} \Delta f(s|e)ds + U'(Y^R)\int_{s'}^{\hat{s}} \Delta f(s|e)ds] \geq 0 \tag{26}$$

From FOC (11) we know that equation (26) is true as for all s at Y^R:

$$f(s|\bar{e}) - \gamma[U'(Y^R)f(s|\bar{e})] - \lambda[U'(Y^R)\Delta f(s|e)] > 0 \tag{27}$$

QED

□

2.1.2 Reservation dependence

Until now we have assumed that the principal offers the agent a contract and the latter compares the payments he receives for each state s to the base wage. He then evaluates the contract and takes his decision. This formulation of the problem does not include the evaluation of the reservation utility; V^* is exogenous and does not depend on the relative magnitude to Y^R. The reservation utility is generally defined as outside opportunities the market offers the agent (Macho-Stadler & Pérez-Castrillo, 2001, p.20). If the agent is offered a fixed wage contract by another principal, this salary would be his reservation income. We think that it is highly unlikely that an agent with reference-dependent preferences will not compare this fixed wage to the base wage of the contract. The intuition behind this is that human beings take several aspects into account in their decision making rather than looking at each possibility separately, e.g. they compare their options. It is also logical, that the agent compares his base wage to the fixed wage of the other contract rather than the weighted expected payments for each state. Therefore, we will incorporate this in the model by changing the reservation utility into an outside option that is compared to the base wage of the contract.

The minimization problem of the principal changes if the reservation utility is dependent on the reference income. We therefore define the reservation utility, V^*, as utility derived from the outside option of the agent, P, evaluated from the view of the reference income Y^R in the contract offered by the principal. The agent derives utility in the same manner from the outside option as from the reference income or the bonus payments, therefore the same utility function is applicable.

Based on these considerations, we will define the reservation utility as follows, keeping the assumption of linear loss aversion:

$$V^* = \begin{cases} U(P) & if \quad P \leq Y^R \\ U(P) - l(U(Y^R) - U(P)) & if \quad P > Y^R \end{cases} \tag{28}$$

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

$$\int_{\underline{s}}^{\bar{s}} (U(Y^R + b(s)) - \theta l(U(Y^R) - U(Y^R + b(s))))f(s|\bar{e})ds - \bar{e} \geq U(P) - \eta l(U(Y^R) - U(P)) \tag{29}$$

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

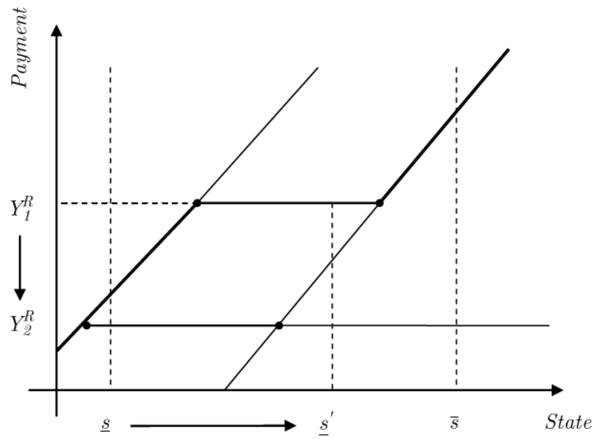

Figure 4: Proposition 1-2.

With:

$$\eta = \begin{cases} 1 & if \quad P > Y^R \\ 0 & if \quad P \leq Y^R \end{cases} \quad (31)$$

Proposition 3 In the model with a value of the outside option that *is* affected by the reference point, the principal may frame payments as losses.

Proof. This proof is basically identical to the one for Proposition 2: As can be seen in equation (30), only an additional negative term enters the first-order condition. This FOC can therefore still be valid with equality which makes contracts with penalties possibly optimal.

QED □

3 Discussion

Assuming that the principal can influence the reference income by setting a base payment is the key point of this paper. The nature of the reference income is matter of much debate in both the theoretical and experimental literature. As framing payments as gains for the agent is optimal, the question arises in this case whether the agent would still see the low base wage as reference income. He will always receive more than the base wage and could rather take as reference income the expected income, the median income, the status quo or the certainty equivalent. Individual decision-makers are subject to some psychology that makes them form the reference point automatically. The alternatives to the base wage are reasonable and worth researching. This would mean for the present model that the reference income becomes endogenous and the principal can only influence the reference income indirectly. It is the question how the principal is able to frame the contract. If the agent works in a team, even peer comparisons can have an influence on the reference income. Consistent experimental testing of these potential explanations is missing, but they are necessary to build meaningful models for optimal compensation schemes. Incorporating the experimental outcomes into this existing model will give more powerful tools for designing efficient contracts.

Connected to the presumption of a reference income for the agent is the question of the possibility of strategic framing. Our model investigates the direct version which would change with an alternative assumption of the reference income. With an endogenous model, the reference income could only be modified indirectly through the relative magnitudes of the base wage and the bonuses or penalties.

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.

Obviously, the optimal contract set out in this paper is not suited for all principal-agent relationships. 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. 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 bonus payments) and effective control mechanisms can 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.

4 Conclusion

Our paper modifies De Meza and Webb's (2007) model by making the exogenous fixed reference income variable and allowing for it to be directly influenced by the principal. Strategically framing the contract by setting a base wage deletes four of the five possible payment schemes in De Meza and Webb. The cost-minimizing solution for the principal is to pay only bonuses to the agent when the outside option is independent of the reference income. This still stays optimal if the outside option is evaluated against the reference income, though it might not be possible to do so. Then penalty contracts can become optimal.

Further research focuses on testing the five possible payment schemes in practice. Experiments will 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 is correct or not. Still the question as to the nature of the reference point remains. It needs to be considered whether further psychological factors influence the formation of the reference point.

Furthermore, important other specifications are necessary to be investigated. 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 diverse options regarding the underlying performance variable also alters the payment scheme. 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.

5 Appendix

5.1 Asian Disease Experiment (Kahneman & Tversky, 1984)

The participants in this experiment were given the following problem description and they were asked to choose between two alternatives in two versions with different frames:

> Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed.

Version 1 of the decision problem is framed with the reference point of saving people:

> If Program A is adopted, 200 people will be saved. If Program B is adopted, there is a $\frac{1}{3}$ probability that 600 people will be saved and $\frac{2}{3}$ probability that no people will be saved.

Version 2 of the decision problem is framed with the reference point of people dying:

If Program C is adopted, 400 people will die. If Program D is adopted, there is a $\frac{1}{3}$ probability that nobody will die and $\frac{2}{3}$ probability that 600 people will die.

Although choices A and C, B and D are equivalent, 72% of the participants chose A and in the second version 78% chose D.

5.2 Multipliers

Proof. The multipliers are not zero.

Suppose that $\gamma = 0$; by first-order stochastic domination (FOSD), there must be a range of low states s where $f(s|\underline{e}) > f(s|\bar{e})$ and it follows that then $\Delta f(s|e) < 0$. The right-hand side of (13) 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 effort, 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 (13) these would be the following:

$$\frac{1}{U'(Y(\underline{s}))} = \gamma(1+l) \tag{32}$$

as there is loss aversion and

$$\frac{1}{U'(Y(\bar{s}))} = \gamma \tag{33}$$

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

References

Barberis N., M., Huang. (2008). Handbook of the equity risk premium: The loss aversion / narrow framing approach to the equity premium puzzle. In R. Mehra (Ed.), (pp. 199–229). Elsevier Science.

Bebchuk, L., & Grinstein, Y. (2005). The growth of executive pay. *Oxford Review of Economic Policy*, *21*(2), 283-303.

Benartzi, S., & Thaler, R. (1995). Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics*, *110*(1), 73-92.

Camerer, C. (2000). Prospect theory in the wild: Evidence from the field. In D. Kahneman & A. Tversky (Eds.), *Choices, values and frames* (10th ed., p. 288-300). Cambridge: Cambridge University Press.

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

De Meza, D., & Webb, D. (2007). 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.

Genesove, D., & Mayer, C. (2001). Loss aversion and seller behavior: Evidence from the housing market. *The Quarterly Journal of Eocnomics*, *116*(4), 1233-1260.

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

Hallahan, K. (1999). Seven models of framing: Implications for public relations. *Journal of Public Relations Research*, *11*(3), 205-242.

Harris, J. (2009). What's wrong with executive compensation? *Journal of Business Ethics*, *85*, 147-156.

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.

Kahneman, D., & Tversky, A. (1984). Choices, values and frames. In D. Kahneman & A. Tversky (Eds.), *Choices, values and frames* (10th ed., p. 1-16). Cambridge University Press.

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). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, *76*(2), 149-188.

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.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, *106*(4), 1039-1061.