

Moral hazard: Base models and two extensions*

Inés Macho-Stadler[†] and David Pérez-Castrillo[‡]

Abstract

We analyze the optimal contract in static moral hazard situations, where the agent's effort is not verifiable. We first present the main trade-offs of the principal-agent model. We cover the trade-off of incentives (motivation) vs. risk-sharing (efficiency), incentives vs. rents (when the agent is protected by limited liability), incentives to a task vs. incentives to another (in a multitask situation), and incentives to the agent vs. incentives to the principal (when both exert a non-verifiable effort). Then, we discuss two recent extensions: how incorporating behavioral biases in the analysis of incentives and inserting the principal-agent problem in a matching market affect the predictions of the classical moral hazard model.

JEL Classification numbers: D86, D03, C78.

Keywords: moral hazard, behavioral approach, matching.

*We thank Kaniska Dam, Mikhail Drugov, Pau Olivella, Joaquin Poblete, Pedro Rey-Biel, and Nicolas Roux for very useful comments. We also thank the Department FAE2 of the Universidad del País Vasco where this project was partially developed. We gratefully acknowledge financial support from the research grants (ECO2012-31962 and ECO2015-63679-P), Generalitat de Catalunya (2014SGR-142), ICREA Academia, and Severo Ochoa Programme (SEV-2015-0563). The authors are fellows of CESifo and MOVE.

[†]Universitat Autònoma de Barcelona and Barcelona GSE. Email: ines.macho@uab.cat

[‡]Universitat Autònoma de Barcelona and Barcelona GSE. Email: david.perez@uab.cat

1 Introduction

Moral hazard (also called hidden action), the informational asymmetry related to the agent's behavior during a relationship, has been a long-time concern for insurance. It is said that the term moral hazard was coined in the nineteenth century by fire insurers to differentiate among the various hazards that cause a fire: physical hazards, both the ones related to the causes (lightning, short circuits) and the ones affecting the magnitude of the loss (type of construction), and moral hazards associated with insurees' behavior (less precautions or careless behavior).¹ Since Arrow (1963, 1968, 1971) and Pauly (1968), models of moral hazard and its applications have increasingly been recognized as key elements in understanding sharecropping contracts, corporate governance, licensing agreements, and executive compensations, to cite just a few examples.² Moral hazard models are now taught in many undergraduate majors and most graduate programs.³

In this chapter we review the literature on moral hazard in static environments. In its simplest version, a moral hazard problem is presented in the contractual relationship of a principal (she) and an agent (he) that works for her on a project.⁴ The effort of the agent determines the probability distribution of the project's outcome.⁵ There is a moral hazard problem when it is not possible to verify the agent's effort. This implies that effort cannot be contracted upon, because in the case of breach of contract, no court of law could know if the contract had really been breached or not. In this case, once hired, the agent will decide the level of effort that he prefers, taking into account how payments change with the outcome, that is, given the payment scheme that he has accepted. The payment scheme is the indirect way in which the principal can sway the agent's behavior.

The shape of the optimal payment scheme comes from the maximization of the principal's benefit subject to two constraints: the agent

¹Aetna Insurance co. (1867).

²The first efforts toward understanding and solving the principal-agent problem were due to Zeckhauser (1970), Spence and Zeckhauser (1971), Ross (1973), Stiglitz (1974), Mirrlees (1975, 1999), Harris and Raviv (1979), Holmström (1979), and Shavell (1979a, b).

³Several textbooks cover moral hazard problems along with adverse selection situations (Mas-Colell, Whinston, and Green, 1985, Macho-Stadler and Pérez-Castrillo, 1997, Salanié, 1997, Laffont and Martimort, 2002, and Bolton and Dewatripont, 2005).

⁴The participants can be individuals or institutions. Examples are bank regulator and bank, shareholders and manager, and insurer and insuree.

⁵In this chapter we will refer to the agent's effort, but the agent may be taking a decision or an action.

participation constraint (the agent will only sign the contract if by doing so he obtains at least as much as his opportunities outside this relationship), and the incentive compatibility constraint (that recognizes that the agent will choose the effort that is best for him given the contract).

The general moral hazard problem is not easy to analyze. However, some simple set-ups have been very successful when adopted to study particular situations. First, it is generally assumed that the principal is risk neutral, and the agent's utility is separable in payment and effort. Still, these hypotheses do not allow us to have a simple enough problem. Second, it is often assumed that the "first-order approach" (FOA) is valid or that the agent chooses among a finite number (usually two) of possible efforts. While several interesting properties of the optimal contract can be derived thanks to these hypotheses, they do not allow the general derivation of explicit solutions. Thus, in many extensions and applications, further simplifications are used in order to find specific solutions. We will describe and bring into play two of these specifications that consider particular functional forms for the agent's utility function combined with certain assumptions on the payment scheme: the case of constant absolute risk aversion (CARA) utility function with linear contracts and the case of risk neutrality with limited liability.

The purpose of this chapter is neither to explain every aspect of the moral hazard problem nor to review each extension or topic. Moreover, for the sake of space, we focus on theoretical models, and we do not cover empirical or experimental results. We have chosen to present the main trade-offs of the principal-agent model and to discuss two extensions that we find particularly interesting: including behavioral considerations and an analysis of the market assignment that determines the partnerships that are formed. The first extension aims to discuss how incorporating behavioral biases in the analysis of incentives may affect the predictions of the classical moral hazard model. We discuss the effect of some of the strands of the literature. We start by considering an agent who not only takes into account his own well-being but also has other-regarding preferences. We then discuss the role of extrinsic and intrinsic motivations and the consequences on the optimal contract. We also cover the literature that concentrates on loss aversion, where the agent evaluates his payoffs not in absolute terms but in comparison with some reference. Finally, we consider the papers which focus on the idea that agents may be optimistic about the production process or overconfident about their ability.

The second extension we present relates to the insertion of the principal-agent problem in a matching market. It is easy to motivate this avenue from the point of view of the agency models. The partial equilibrium

approach characterizes the optimal wage scheme when a principal hires an agent (a given pair Principal-Agent). This approach is well-defined for the case where there is a single principal in the economy or when principals are perfectly competitive and hence get zero profits. In the classical approach, the bargaining power is given to principals or agents by assumption, which implies that the reservation utility or the zero profit condition determines the distribution of surplus. In other words, the effects of competition are summarized by a single parameter of the agent's outside option (his reservation utility) or the principals' zero profit condition. However, when we consider explicitly the existence of several heterogeneous principals and several heterogeneous agents, some of the properties obtained in the simple version of the agency problem do not necessarily hold. Thus, empirical work and policy recommendations may be based on the wrong arguments. Moreover, we can address the endogenous determination of the principals and agents that meet.

2 Base Moral Hazard Models

A principal hires an agent to perform a task that we refer to as effort, $e \in E$, in exchange for a wage, w . The final outcome of the relationship, x , depends on the effort e that the agent devotes to the task and some random variable for which both participants have the same prior distribution. The set of possible outcomes is denoted by X , which can be a continuous set, in which case we denote $X = [\underline{x}, \bar{x}]$, or a discrete set. The distribution of the random variable induces a probability $p(x | e) > 0$ of outcome $x \in X$ conditional on effort e , where $p(\cdot | e)$ is a twice continuously differentiable density function if X is a continuous set and it is a vector of probabilities if X is discrete. We denote by $P(x | e)$ the cumulative distribution function, that is, $P(x | e) = \int_{y \leq x} p(y | e)$ (or $P(x | e) = \sum_{y \leq x} p(y | e)$ if X is discrete).

Since uncertainty exists, participants may react to risk. We concentrate on the case of a risk-neutral principal and a (possibly) risk-averse agent. Risk preferences are expressed by the shape of their (von Neumann-Morgenstern) utility functions. The principal, who owns the outcome x and must pay the agent w , has preferences represented by the utility function $x - w$. The agent, who receives a monetary payoff w for his participation in the relationship and supplies an effort e , has an additively separable utility function: $U(w, e) = u(w) - v(e)$, where $u(w)$ is assumed to be increasing and concave and $v(e)$ is increasing and convex.⁶ The agent can obtain a utility level \underline{U} outside the relationship

⁶The key characteristic of this class of utility functions is that the agent's risk aversion (preferences over lotteries) is independent of the effort supplied. Grossman and Hart (1983) assume the most general utility function by considering that the

with the principal. Therefore, he only accepts contracts that give him, in expectation, at least \underline{U} .

Under symmetric information, that is, when effort is verifiable, the optimal (first-best) contract includes the first-best effort e^{FB} and the payment scheme $(w^{FB}(x))_{x \in X}$, which incorporates the optimal risk sharing among the two participants. If the agent is risk averse then the payment mechanism completely insures the agent: he receives a fixed payment.

Under moral hazard, the effort is not contractible and the agent can choose the effort that is best for him, given the contract. Thus, if the principal proposes a fixed wage, the agent's payment does not depend on his effort and he will choose the effort that is least costly for him, that is, the lowest possible level of effort.

When designing the optimal incentive contract for the moral hazard problem, the principal gets the agent interested in the consequences of his own behavior by making his payoff dependent on the outcome. If the agent is risk averse, given that the outcome is noisy, this entails the cost of distorting the optimal risk sharing among both participants. In this case, the optimal contract solves the trade-off between distorting the efficient allocation of risk and providing incentives.

The optimal contract under moral hazard takes into account the acceptance condition for the agent and his choice of effort. Moreover, it is often the case that arbitrarily low or high payments are not feasible, which would introduce additional constraints into the principal's program. For example, the agent may have limited liability so that it is not possible to impose a penalty on him (or he should receive a minimum legal wage independent of the outcome). Similarly, it may not be possible for the principal to pay the agent more than the value of the outcome, or she may be constrained (by law or by norm) not to pay too much to the agent. An example of upper bounds are the European Union regulatory cap on bankers' bonus payments such that "the maximum ratio between the variable and the fixed part of the total remuneration is limited to 100%." When there are lower and/or upper bounds to the agent's payment, new trade-offs may appear. For example, the implementation of some (high) efforts may not be possible because there is no room for enough variation in payments or it may become very expensive as it requires awarding the agent an expected utility that is higher than his reservation utility. In the latter case, there is a trade-off between giving extra rents and providing incentives.

agent's utility has the form $U(w, e) = K(e)u(w) - v(e)$. Special cases are $K(e) = 1$, i.e., additively separable preferences, and $v(e) = 0$, i.e., multiplicatively separable preferences.

The timing of the relationship between the principal and the agent is the following. First, the principal decides on the contract she offers to the agent, in particular on the agent's payment scheme $(w(x))_{x \in X}$ as a function of the outcome of the relationship. Then the agent decides whether or not to enter the relationship. Finally, if the contract is accepted, the agent chooses the effort level e that he most desires, given the agreed contract. This is a free decision by the agent because effort is not a verifiable variable. The principal bears this fact in mind when she designs the contract that defines the relationship, and the "game" can be solved by backward induction. Formally, if we first consider a situation without lower or upper bounds on salaries, the optimal contract under moral hazard is the solution to the maximization problem (*P1*):

$$\begin{aligned} & \underset{\{(w(x))_{x \in X}, e\}}{\text{Max}} \quad \{E(x - w(x) \mid e)\} \\ & \text{s.t.} \quad E(u(w(x)) \mid e) - v(e) \geq \underline{U} \quad (\text{PC}) \\ & e \in \arg \max_{\hat{e} \in E} \{E(u(w(x)) \mid \hat{e}) - v(\hat{e})\}, \quad (\text{ICC}) \end{aligned}$$

where $E(y \mid e)$ denotes the expectation of y conditional on the effort e . The first restriction of the program is the *participation constraint* (PC), which states that the agent will not sign a contract that gives him lower expected utility than the alternative market opportunities. The second restriction is the *incentive compatibility constraint* (ICC), which determines the agent's effort under moral hazard. If the ICC is not relevant (either because there is symmetric information or because ICC is not binding at the optimum) then the solution to the program is the first-best contract $\left((w^{FB}(x))_{x \in X}, e^{FB} \right)$.

The solution to program (*P1*) provides the optimal contract under moral hazard and the optimal level of the principal's utility for a given level of the reservation utility \underline{U} . As the level of \underline{U} changes, we obtain the Pareto frontier in the space of the utilities of the two participants. Thus, the main quality properties of the optimal contract hold if instead of considering (PC) we maximize the agent's utility subject to a participation constraint for the principal.

The main difficulty in solving the general program (*P1*) is related to the fact that the incentive compatibility constraint is itself a maximization problem. To overcome this obstacle, the literature has adopted two solutions. (a) If the set E is finite (most papers that follow this approach consider E to include two levels of effort) then the ICC can be replaced by a finite set of inequalities (just one inequality in the case of two efforts). (b) If the set E is a continuum, say $E = [\underline{e}, \bar{e}]$, then we can try to substitute the ICC by its first-order condition, which is

a necessary condition of the optimal e if it is interior. This is called the *first-order approach* (FOA). One has to be careful if one follows this approach because the agent's expected utility may fail to be concave in effort. Hence, using the FOA may be incorrect, and finding the optimal effort in this program difficult.⁷ A possible way out proposed by Grossman and Hart (1983) is to solve the problem in steps, first identifying the optimal payment mechanism for any effort and then, if possible, finding the optimal effort.⁸ The other possibility is to consider situations where the agent's maximization problem is well-defined, which requires introducing assumptions for the FOA to be valid.

The moral hazard problem may give rise to several distortions in the optimal contract because it forces the principal to trade-off incentives for the effort of the agent and other objectives. We now discuss characteristics of the solution of (P1) for several cases, emphasizing the trade-offs faced by the principal. We will make it clear in some cases that additional constraints are added to (P1) owing for example to the existence of bounds on the payments.

2.1 Incentives vs. risk-sharing

We first analyze the consequences of moral hazard in situations where the agent is risk averse, that is, $u(w)$ is strictly concave. In this case, the optimal, first-best contract fully insures the agent. However, providing incentives requires that the agent's salary depends on the outcome. Thus, the principal needs to trade-off incentives vs. risk-sharing.

We develop the analysis for three different models.

2.1.1 Model 1: Continuous effort

Consider a situation where E is continuous and the FOA is valid. Denote by λ (resp., μ) the Lagrangian multiplier of the PC (resp., the ICC). Then, for a given effort e , Holmström (1979) tells us that the solution to the principal's program (P1) with respect to the payoff scheme

⁷Mirrlees (1975) shows that the FOA is generally invalid unless the optimum effort derived from the ICC (the solution to the agent's maximum problem) is unique. In the absence of uniqueness, the first-order conditions of the principal's problem when the ICC is substituted by its first-order condition are not even necessary conditions for the optimality of the incentive contract. We describe the conditions at the end of section 2.1.

⁸Grossman and Hart (1983) show that this can always be done for additively or multiplicatively separable utility functions. By using the utilities of the wages instead of the wages, the principal's program with respect to the payment scheme for any effort can be rewritten as a minimization problem where the objective function is a convex cost function subject to (a possibly infinite number of) linear constraints. In particular, when the set of possible efforts E is a finite set, using Kuhn-Tucker one obtains necessary and sufficient conditions for optimality.

$(w(x))_{x \in X}$ satisfies, for all $x \in X$,

$$\frac{1}{u'(w^*(x))} = \lambda + \mu \frac{p_e(x | e)}{p(x | e)}, \quad (2)$$

where $p_e(x | e)$ is the partial derivative of $p(x | e)$ with respect to e . In the optimal contract, both PC and ICC are binding, that is, λ and μ are strictly positive. Their value depend on the effort e .⁹ The ratio $\frac{p_e(x|e)}{p(x|e)}$ is the *likelihood ratio* of obtaining outcome x when the effort is e .

The optimal scheme $(w^*(x))_{x \in X}$ and the multipliers λ and μ are characterized by the condition (2) for all $x \in X$ together with (PC) and (ICC). Therefore, the optimal wage scheme for a given effort e does not depend on the value that the principal places on the outcome: the value of x does not enter directly into any of these equations. If the wage is a function of the outcome it serves only as an incentive for the agent. Hence, it only depends on the outcome as long as the outcome is informative about the effort. In particular, the necessary and sufficient condition for a better outcome to always lead to a higher wage, that is, $w^*(x) > 0$, is that the likelihood ratio is increasing in x . This condition is called the *monotone likelihood ratio property* (MLRP), which holds when

$$\frac{p_e(x | e)}{p(x | e)} \text{ is strictly increasing in } x \quad (\text{MLRP})$$

for all $e > \underline{e}$.¹⁰

Moreover, MLRP together with CDFC (*convexity of the distribution function condition*), which are often called the Mirrlees-Rogerson sufficient conditions, are sufficient conditions for the validity of the FOA (Mirrlees, 1976, Rogerson, 1985, and Jewitt, 1988). We say that a distribution function satisfies CDFC if the second derivative of the cumulative distribution function $P(x | e)$ with respect to e is non-negative, that is,

$$\frac{\partial^2 P(x | e)}{\partial e^2} \geq 0.$$

Hence, the validity of the FOA requires demanding conditions on the

⁹If the agent is risk neutral, then the multiplier μ is zero and equation (2) only gives the value of the multiplier λ . In this case, any payment scheme whose expected payoff ensures the agent an expected utility level of \underline{U} is optimal.

¹⁰Holmström (1979), Shavell (1979a), and Milgrom (1981) show that under the FOA, if the distribution function satisfies MLRP then the wage scheme is increasing in output. Note that MLRP is stronger than first-order stochastic dominance, which requires that $\frac{\partial}{\partial e} P(x | e) < 0$ for all $x \in (\underline{x}, \bar{x})$.

probability function (MLRP and CDFC).^{11,12}

We make two additional remarks about the optimal contract. First, the wage scheme needs to be simpler as the agent has more room to manipulate the outcome. For example, if the agent can freely dispose of the output, the optimal payment mechanism is necessarily monotonic even if the MLRP does not hold. Alternatively, if there are several agents who can trade output among themselves, then only a linear scheme is feasible (any non-linear scheme will be “linearized” by arbitrage).¹³

Second, we have considered payment schemes that only depend on the outcome of the relationship (the outcome related to the effort is the only verifiable variable). However, the principal will base the contract on any signals that reveal information on the agent’s effort. Hence, if possible, the contract should be contingent on many other variables. Information related to the state of nature may be useful if it allows better estimates of the agent’s effort, thus reducing the risk inherent in the relationship. This is known as *the sufficient statistic result*, and it is perhaps the most important conclusion in the moral hazard literature (Holmström, 1979). Formally, we say that x is sufficient for $\{x, y\}$ with respect to $e \in E$ if and only if the distribution function p is multiplicatively separable in y and e :

$$p(x, y | e) \equiv g(x, e)h(y, x).$$

We say that y is informative about $e \in E$ if x is not sufficient for $\{x, y\}$ with respect to $e \in E$. Finally, if y is informative about $e \in E$ then there

¹¹MLRP and CDFC are very strong conditions and it is difficult to find distributions which satisfy both of them. The two-step procedure proposed by Grossman and Hart (1983) provide a way of proceeding when the FOA is not valid.

¹²Kirkegaard (2014) recently proposed a reformulation of the moral hazard problem that allows the use of results from the areas of choice under uncertainty. In this way, he can prove the classic results using an unifying methodology and also extend the analysis to larger domains than previous work.

¹³By considering additional properties of the participants’ objective function, more information on the optimal contract can be obtained. Imagine that the agent is “prudent,” in the sense that $u'' < 0$ and $u''' > 0$. A prudent agent is risk-averse and his marginal utility is strictly convex so he is downside risk-averse (Menezes et al., 1980). This agent applies a heavier discount to downward variations than to upward variations of the payment scheme. Chaigneau (2014) shows that concave contracts tend to provide more incentives to risk averse agents, while convex contracts tend to be more profitable to motivate prudent ones. The intuition is that concave payment schemes concentrate incentives where the marginal utility of risk averse agents is highest, while convex contracts protect against downside risk. However, when the principal is also risk averse and prudent, convex contracts are not optimal if the principal is sufficiently prudent relative to the agent (Sinclair-Desgagné and Spaeter, 2015).

is a payment mechanism $w(x, y)$ that strictly Pareto dominates the best $w(x)$.

The empirical content of the sufficient statistic argument is that the optimal contract should exploit all available information in order to optimally filter out risk.¹⁴ In the limit, if by including many variables the agent's effort can be inferred with certainty, then the symmetric information effort can be implemented at no extra cost.

Finally, once we have computed the optimal scheme for each e , that we denote $w^*(x, e)_{x \in X}$, the principal can find the *optimal effort* under the moral hazard problem by solving

$$\max_{e \in E} \{E(x - w^*(x, e) \mid e)\}.$$

The main difficulty of this program is that it is not generally concave in effort. If the principal's problem is well-defined and has a solution, the optimal effort e^* is determined by the usual condition of equality between marginal revenues and the marginal costs of increasing the effort, which includes the increase in average wages plus the extra cost in terms of the incentives needed to increase the effort.¹⁵ We notice that some efforts may not be implementable under moral hazard and that the lowest effort \underline{e} can always be implemented at no extra cost using the symmetric information wage scheme.

Example 1: CARA risk preferences and linear contract. A particularly simple, and very popular, model is one where the principal is risk neutral and the agent has CARA risk preferences:

$$u(w, e) = -\exp[-r(w - v(e))],$$

where r is the coefficient of absolute risk aversion. Additionally, assume that the cost of effort is a quadratic function

$$v(e) = \frac{1}{2}ve^2.$$

¹⁴For example, when the principal hires several agents, the central question is whether incentives should be provided as a function of all agents' performance. The answer comes from the sufficient statistic result and depends on the linkage of the agents' situation, in particular on whether the agents' outcomes are subject to correlated shocks (informational linkage) or whether the performance of an agent depends on the effort of other agents (technological linkage). See Holmström (1982) and Mookherjee (1984).

¹⁵It is interesting to note that under symmetric information the PC determines the optimal effort level, while it is the cost implied by the ICC which determines the effort when there is moral hazard. The reason is that under moral hazard and using the FOA the ICC implies that the derivative of the PC with respect to the effort is zero.

The output x depends on the agent's effort e and a random variable ε that is normally distributed with mean zero and variance σ^2 :

$$x = e + \varepsilon.$$

Finally, we restrict attention to linear wage schemes of the form $w = F + sx$, where F is a fixed payment and s is the share of the output that goes to the agent.¹⁶ In this case, it is convenient to solve the program by using the agent's certain equivalent income

$$F + se - \frac{1}{2}ve^2 - \frac{r}{2}\sigma^2,$$

in which case the ICC becomes very easy to write: $e = \frac{s}{v}$.

Solving the principal's program, the PC determines the fixed part of the contract F and the variable performance part of the contract is

$$s^* = \frac{1}{1 + rv\sigma^2}, \quad (3)$$

which is decreasing in the cost of the effort v , the agent's risk aversion (measured by r), and the variance of the outcome σ^2 . Since a higher s translates into a higher effort, the previous expressions reflect the trade-off between efficiency (optimal risk sharing would require $s = 0$) and incentives.¹⁷

¹⁶Although linear contracts are generally not optimal in the static setting (see Mirrlees, 1975), Holmström and Milgrom (1987) show that the optimal contract is linear in the final outcome if the agent chooses efforts continuously to control the drift vector of a Brownian motion process and he observes his accumulated performance before acting. Linear contracts are also shown to be optimal in models with limited liability and risk neutrality if the principal is uncertain about the technology available to the agent (see Carroll, 2015).

¹⁷In a multiagent situation, the sufficient statistic result is easy to illustrate when the principal hires two agents with CARA risk preferences and non-cooperative behavior. Linear contracts would have the form

$$w_i = F_i + s_i x_i + z_i x_{-i} \quad \text{for } i = 1, 2.$$

When $z_i \neq 0$ there is relative performance evaluation. Suppose that each agent's individual outcome depends on the other agent's random shock:

$$x_i = e_i + \varepsilon_i + \rho\varepsilon_{-i} \quad \text{for } i = 1, 2,$$

where ε_i , $i = 1, 2$, follows a distribution $N(0, \sigma^2)$, and ρ is the degree of correlation among the agents' outcomes. Then, in the optimal contract,

$$s_i^* = \frac{1 + \rho^2}{1 + \rho^2 + rv\sigma^2(1 - \rho^2)^2}, \quad z_i^* = -\frac{2\rho}{1 + \rho^2 + rv\sigma^2(1 - \rho^2)^2}.$$

Thus, for $\rho \neq 0$, there is relative performance evaluation, since the wage of agent i depends on the individual outcome of agent $-i$.

2.1.2 Model 2: Two efforts

Consider a situation similar to the one discussed in Model 1 but with $E = \{e^H, e^L\}$, that is, there are only two possible levels of effort: a high effort whose cost for the agent is $v(e^H)$ and a low effort with a cost of $v(e^L) < v(e^H)$.¹⁸ Implementing e^L is easy because the same fixed-wage contract that is optimal under symmetric information is also optimal under moral hazard (the Lagrange multiplier of the ICC is zero). On the other hand, implementing e^H requires taking into account the ICC that, in this case, can be written as

$$E(u(w(x)) | e^H) - v(e^H) \geq E(u(w(x)) | e^L) - v(e^L). \quad (\text{ICC2})$$

The solution to (P1) satisfies, for all $x \in X$,

$$\frac{1}{u'(w^*(x))} = \lambda + \mu \frac{(p(x | e^H) - p(x | e^L))}{p(x | e^H)}, \quad (4)$$

where $\frac{(p(x|e^H)-p(x|e^L))}{p(x|e^H)}$ is the likelihood ratio in the discrete case.

In this model, once the optimal payment scheme that allows the implementation of each effort has been obtained, finding the optimal effort is straightforward. It comes from the comparison of the principal's profits for each effort.

2.1.3 Model 3: Bounded feasible payments

As discussed above, there are important real-life situations where the principal cannot base the incentives on arbitrarily large bonuses (“carrots”) or fines (“sticks”). We consider now a situation that shares all the assumptions of Model 1 but where there are lower and upper bounds for the feasible payments. For each outcome $x \in X$, the salary $w(x)$ must satisfy

$$\underline{w}(x) \leq w(x) \leq \bar{w}(x) \quad (5)$$

where $\underline{w}(x)$ and $\bar{w}(x)$ are continuous, non-decreasing, and piecewise differentiable, with $\underline{w}(x) < \bar{w}(x)$ for all $x \in X$. Moreover, assume that the MLRP holds. Then, the analysis of Jewitt et al. (2008) ensures that the optimal payment scheme $(w^*(x))_{x \in X}$ to implement an effort e satisfies conditions similar to (2) “as much as possible”:

$$\frac{1}{u'(w^*(x))} = \begin{cases} \frac{1}{u'(\bar{w}(x))}, & \text{if } \frac{1}{u'(\bar{w}(x))} < \lambda + \mu \frac{p_e(x|e)}{p(x|e)} \\ \lambda + \mu \frac{p_e(x|e)}{p(x|e)}, & \text{if } \frac{1}{u'(\underline{w}(x))} \leq \lambda + \mu \frac{p_e(x|e)}{p(x|e)} \leq \frac{1}{u'(\bar{w}(x))} \\ \frac{1}{u'(\underline{w}(x))}, & \text{if } \lambda + \mu \frac{p_e(x|e)}{p(x|e)} < \frac{1}{u'(\underline{w}(x))} \end{cases}, \quad (6)$$

¹⁸See, e.g., Grossman and Hart (1983).

for some $\lambda \geq 0$ and $\mu \geq 0$. A particularly interesting example corresponds to a situation where there is no upper bound on salaries but there is a minimum wage \underline{w} (that is, the lower bound is independent of the outcome). This may be the case, for example, because of the agent's limited liability. In this case, the first line of (6) has no bite and the third line of (6) matters for low levels of the outcome, because the MLRP implies that the wage scheme is monotone. Thus, the optimal contract offers the minimum salary wage \underline{w} until some minimum outcome \hat{x} is reached and, from this level on, the contract follows a pattern similar to that without bounds.

2.2 Incentives vs. rents

We now assume that both the principal and the agent are risk neutral and that the sets X and E are continuous. Moreover, the payments to the agent are subject to lower and upper bounds.

Without limited liability, and because of the agent's risk neutrality, there is no benefit in insuring the agent and the solution of (P1) would be a franchise contract that would lead to the first-best. The franchise contract has the form $w(x) = w - k$, where k is the constant that makes (PC) binding. However, with limited liability, the principal is often forced to give the agent additional rents so that he has an incentive to provide a high effort.¹⁹ Thus, the optimal contract trades off incentives vs. rents.

2.2.1 Model 4: Limited liability

The participants are subject to limited liability so that, in the same spirit as in (5), the wage can neither be negative nor higher than the outcome, that is,

$$0 \leq w(x) \leq x \tag{7}$$

for all $x \in X$. Following the steps in Innes (1990),²⁰ we assume that the MLRP holds, $E\{x \mid e = 0\} = 0$, a profitable contract exists involving a

¹⁹When the agent has limited wealth, his level of effort may be constraint. Qu erou, Soubeyran, and Soubeyran (20015) study a situation where the principal may need to make an up-front transfer to the agent because the agent may not have enough resources to pay for the cost of the effort, when this cost is monetary.

²⁰Holmstr om (1979) and Lewis (1980) already noted the potential importance of limited liability constraints. However, Innes (1990) is the first paper to study the impact of liability limits on the qualitative properties of the optimal contract. Sappington (1983) and Demski, Sappington, and Spiller (1988) also bring in a limited liability constraint but they assume that the agent chooses the effort e after observing the state of nature. Other papers that study moral hazard problems with a minimum bound on payments under different assumptions are Park (1995), Kim (1997), Oyer (2000), Matthews (2001), and Jewitt, Kadan, and Swinkels (2008).

positive effort, and the total value of the relationship is strictly concave in e .

Given that higher effort increases the probability of higher outcomes, the contract should give the agent maximal payoffs in high outcomes. A particularly simple contract emerges under the additional monotonicity constraint that the principal's profit cannot be decreasing in the outcome, that is, $x - w(x)$ is non-decreasing in x .²¹ In this case, the optimal contract is a “debt contract” for the principal where she obtains $\min\{x, z\}$, for some $z > 0$. Thus, the optimal salary scheme is

$$w^*(x) = \max\{x - z, 0\}$$

for all $x \in X$. The value z corresponds to the one that makes (PC) binding if \underline{U} is high enough.²² If \underline{U} is very low then (PC) is not binding because (7) constrains the payoffs so much that the principal prefers to give the agent some extra rent to better provide incentives for effort. In all the cases, the effort implemented under moral hazard is lower than the first-best level e^{FB} .²³

Without the monotonicity constraint, the optimal contract takes the extreme “live-or-die” contract of the form

$$w^*(x) = \begin{cases} 0, & \text{if } x \leq z \\ x, & \text{if } x > z \end{cases}$$

for some value $z > 0$, whenever this contract leads to an effort level lower than e^{FB} . Otherwise, a contract proportional to the previous one (that is, a contract that gives sx for $x > z$) is a solution to the program (with z and s chosen in an appropriate way) and it implements the first best.

The PC is often not binding in situations where there is limited liability. In other words, when the limited liability constraint is binding the agent may obtain some rents, making the participation constraint slack. This is in contrast to the case without limited liability constraints

²¹The monotonicity constraint may be due to the possibility for the principal to “burn” or “hide” profits, or to the possibility for the agent to costly inflate the outcome if the payment does not satisfy the constraint.

²²Matthews (2001) shows that, under the same restrictions as Innes (1990), debt is still the optimal incentive contract if the agent is risk averse, renegotiation cannot be prevented, and the agent has all the bargaining power in the renegotiation game.

²³Poblete and Spulber (2012) extend the analysis of Innes (1990) by characterizing the optimal agency contract in more general environments using the state-space (or parametric) representation. They assume a technology $x = x(\theta, e)$, where θ is the state, a random variable with some density and distribution function. They do not assume the MLRP and introduce a “critical ratio” from which the form of the optimal contract easily follows. In particular, they provide a weaker condition than the MLRP under which the optimal contract is a debt contract.

where, at least when the agent's utility is additively separable, the agent never receives rents. Thus, the non-verifiability of the effort when there is limited liability and the agent is risk neutral may imply a cost either because the optimal contract leads to an effort lower than e^{FB} , because it gives the agent an informational rent, or both.²⁴ Example 2 illustrates the trade-offs in a simple model with two outcomes.

Example 2: Limited liability and two outcomes. In this example, we consider that only the agent is subject to limited liability, so the only additional constraint to program (P1) is $w(x) \geq 0$. Moreover, two outcomes are possible: success (a “good” outcome), in which case $x = x_G > 0$, and failure (a “bad” outcome), $x_B = 0$. The probability of success is $p(x_G | e) = e$ and the cost of effort is $v(e) = \frac{1}{2}ve^2$, with $v > 0$.

The agent's ICC implies that, under a contract $(w(x_B), w(x_G))$, he will select effort $e = \frac{w(x_G) - w(x_B)}{v}$. Once we take this constraint into account, together with the participation and limited liability constraints, the program that the principal solves is

$$\begin{aligned} \underset{(w(x_B), w(x_G))}{Max} \quad & \left\{ -w(x_B) + \frac{(w(x_G) - w(x_B))(x_G - (w(x_G) - w(x_B)))}{v} \right\} \\ \text{s.t.} \quad & w(x_B) + \frac{(w(x_G) - w(x_B))^2}{2v} \geq \underline{U} \end{aligned} \quad (8a)$$

$$w(x_B) \geq 0, \quad (8b)$$

which has a solution in which the principal makes no-negative profits if $\underline{U} \leq \frac{x_G^2}{2v}$.

In the optimal contract, the base salary is zero, $w^*(x_B) = 0$, because the limited liability constraint (8b) always binds. The participation constraint (8a) is also binding if the agent's reservation utility \underline{U} is intermediate ($\underline{U} \in \left[\frac{x_G^2}{8v}, \frac{x_G^2}{2v} \right]$). On the other hand, if \underline{U} is low ($\underline{U} < \frac{x_G^2}{8v}$) then the optimal contract gives a rent to the agent: providing incentives requires separating $w(x_G)$ from $w(x_B) = 0$ and the principal prefers to offer a salary $w(x_G)$ higher than the one necessary to satisfy the PC so that the agent chooses a higher effort. The optimal bonus is

$$w^*(x_G) = \begin{cases} \sqrt{2v\underline{U}}, & \text{if } \frac{x_G^2}{8v} \leq \underline{U} \leq \frac{x_G^2}{2v} \\ \frac{x_G}{2}, & \text{if } \underline{U} < \frac{x_G^2}{8v}. \end{cases} \quad (9)$$

²⁴This framework is also useful for studying more complex situations. See Fleckinger and Roux (2012) for a comprehensive review of the literature on performance comparison and competition in motivating agents in the framework where all participants are risk neutral, the agents are protected by limited liability, and they choose their effort non-cooperatively.

For intermediate outside utility, the effort increases (it gets closer to the first-best effort $e^{FB} = \frac{x_G}{2v}$) while the principal's profit decreases with \underline{U} : $e^* = \sqrt{\frac{2\underline{U}}{v}}$ and $\pi^* = x_G \sqrt{\frac{2\underline{U}}{v}} - 2\underline{U}$. When $\underline{U} < \frac{x_G^2}{8v}$ then effort and profits are constant: $e^* = \frac{x_G}{2v}$, the utility of the agent is $\frac{x_G^2}{8v} > \underline{U}$ and the principal profit is $\frac{x_G^2}{4v}$.

2.3 Incentives to a task vs. incentives to another task

The basic theory of moral hazard considers an agent supplying a one-dimensional effort that influences a one-dimensional output. However, relationships are often more complicated and the agent may be responsible for supplying a multi-dimensional effort or performing more than one task. Holmström and Milgrom (1991) study this extension using a model where the agent has CARA utility over wage and effort (as in Example 1) and either the tasks are related in the agent's cost of exerting them or their outcomes may be subject to correlated shocks. In this influential paper they discuss, among other issues, the trade-off between the incentives for different tasks in the extreme case where the outcome is easy to measure in one task while it is very difficult or impossible to measure (or to verify) in another task.²⁵

2.3.1 Model 5: Moral hazard with two tasks

Consider a risk-neutral principal who hires an agent with a CARA utility function to provide a vector (e_1, e_2) of efforts. The cost of the efforts for the agent are summarized in the cost function

$$v(e_1, e_2) = \frac{1}{2}v(e_1^2 + e_2^2) + \delta e_1 e_2,$$

with $|\delta| < v$. The output vector (x_1, x_2) depends on the agent's efforts and some random variable:

$$x_i = e_i + \varepsilon_i \text{ for } i = 1, 2.$$

The noise of the output function is assumed to follow a normal distribution $N(0, \Sigma)$, with $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$. Finally, the principal offers a payment scheme $w(x_1, x_2)$ to the agent, which is assumed to be linear:

$$w(x_1, x_2) = F + s_1 x_1 + s_2 x_2.$$

²⁵Holmström and Milgrom (1991) also consider limits on outside activities and how to allocate tasks between the agents. See also Feltham and Xie (1994).

Focussing in the interior solution, from the ICCs for the two efforts one can derive the agent's decision regarding (e_1, e_2) :

$$e_i(s_1, s_2) = \frac{s_i v - \delta s_j}{v^2 - \delta^2},$$

for $i = 1, 2$. Given the expression for the agent's decision, it is easy to check that if $\delta > 0$ then there is a substitution effect: the effort in one task decreases when incentives provided to the other task increase.

The solution of the principal's problem determines the optimal s_1^* (the term s_2^* of the compensation scheme is symmetric):²⁶

$$s_1^* = \frac{1 + r\delta(\sigma_{12} - \sigma_2^2) + rv(\sigma_2^2 - \sigma_{12})}{1 + r(2\delta\sigma_{12} + v\sigma_1^2 + v\sigma_2^2) + r^2(v^2 - \delta^2)(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)}.$$

Since there are many effects at work in the expression defining the optimal shares s_1^* and s_2^* , we present two extreme cases of the general situation to better understand the effects at work.

(i) If the two tasks are not related to each other in their cost structure, $\delta = 0$, but the random shocks are correlated, $\sigma_{12} \neq 0$, then the incentive mechanism is

$$s_1^* = \frac{1 + r\sigma_2^2 v - rv\sigma_{12}}{(1 + rv\sigma_1^2)(1 + rv\sigma_2^2) - r^2 v^2 \sigma_{12}^2},$$

which depends on both tasks' variance shocks and their covariance. For $\sigma_1^2 = \sigma_2^2$, s_1^* is decreasing in σ_{12} : the higher the covariance of the two tasks, the lower the weight of each outcome on the payment scheme. The reason is that with a high covariance, the outcomes of the two tasks move together and the incentives for the effort on one task derive from the payments on the output of both tasks.

(ii) If the two tasks are related to each other in the cost structure ($\delta \neq 0$) but there is no correlation between the random shocks ($\sigma_{12} = 0$), then in the optimal contract,

$$s_1^* = \frac{1 + r(v - \delta)\sigma_2^2}{1 + rv(\sigma_1^2 + \sigma_2^2) + r^2(v^2 - \delta^2)\sigma_1^2\sigma_2^2}.$$

²⁶If the two tasks are independent ($\delta = 0$) and there is no correlation of the random shocks ($\sigma_{12} = 0$), then the incentives are separable,

$$s_i = \frac{1}{1 + rv\sigma_i^2} \text{ for } i = 1, 2,$$

and the payment scheme is the same as the one obtained in the single-task moral hazard problem.

As expected, s_1^* is decreasing on σ_1^2 . Moreover, it is also decreasing on σ_2^2 . Therefore, when efforts in the two tasks are substitutes then the optimal shares s_1^* and s_2^* are complementary. If $\sigma_1^2 = \sigma_2^2 = 0$ then $s_1^* = s_2^* = 1$ and the agent is the residual claimant for both tasks. But if $\sigma_1^2 = 0$ and $\sigma_2^2 > 0$, then

$$s_1^* = \frac{1 + r(v - \delta)\sigma_2^2}{1 + rv\sigma_2^2} \quad \text{and} \quad s_2^* = \frac{1}{1 + rv\sigma_2^2},$$

where $s_1^* < 1$ if and only if $\delta > 0$. Therefore, even if the outcome of task 1 is a perfect measure of the effort in this task, the principal decreases the incentives associated with the outcome of task 1 when tasks are substitutable not to harm the effort supplied in task 2. On the other hand, if the tasks are complementary then the optimal s_1^* is higher than 1. In contrast, the parameter s_2^* does not depend on δ and it is the same as in the traditional moral hazard.

Finally, if task 1 can be measured and task 2 cannot (which can be represented by $\sigma_2^2 = \infty$) but the agent has some intrinsic motivation for this task, then the optimal scheme is based on

$$s_1^* = \frac{1 - \delta/v}{1 + r\sigma_1^2(v - \delta^2/v)}, \quad \text{and} \quad s_2^* = 0.$$

Here, if the tasks are substitutes ($\delta > 0$) putting effort into one task increases the marginal cost of the other.²⁷ Therefore, the principal gives the agent a lower incentive to exert effort in task 1 even when it is easily measurable because she does not want to discourage the agent's effort in task 2, which cannot be directly motivated. The higher the cross-effort effect is (that is, the more substitutable the effort levels are), the lower the optimal s_1^* .²⁸ In contrast, if the tasks are complements ($\delta < 0$) the opposite happens, and the agent will be highly motivated to perform task 1 to encourage effort in the unmeasurable task.

²⁷In a situation where the principal cares specially about the non-measurable task, Holmström and Milgrom (1991) show that it is best not to provide any incentive to the task with measurable output.

²⁸Dam and Ruiz-Pérez (2012) study a model where a risk-neutral agent subject to limited liability exerts effort in two tasks. When the efforts in the two tasks are independent of each other, the optimal contract is a debt contract. However, if the tasks are substitutes, then revenue sharing emerges as an optimal agreement. Ghatak and Pandey (2000) also show the optimality of sharing contracts when the risk-neutral agent has to supply an effort and to choose the riskiness of the production technique.

2.4 Incentives to the agent vs. incentives to the principal

In many situations, it is not only the agent who must submit an effort or take a decision, but the principal's contribution is crucial for the relationship and, just like the agent's, it is not verifiable. In these situations, the stronger the incentives to the agent (that is, the more the salary depends on the outcome) the weaker the incentives to the principal (because the less the principal's benefit depends on the outcome).

2.4.1 Model 6: Double-sided moral hazard with risk neutrality

When both the principal and the agent are risk neutral, program ($P1$) is still interesting if we consider a double-sided moral hazard problem. In this environment, the agent chooses e and, simultaneously, the principal decides on her effort $a \in A$, at a cost of $c(a)$, with $c(\cdot)$ increasing and convex. Following the analysis of Bhattacharyya and Lafontaine (1995), assume that the outcome of the relationship depends on both e and a according to

$$x = h(e, a) + \varepsilon$$

where the function $h(\cdot, \cdot)$ is increasing and concave in both arguments, the cross-partial derivative is positive, $h(0, a) = h(e, 0) = 0$, and ε is a random term with mean zero and variance σ^2 .

The new maximization problem ($P1'$) takes into account that the outcome depends on both efforts and that there is also an ICC for the principal:

$$\begin{aligned} & \underset{\{(w(x))_{x \in X}, e, a\}}{\text{Max}} \quad \{E(x - w(x) \mid e, a)\} \\ & \text{s.t.} \quad E(u(w(x)) \mid e, a) - v(e) \geq \underline{U} \\ & e \in \arg \max_{\hat{e} \in E} \{E(u(w(x)) \mid \hat{e}, a) - v(\hat{e})\} \\ & a \in \arg \max_{\hat{a} \in E} \{E(x - w(x) \mid e, \hat{a}) - c(\hat{a})\}. \end{aligned}$$

Bhattacharyya and Lafontaine (1995) show that, without loss of generality, the optimal sharing rule can be represented by a linear contract

$$w(x) = F + sx$$

for some sharing $s \in (0, 1)$. A linear contract is not the unique way to achieve the optimal solution for ($P1'$) but there is always an optimal solution that is linear.²⁹ In terms of incentives, the crucial element of

²⁹Romano (1994) obtains a similar result.

any optimal contract is its slope at the optimum. By choosing a linear rule with the slope of any optimal sharing rule (and adjusting the fixed fee), exactly the same incentives and total payments can be achieved as with the initial rule.

The optimal sharing s^* makes a trade-off between providing incentives to the agent and providing incentives to the principal. Once s^* is determined, the fixed part of the contract F^* is easily obtained because the agent's participation constraint is binding. While it is not possible to obtain simple, closed-form expressions for the optimal sharing rule and the levels of the optimal efforts on the part of the two parties in general, a very simple example allows us to grasp most of the intuitions.

Example 3: Double moral hazard with linear outcome. Following Ghatak and Karaivanov (2014), consider³⁰

$$h(e, a) = \alpha\theta_A\theta_P + \theta_A e + \theta_P a$$

where $\theta_A \geq 1$ and $\theta_P \geq 1$ represent the agent's and principal's ability to perform his or her task, respectively, and α is a parameter capturing the extent of the types' complementarity in production. Moreover, $v(e) = \frac{1}{2}e^2$ and $c(a) = \frac{1}{2}a^2$.³¹ Then, the optimal sharing s^* derived from (P1') is

$$s^* = \frac{\theta_A^2}{\theta_A^2 + \theta_P^2},$$

which gives more weight to the relatively more important participant: $s^* > 1/2$ (and the share that goes to the principal satisfies $(1 - s^*) < 1/2$) if and only if θ_A is larger than θ_P , and s^* is increasing in θ_A and decreasing in θ_P . Given the optimal contract, the efforts under double-sided moral hazard are

$$e^* = \theta_A s^* = \frac{\theta_A^3}{\theta_A^2 + \theta_P^2} \text{ and } a^* = \theta_P (1 - s^*) = \frac{\theta_P^3}{\theta_A^2 + \theta_P^2}. \quad (11)$$

Both efforts are lower than the corresponding first-best efforts that in this model are $e^{FB} = \theta_A$ and $a^{FB} = \theta_P$.³² However, the optimal sharing

³⁰We note that Example 3 does not satisfy all the assumptions of Model 6 because $h(0, a) > 0$, $h(e, 0) > 0$, and $\frac{\partial^2 h}{\partial a \partial e}(e, a) > 0$.

³¹The objective in Ghatak and Karaivanov (2014) was to study the contractual choice in agriculture, taking into account that two different tasks are necessary, following a model in the spirit of the classic Eswaran and Kotwal (1985) model.

³²The first-best cannot be achieved (even though both partners are risk neutral) because there is no "budget-breaker," (or residual claimant) that is, it is not possible to propose a contract where the total remuneration of the principal and agent is higher than the outcome sometimes and lower other times. This is similar to Holmström (1982) who shows that joint production cannot lead to efficiency when all the income is distributed amongst the agents, i.e., if the budget constraint always binds.

rule solves the trade-off with respect to the incentives for the principal and the agent by inducing a smaller distortion to the most important participant, that is,

$$e^{FB} - e^* > a^{FB} - a^* \iff \theta_A < \theta_P.$$

3 Behavioral approach

The classical moral hazard problem assumes full rationality and standard preferences. Recent behavioral research on moral hazard, encouraged by experimental results, attempts to understand the implications of agents' non-fully rational and non-purely selfish preferences on the shape of the incentive contracts. What follows is not a review of the behavioral literature but provides some examples of how departures from the classical model affect the conclusions obtained in section 2. We briefly present the consequences of considering other-regarding preferences, intrinsic motivation, loss aversion, and overconfidence.³³ In each of the following behavioral approaches, a new effect appears. For example, in the inequality aversion extension, the incentive contract is the issue of the trade-off between insurance, incentives, and fairness. Similarly, in the overconfidence extension the optimal payoff scheme makes a trade-off between optimal risk sharing, incentives, and gambling.

3.1 Other-regarding preferences

There is evidence pointing to the existence of people or institutions who are not just concerned about their own payment scheme but also care about other participants' well-being (see Rabin, 2002, Englmaier, 2005, and Sobel, 2005, for reviews of the literature). To illustrate the meaning of this type of preferences in our framework, consider the interaction between a principal and an agent in which case the payment of the participants are described by the vector $(x - w(x), w(x))_{x \in X}$. If the agent (a similar argument can be done for the principal) is only concerned about the second element of this vector then we are in the classical framework. In contrast, when he cares about the whole vector of payoffs (or of utilities) then we say that the agent has other-regarding preferences.

There are several ways in which the agent can experience other-regarding preferences. A first possibility is that he has a utility function similar to a *weighted social preference* (Segal and Sobel, 2007), such as

$$U(w(x), x - w(x), e) = u(w(x) + \delta(x - w(x))) - v(e).$$

In this case, the agent is altruistic if δ is positive, while he is spiteful if δ

³³See Köszegi (2014) for the latest survey of behavioral contract design.

is negative.³⁴ The sign of δ also determines whether the contract is more or less costly for the principal, as compared to the classical framework where $\delta = 0$.

Dur and Glazer (2008) study an environment where the agent is other-regarding because he envies the principal. They show that envy tightens the agent's participation constraint and the optimal contract calls for higher wages and lower effort requirements.

Inequality aversion is a second form of other-regarding preferences (see Fehr and Schmidt, 1999, and Bolton and Ockenfels, 2000). Focusing again on the agent's behavior and denoting the difference between the principal and the agent's payoffs as $d(x) \equiv x - 2w(x)$, an easy example of a utility function that represents an inequality-averse agent is

$$u(w(x), x-w(x), e) = w(x) - \delta (\max \{d(x), 0\} + \gamma \max \{-d(x), 0\}) - v(e). \quad (12)$$

The parameter δ , with $\delta \geq 0$, measures the extent of the agent's concern about the difference in earnings. If $\delta > 0$, the agent is inequality averse and if $\gamma \in [0, 1)$ the agent suffers more from inequality when he is behind than when he is ahead.

Providing incentives to an inequality-averse agent is often more costly than to a classical agent. The intuition is that as the project is more profitable, more inequality is created and it is more expensive to satisfy the ICC. To illustrate this effect, we follow Itoh (2004) and consider a simple model where both participants are risk neutral, the agent is protected by limited liability, there are two efforts, $E = \{e^H, e^L\}$, and two outcomes, $X = \{x_G, x_B = 0\}$. The probabilities of success are $p(x_G | e^H) = p^H$ and $p(x_G | e^L) = p^L$, and the cost of effort is $v(e^H) = v > 0 = v(e^L)$. Assume that the principal wants the agent to exert effort e^H . Then, taking into account that the limited liability binds ($w_B = 0$), the ICC takes the form

$$(p^H - p^L) [w_G - \delta (\max \{d(x_G), 0\} + \gamma \max \{-d(x_G), 0\})] \geq v.$$

The incentives (the left-hand side of the ICC) are decreasing in δ ; hence,

³⁴We concentrate the discussion on the cases where the agent is other-regarding. One can similarly define an other-regarding preference principal as

$$B(x - w(x), w(x)) = x - w(x) + \delta_P (w(x) - v(e)).$$

Also notice that the coefficient δ_i , for $i = P, A$, could be a function of the distance between $x - w(x)$ and $w(x) - v(e)$, in such a way that, for example, i may care more about the difference in earnings if s/he is the one getting less than if s/he is the one getting more as in Fehr and Schmidt (1999) and Bolton and Ockensfels (2000) discussed below.

the principal's profits are decreasing in δ and she is in general worse off when hiring an inequality averse rather than a classical agent.³⁵

In a model where the inequity aversion is convex in the difference in the payoffs, Englmaier and Wambach (2010) find a tendency toward linear sharing rules as the agent's concern for inequity become more important, in line with other findings that the more complex the situation is the simpler the optimal incentive scheme tends to be.³⁶ Interestingly, given that the contract now has to balance among three objectives (risk sharing, incentives, and inequality concerns) Englmaier and Wambach (2010) also find that the sufficient statistics result is violated because optimal contracts may be overdetermined or incomplete. To understand the intuition, consider a situation with two sources of information: the outcome related to the agent's effort and another variable. First, if this other information is not related to the agent's effort but just to the principal's profit, then the second measure will be included in the contract (which will thus include non-informative performance measures) because the set of variables used in the payment scheme no longer serve only as a signal of the agent's effort but also deal with the agent's concern about an equitable treatment. Second, if the second variable (second-order stochastically) dominates the outcome, it may be optimal to concentrate the incentives on the outcome (neglecting informative performance measures) because this is the variable the agent is interested in when he is inequality concerned.

A third form of other-regarding preferences is *reciprocal behavior*, where the agent may take into account the behavior of the principal, in such a way that the agent will reciprocate and take a decision that also benefits the principal (supplying higher effort) if she takes a decision that benefits the agent (paying higher wages) (Rabin, 1993).³⁷ If an agent follows the previous behavior, monetary incentive and reciprocal motivation are substitute and the agent's PC may not be binding at the optimal contract. To illustrate how these two incentive tools are combined in the optimal contract, we present a simplified version of Englmaier and Leider's (2012) model, where the agent has reciprocal references.

We consider a risk-neutral principal and a risk-averse agent. There

³⁵In this model, the effect is particularly straightforward because there is no inequality if the outcome is x_B .

³⁶The convexity of the inequality term $\delta(d(x))$ implies an aversion to loteries over different levels of inequity.

³⁷Akerlof (1982) explained the labor relation as a gift exchange where agents respond to a generous wage scheme offered by the principal by exerting more than minimal effort. See, e.g., Dufwenberg and Kirchsteiger (2005) and Falk and Fischbacher (2006) on reciprocity in sequential games.

are two possible outcomes: success (x_G) and failure (x_B), with $x_G > x_B$. The agent chooses between two efforts: $e \in \{e^L, e^H\}$, with $p(x_G | e^H) = p^H > p^L = p(x_G | e^L)$. A contract takes the form (w_G, w_B, e^H) , where we assume that the principal is interested in obtaining effort e^H . The effort e^H is not enforceable but is “the job description.” The agent is reciprocal, in the sense that his expected utility under (w_G, w_B, e^H) if he provides effort e is

$$p(x_G | e)u(w_G) + (1 - p(x_G | e))u(w_B) - ve + \delta R(e^H)r(e)$$

where $R(e^H) \equiv p(x_G | e^H)u(w_G) + (1 - p(x_G | e^H))u(w_B) - ve^H - \underline{U}$ is the agent’s expected rent under the job description, $r(e) \equiv p(x_G | e)x_G + (1 - p(x_G | e))x_B$ is the principal’s expected revenue if the agent chooses e , and δ is the intensity of the agent’s reciprocal behavior. With this utility function, the agent experiences reciprocal motivation only if the contract gives him an expected utility higher than his reservation utility (that is, if the PC does not bind). The agent’s ICC is

$$(p^H - p^L)(u(w_G) - u(w_B)) + \delta R(e^H)(p^H - p^L)(x_G - x_B) \geq v(e^H - e^L).$$

It is worthwhile noticing that a high enough fixed payment \tilde{w} can implement the effort e^H . The condition is that \tilde{w} satisfies

$$\delta [u(\tilde{w}) - ve^H - \underline{U}] (p^H - p^L)(x_G - x_B) \geq v(e^H - e^L).$$

The ICC is always binding at the optimal contract, but the agent’s PC may be binding or not. The optimal contract is a standard one for low values of δ , providing no rents to the risk-averse agent, whereas it is a reciprocity contract that gives the agent a utility larger than his reservation utility for large values of δ .³⁸ When rents are provided to the agent, the FOC of the principal’s program with respect to $w(x)$ can be written as

$$\frac{1}{u'(w^*(x))} = \mu \left(\frac{p(x | e^H) - p(x | e^L)}{p(x | e^H)} + \delta (p^H - p^L)(x_G - x_B) \right)$$

which implies that monetary and reciprocity motivations are substitutable.³⁹

³⁸In fact, for very large values of δ the first-best solution can be arbitrarily closely approximated with a contract that gives the agent an infinitesimal rent (Englmaier and Leider, 2012).

³⁹Behavioral models have been very useful in analyzing multiagent situations. It is interesting to note that since inequality-averse agents care about other agents’ remuneration, to reduce inequity among agents, their payments will tend to depend on

3.2 Extrinsic and intrinsic motivation

The classical moral hazard model is based on designing incentives to provide extrinsic motivation to the agent. The so-called extrinsic motivation is the one that is derived from the monetary incentive scheme. In contrast, an agent's *intrinsic motivation* comes from the utility obtained from achieving some goal set by himself, the society, the principal, or from working for a particular type of principal, such as one who honors some community (environmental or another form of social) standards. The simplest agent's utility function that represents an agent with both intrinsic and extrinsic motivation is

$$U(w, e, s) = w + Im - v(e),$$

where $I \geq 0$ is the intrinsic motivation ($I = 0$ in the classical model), and m is the source of this motivation. We briefly discuss the consequences of some sources of intrinsic motivation.⁴⁰

Intrinsic motivation may come from the agent's perception of the world, which may depend on the contract the principal offers. The underlying idea in this approach is the following. The agent expects to offer a predetermined effort and to receive a fixed-fee (first-best) payment. However, if he is offered an incentive contract instead, his perception of the relationship changes and he becomes aware of the possibility of shirking.⁴¹ Thus, providing extrinsic incentives for the agent can be counterproductive because it may crowd out his intrinsic motivation, leading to lower effort levels and lower profits for the principal (Kreps, 1997, and James, 2005). For example, Auster (2013) and von Thadden and Zhao (2012, 2014) study a situation where agents are *unaware* of the full effort problem and they make a default effort when offered an (incomplete) full insurance payoff, while they become aware of the effort problem and behave strategically if they are offered the optimal moral hazard contract (see also James, 2005). Similarly, Bénabou and Tirole

others agents' performance, even if they are statistically and technologically independent. For example, Englmaier and Wambach (2010), Goel and Thakor (2006), and Bartling (2011) show that inequity aversion or envy among agents may render team incentives optimal. Itoh (2004) finds that inequity aversion when agents are subject to limited liability may allow agency costs to be reduced. Rey-Biel (2003) finds that the principal can always exploit inequity aversion to extract more rents from her agents. Demougin and Fluet (2006) compare group and individual bonus schemes for behindness-averse agents and derive conditions under which either scheme implements a given effort level at least costs.

⁴⁰Some cases of other-regarding behavior, such as that of social preferences (for $m = x - w(x)$) or the reciprocal motivation model presented in the previous subsection, can also be understood as models of intrinsic motivation.

⁴¹This can be seen as a form of bounded rationality.

(2003) consider a principal agent model where the principal is better-informed than the agent about the agent’s characteristic, and show that performance incentives lead to an increase of the agent’s effort in the short run but they are negative reinforcements in the long run.⁴²

Another source of intrinsic motivation may be due to some characteristic (or to a verifiable decision) of the principal (Murdock, 2002, and Besley and Ghatak, 2005). To better explain this approach, suppose that the two participants are risk neutral, the agent’s payoff is constrained to be non-negative (that is, there is limited liability), and there are two possible outcomes: success (x_G) and failure (x_B), with $x_G > x_B = 0$. The set of possible efforts is $E = [0, 1]$ and $p(x_G | e) = e$. A principal has a certain type, or a mission orientation, and the agent’s preferences can be aligned with a particular mission or with none of them. The public type of the principal is τ , with $\tau \in \{0, M\}$. A type-0 principal has no mission and is the traditional profit-maximizing partner whereas type- M principals have a mission. The agent is mission-oriented. The source of his intrinsic motivation is that he cares about the success of his job when he works for a principal with a mission.⁴³ His utility function can be represented by⁴⁴

$$U(w, e, x) = w + I(x, \tau) - \frac{1}{2}e^2, \quad (13)$$

where $I(x, \tau)$ depends on the outcome x and the type of the principal τ . In case of failure $I(x_B, \tau) = 0$, for all τ . In case of success, $I(x_G, \tau = M) > I(x_G, \tau = 0) = 0$, that is, when the mission-oriented agent works for a type-0 principal, he behaves as a traditional agent. $I(x_G, \tau = M)$ is the intrinsic utility that the agent derives from the success of his work for a type- M principal.

This model is an extension of Example 2 (with $v = 1$) and the expression for the optimal bonus in this environment is also very similar

⁴²The intuition is that when the principal pays a bonus to induce low ability agents’ to work (the principal increases the agent’s extrinsic motivation), then the agent perceives the bonus as a bad signal about his own ability (she reduces the agent’s intrinsic motivation). Kirkegaard (2015) studies a model where the agent works for the principal and simultaneously pursues private benefits. He shows that the optimal contract may offer high rewards but flat incentives to lessen the agent’s incentive to pursue private benefits, his intrinsic motivation.

⁴³A type-0 agent would be the traditional agent who does not care about the type of the principal and we would be back to the traditional moral hazard problem, whatever the type (mission) of the principal is.

⁴⁴The model can be extended by allowing the principal to choose the “mission,” taking into account the effect of the choice in the agent’s incentives (Besley and Ghatak, 2005).

to (9):

$$w^*(x_G) = \begin{cases} \sqrt{2\underline{U}} - I(x_G, \tau), & \text{if } \frac{(x_G + I(x_G, \tau))^2}{8} \leq \underline{U} \leq \frac{(x_G + I(x_G, \tau))^2}{2} \\ \max \left\{ 0, \frac{x_G - I(x_G, \tau)}{2} \right\}, & \text{if } \underline{U} < \frac{(x_G + I(x_G, \tau))^2}{8} \end{cases}. \quad (14)$$

The effort implemented by the agent is $e^* = \sqrt{2\underline{U}}$ for the intermediate region of \underline{U} whereas it is $e^* = \max \left\{ I(x_G, \tau), \frac{x_G + I(x_G, \tau)}{2} \right\}$ when \underline{U} is low. Thus, a higher intrinsic motivation $I(x_G, \tau)$ results in a higher (or equal) effort by the agent at a lower cost in terms of bonus.⁴⁵

3.3 Loss aversion

There is also evidence that some individuals do not evaluate payoff in absolute terms but in comparison with some reference point (Kahneman and Tversky, 1984). *Loss aversion* is the reference-dependent preference that has been most studied, a type of preference that may explain why contracts framed as bonuses are much more prevalent than contracts framed as penalties (see, e.g., Aron and Olivella, 1994). The idea is that, evaluated at the reference point, the marginal utility of a loss is larger than the marginal utility of a gain, so that the agent's utility function has a kink at this reference point. We present the basic loss aversion model by De Meza and Webb (2007) with an exogenous reference wage, which fits within the structure of Model 2, where $e \in \{e^L, e^H\}$. Consider that the principal is risk neutral and the agent is risk averse with loss aversion with respect to a reference wage w^R :

$$U(w, e) = u(w) - 1_{w < w^R} l(u(w^R) - u(w)) - v(e),$$

where the index $1_{w < w^R} = 1$ if $w < w^R$ and $1_{w < w^R} = 0$ if $w \geq w^R$; and $l > 0$ is the loss the agent suffers when the wage is lower than the reference wage w^R .

Assume that MLRP is satisfied and that the principal aims to implement the high effort. Then, from the first-order condition of the

⁴⁵Guo (forthcoming) also analyzes a model where the agent has extrinsic motivation, in addition to the monetary incentives, associated with a principal's decision. In her paper, the extra motivation of the agent (an employee) comes when his principal (a manager) invests in a non-contractible employee-friendly relationship. In Guo (forthcoming), the utility function of the agent has the form

$$U(w, e, m) = u(w) - v(e, I(m))$$

because the agent's extrinsic motivation influences his effort disutility. The investment in the relationship can also benefit the principal because, for example, the employee may support the manager if she faces a replacement threat.

principal's problem, we obtain

$$\frac{1}{u'(w^*(x))} = (1 + \nu(x)l) \left(\lambda + \mu \frac{p(x | e^H) - p(x | e^L)}{p(x | e^H)} \right). \quad (15)$$

where $\nu(x) \in [0, 1]$ is an instrument to handle the kink in the agent's utility. If $w < w^R$, loss aversion applies and $\nu(x) = 1$; if $w > w^R$, loss aversion does not apply and $\nu(x) = 0$; and if $w = w^R$, then the ICC holds with equality for $\nu(x) \in [0, 1]$.

From (15), De Meza and Webb (2007) derive that loss aversion does not affect the condition if w^R is very low ($\nu(x) = 0$ for all x) and, as in Model 2, $w^*(x)$ is strictly increasing in outcome. Similarly, if w^R is very high ($\nu(x) = 1$ for all x) then $w^*(x)$ is also strictly increasing in outcome. However, in the remaining cases loss aversion affects the optimal payment scheme, and there are zones (for the lower, the intermediary, or the higher outcomes) where the agent receives a flat wage equal to w^R .⁴⁶

In the presence of loss aversion or reference-dependent preferences, the principal designs incentives by taking into account both the induced risk sharing and the agent's loss aversion. The wage scheme will be a function of the outcome at least for certain outcomes, but it tends to have a significant number of outcomes where the payment is flat.⁴⁷

3.4 Overconfidence

Contracts are based on the principal and agent's beliefs (correct or incorrect) and, in the standard moral hazard model, it is customary to assume that both participants share the same beliefs about the uncertain elements of the relationship. However, we may think of situations where beliefs are different but each knows the view of the other (they "agree to disagree"). This may correspond to environments where the agent is "optimistic" and "overconfident" or he has different beliefs to the principal concerning his abilities (he can have a positive or negative self-image as compared to what the principal thinks). Santos-Pinto (2008) and De la Rosa (2011) consider a moral hazard model when the principal and agent have (public) asymmetric beliefs.

⁴⁶They also show that if, in addition, there is a limited liability constraint $w \geq \underline{w}$ (with $\underline{w} < w^R$), then it is optimal not to have payments in the interval (\underline{w}, w^R) and $w^*(x)$ is discontinuous.

⁴⁷Kőszegi and Rabin (2006) and (2007) show that the optimal payment scheme often has two wages (and incentives are based on a bonus). De Meza and Webb (2007) find that when the reference wage is the median wage, the incentives are based on performances over the median. When the reference point is endogenous, Herweg, Müller, and Weinschenk (2010) show that the rational expectation about the wage is the expected wage.

We present the basic elements and results within a structure close to Model 1. Consider that the principal is risk neutral and the agent is risk averse, with utility function $U(w, e) = u(w) - ve$, and that there are two possible outcomes: success (x_G) and failure (x_B), with $x_G > x_B$. We denote $p_P(x | e)$ and $p_A(x | e)$ the principal's and the agent's beliefs for outcome $x \in \{x_G, x_B\}$ for a given effort $e \in E$. The beliefs are asymmetric if $p_A(x | e) \neq p_P(x | e)$ for at least some $e \in E$, and we can say that the agent exhibits a positive self-image of own ability (or he is optimistic) if $p_A(x_G | e) > p_P(x_G | e)$, for all $e \in E$.

The existence of different beliefs affects the contract even under symmetric information. In this case, the wage scheme to implement an effort e satisfies

$$\frac{1}{u'(w^*(x))} = \lambda \frac{p_A(x | e)}{p_P(x | e)}.$$

Hence, full insurance (which results when $p_A(x | e) = p_P(x | e)$) no longer holds. Since the principal and agent have different views of the uncertain situation they are involved in, they can agree on a side-bet in such a way that both think they can gain. In the first-best, an optimistic agent will be paid more in the case of success (because the principal thinks this bonus will not be paid that often) and a pessimist will be paid more in the case of failure.⁴⁸ In addition, whether the agent is right or wrong in his beliefs, with the contract he will obtain his reservation utility according to his subjective beliefs (PC always binds). From her perspective, the cost to the principal of implementing high efforts is lower than in the standard model and it decreases with the agent's optimism and overconfidence.⁴⁹

After the analysis of the contracts under symmetric information, it is easy to see that under moral hazard it can be the case that it is less expensive for the principal to implement the high rather than the low effort.⁵⁰ If the agent is optimistic or overconfident enough, the first-best risk-sharing incentive scheme may induce the agent to exert high effort under moral hazard. In general, as shown by Santos-Pinto (2008), to induce the agent to work the asymmetry of beliefs can be either favorable

⁴⁸As Santos-Pinto (2008) points out, if an agent is risk neutral and has mistaken beliefs, the principal's problem does not have a solution because the principal can always increase her profits by raising the stakes of the side-bet. This implies that when the agent is risk neutral but is protected by limited liability then, in the optimal contract, the limited liability constraint is binding.

⁴⁹It is usual to assume that the agent is the one mistaken about the real technical conditions of the production process, but it is also possible that the opposite is true.

⁵⁰Another classical result in moral hazard with symmetric beliefs is that for the lowest effort the optimal contract under moral hazard and under symmetric information coincide. This result may not hold under overconfidence.

or unfavorable, depending on whether the agent is overconfident or the opposite. De la Rosa (2011) and Gervais et al. (2011) highlight that the reason for the asymmetries also matters. Incentive contracts are sensitive to the kind and level of overconfidence, not only to the presence of overconfidence per se. For example, in De la Rosa (2011) beliefs take the functional forms

$$p_P(x_G | e) = q_P + \theta_P e, \text{ and } p_A(x_G | e) = q_A + \theta_A e,$$

with $q_I > 0$, $\theta_I > 0$ and $q_I + \theta_I < 1$ for $I = P, A$. Then, assuming $e \in \{e^L = 0, e^H = 1\}$, if $q_A > q_P$ the agent is optimistic, if $\theta_A > \theta_P$ the agent is overconfident, and he is overconfident overall if $q_A > q_P$ and $q_A + \theta_A > q_P + \theta_P$. If the high effort is implemented, the principal's expected profit increases in both the agent's level of optimism and overconfidence. But if the low effort is implemented, the principal's expected profit increases in the agent's level of optimism or pessimism, for an optimistic or a pessimistic agent, respectively, and it decreases in the agent's overconfidence if the agent is significantly optimistic.

4 Principal-agent markets

The models that we have discussed above, and almost all the papers that study settings involving a moral hazard problem, take the identity and the characteristics of the participants in the relationship as given. They consider an isolated principal-agent situation (or an isolated relationship among several principals and/or several agents) and analyze the optimal contract (contracts) in this relationship. The principal assumes all the bargaining power, and the agent is ready to accept a contract as long as it guarantees him his exogenously given reservation utility.

The previous description is a good fit for situations where the participants in a contract cannot be easily replaced, as is the case for the relationship between a regulator and a firm. However, most often, a principal can look for alternative agents and an agent can look for alternative principals. When several principals and several agents exist in this "market," in addition to the question about the optimal contracts, we can address the endogenous determination of the identity of the pairs that meet (i.e., the matching between principals and agents). In particular, we can study whether, at equilibrium, there is *positive assortative matching* (PAM) or *negative assortative matching* (NAM). A PAM between principals and agents with respect to, say, ability (or any other characteristic, such as risk, type, etc.) exists if the partner A of a principal P with a higher ability than another principal P' has a higher or equal level of ability than the partner A' of the principal P' . A negative assortative matching is defined in a similar manner.

Furthermore, the alternative relationships that could be formed in the market are crucial to understanding the endogenous level of payoffs that each principal and agent obtain and some of the properties of the contract.

The theory of “two-sided matching models” provides the tool to study markets where heterogeneous players from one side (principals) meet with heterogeneous players from the other side (agents).⁵¹ The equilibrium of the market determines the identity of the partners that actually sign contracts (that is, the “matching”) together with the profits that they obtain and the characteristics of the contracts.⁵² Equilibrium outcomes satisfy two useful properties. First, equilibrium contracts are always Pareto optimal; hence, we can use what we have learned from the analysis of isolated relationships. Second, if utility is transferable (that is, it is possible to transfer one unit of utility from the principal to the agent) then any equilibrium matching is efficient in the sense that it maximizes “total surplus”: the sum of all the profits in the market cannot be increased by reassigning principals and agents.

We now discuss some of the new lessons from matching models with contracts.

4.1 The relationship between risk and incentives

A quite robust prediction of the moral hazard literature is the negative relationship between risk and performance pay (e.g., Holmström and Milgrom, 1987): the higher the risk of the project the agent is working on, the lower the incentives included in the contract. When the agent has CARA risk preferences, we have seen in (3) that the variable part of the contract s^* is decreasing in the risk of the relationship, represented

⁵¹The book by Roth and Sotomayor (1990) made the theory of two-sided matching models popular and accessible. Gale and Shapley (1962) started it by studying “the marriage market,” where each participant (in their case, a man or a woman) is only concerned about the characteristics of the members of the other side of the market (women or men, respectively). Shapley and Shubik (1972) broaden the set of applications of this theory by considering, in “the assignment model” that the utility derived from a relationship not only depends on the characteristics of the partner but also on money, which can be exchanged among partners as part of the agreement.

The assignment model can be easily extended to situations where not only money but contracts are endogeneously decided simultaneously with the matching, as long as utility is transferable one-to-one between principals and agents. In other cases, the analysis is more complex but the contribution by Kaneko (1982) allows us to also use this tool.

⁵²Any competitive equilibrium is also a stable outcome and vice-versa, where stability means individual rationality together with the property that no principal-agent pair can be better off by leaving their current partners and signing a new contract among them.

by the variance σ^2 . The CARA assumption also implies that utility is transferable because the principal can give or take away utility directly through the fixed part of the contract F .

To study whether this conclusion also holds when principals and agents interact in a market, consider that there exists a set of principals who are heterogeneous in the risk (variance) of their production process: each principal is associated with the variance of her project σ^2 , with $\sigma^2 \in [\sigma_L^2, \sigma_H^2]$. There is also a set of agents, heterogeneous in their risk aversion attitude: each agent is identified by his degree of risk aversion r , with $r \in [r_L, r_H]$. Both populations have the same mass.⁵³

Serfes (2005) analyzes how the degree of risk aversion of an agent relates to the risk of the project he is involved in at equilibrium. He provides the answer for two interesting cases: (a) if $\sigma_L^2 r_L \geq 1/v$ (that is, the risk and/or the degree of the agent's risk aversion are always large) then there is PAM: low-risk-averse agents are matched with low-risk principals (projects) and vice-versa; and (b) if $\sigma_H^2 r_H \leq 1/v$ (that is, the risk and/or the degree of the agent's risk aversion are always small) then there is NAM: low-risk-averse agents are matched with high-risk principals (projects) and vice-versa.

If we now rethink the relationship between risk and performance pay, there are two effects. There is the direct effect of σ^2 on s , the same that is present in the standard principal-agent model, which is always negative. There is also an indirect effect of σ^2 on s through the assignment that may be negative (if PAM, because a high σ^2 is matched with a high r , which leads to a low s) or positive (if NAM). Thus, while the relationship between risk and performance pay is certainly negative if $\sigma_L^2 r_L \geq 1/v$ (because of PAM), it can be positive or have any other shape (like a U shape), otherwise.

Using a similar model, Li and Ueda (2009) analyze the relationship between risk and ability. As in Serfes (2005), each principal is characterized by the variance of her project but, in contrast with that paper, Li and Ueda (2009) assume that agents are heterogeneous in terms of ability. At equilibrium a better agent is matched with a firm whose project has lower variance. In their set-up, this provides an explanation for the fact that safer firms receive funding from more reputable venture capitalists.

⁵³ Although the matching models typically involve a finite set of members on both sides, here we present the continuous model, as in Serfes (2005), because the conditions are easier to write. See also Serfes (2008) for a similar analysis with discrete sets.

4.2 The nature of the matching between principals and agents under moral hazard

The presence of moral hazard in a relationship not only changes the characteristics and efficiency of the contract, it may also influence the identity of the principals and agents that decide to establish a partnership. We illustrate a reversal in the nature of the matching using the model introduced as Example 3, due to Ghatak and Karivanov (2014), where principals and agents are risk neutral and the relationship of a principal with characteristic θ_P and an agent with characteristic θ_A produces an output of $h(e, a) = \alpha\theta_A\theta_B + \theta_Ae + \theta_Pa + \varepsilon$.

We now consider a finite set of heterogeneous principals, each endowed with a characteristic θ_P and a finite set of heterogeneous agents, each endowed with a characteristic θ_A . Assume for simplicity that the size of the two sets is the same, and θ_A and θ_P are always higher than 1.

If efforts are contractible, then the first-best efforts are $e^{FB} = \theta_A$ and $a^{FB} = \theta_B$, and the expected value of the outcome is $h(e^{FB}, a^{FB}) = \alpha\theta_A\theta_B + \theta_A^2 + \theta_P^2$. Thus, if we take into account the cost of the effort, the joint surplus in the relationship, as a function of the characteristics (θ_A, θ_P) , is

$$S^{FB}(\theta_A, \theta_P) = \alpha\theta_A\theta_B + \frac{1}{2}(\theta_A^2 + \theta_P^2).$$

For every $\alpha \geq 0$ the function is increasing in the characteristics θ_A and θ_P and the cross-partial derivative $\frac{\partial^2 S^{FB}}{\partial \theta_A \partial \theta_P}(\theta_A, \theta_P)$ is non-negative. Then, applying results by Legros and Newman (2002) (see also Becker, 1973), the equilibrium satisfies PAM: principals with a high characteristic θ_P end up working with agents with a high characteristic θ_A , and vice-versa.⁵⁴

If efforts are not contractible, then the optimal sharing rule decided by any partnership formed makes a trade-off between providing incentives to the principal and the agent. The second-best efforts are given by (11) and the joint surplus in the relationship is

$$S(\theta_A, \theta_P) = \alpha\theta_A\theta_B + \frac{1}{2}(\theta_A^2 + \theta_P^2) - \frac{1}{2} \frac{\theta_A^2 \theta_P^2}{(\theta_A^2 + \theta_P^2)}.$$

⁵⁴Most models that analyze whether the equilibrium satisfies PAM or NAM consider joint surplus functions that are twice differentiable in the characteristics and hence, they use the cross-partial derivative of the joint surplus to assert the nature of the matching. However, as Besley and Ghatak (2005) state, nonstandard matching arguments are needed in the analysis of horizontal characteristics, that is, when the value function is not twice differentiable in the arguments, for example because it depends on the distance between the characteristics of the principal and the agent.

The cross-partial derivative of $S(\theta_A, \theta_P)$ is now negative for positive but low values of α . Therefore, if α is low then the equilibrium satisfies NAM: principals with a high characteristic θ_P end up working with agents with low characteristics θ_A and vice-versa.

Due to the incentive problem, the modularity of the joint surplus under moral hazard depends on both the complementarity of the characteristics in the production function and the endogenous efforts, which depend on the optimal sharing rule. This rule provides incentives to each participant as a function of the magnitude of his/her type relative to the other. Better incentives are provided to θ_P when θ_A is low rather than when it is high. Therefore, the positive effect of an increase in, say, θ_P on e_P is lower the higher θ_A is. This effect induces a certain substitutability between the types that more than compensates the complementarity in the production function when α is low.

Chakraborty and Citanna (2005) and Kaya and Vereshchagina (2015) also study the nature of the matching in a market where each partnership is subject to double-sided moral hazard. In their contributions, the market has only “one side,” instead of “two sides,” that is, each of the participants can play either of the two roles in the partnerships. Chakraborty and Citanna (2005) propose a model where individuals are heterogeneous in wealth and are subject to limited liability. The wealth level of the individual can matter because of the limited liability, but everyone has identical incentives to hire a rich individual. Thus, under symmetric information, any matching is efficient. However, one of the tasks in the partnership is more effort-intensive than the other. Under moral hazard, to facilitate incentive provision, richer individuals have to be allocated to more effort-intensive tasks, which results in NAM at equilibrium. In Kaya and Vereshchagina (2015), individuals are heterogeneous because (for a given effort) their contribution is different: some individuals are better than others. They study a repeated interaction where, once a partnership is formed, the partners produce a stochastic output in each period. As before, in the absence of moral hazard, equilibrium sorting is indeterminate. However, there are two cases in which moral hazard leads to the formation of heterogeneous teams: when one of the partners makes an inefficient effort (NAM is due to the same reasons as above), and when the optimal level of effort can be sustained by both partners at the beginning of the relationship, and the partners’ types either increase the marginal product of effort or have little impact on the output, so that the output realization is a very informative signal of the effort (NAM is efficient in this case because it allows better punishment strategies).

The moral hazard problem also has an influence on the equilibrium

sharing of the surplus between principals and agents. Although total surplus is reduced because of the moral hazard, an agent with a high θ_A may end up obtaining higher rents because of the existence of the moral hazard problem. When there is moral hazard, a “good” agent is more appealing for a principal with low θ_P , who would be ready to pay him more, increasing his “market bargaining power” (and his expected payoff) with a principal with a high θ_P (see Macho-Stadler and Pérez-Castrillo, 2014).

In the previous model, the moral hazard problem induces a reversal of the nature of the partnerships compared to the first-best matching because the need to provide incentives to both participants makes “asymmetric” partnerships more profitable. When only the agent is subject to moral hazard, a reversal may also happen when the principal can choose between two different instruments. Alonso-Paulí and Pérez-Castrillo (2012) study a situation where the agent receives information about the state of the world after having signed the contract, and this information is relevant for the choice of the optimal effort. The principal can offer either an incentive contract or a contract with a verifiable, but rigid, effort. The second type of contract allows for better management control, but makes it hard for the agent to react to market conditions. Although the matching between principals and agents is PAM when only one type of contract is used in all the partnerships,⁵⁵ the best principals might be willing to renounce hiring the best agents through incentive contracts, signing rigid contracts with lower-ability agents instead.⁵⁶

4.3 Heterogeneity, profits, and efficiency

The utilities that principals and agents obtain at equilibrium are endogenous and depend on the sizes of the populations of principals and agents as well as on their characteristics. To highlight some of the implications of the endogenous market power of principals and agents, consider a simple modification of the model studied in Dam and Pérez-Castrillo (2006). There are n_P homogeneous principals and n_A heterogeneous agents. All participants are risk neutral. Agents differ with respect to their initial

⁵⁵Legros and Newman (2007) provide sufficient conditions for monotone matchings in environments where, as is the case in the framework of Alonso-Paulí and Pérez-Castrillo (2012), utility is not fully transferable.

⁵⁶In dynamic relationships, the agreements can also be governed by two types of contracts: short-term and long-term contracts. When information on the workers’ ability is revealed during the relationships, the market dictates a trade-off between the optimal matching (which requires that principals sign short-term contracts) and incentives (which requires long-term contracts). At equilibrium, the matching is not necessarily PAM because both types of contract can coexist (see Macho-Stadler, Pérez-Castrillo, and Porteiro, 2014).

wealth. An agent a^j has an initial wealth \underline{w}^j , which is known to the principals, with $\underline{w}^1 \geq \underline{w}^2 \geq \dots \geq \underline{w}^{n_A} \geq 0$. Therefore, a contract with agent a^j needs to satisfy the limited liability constraint $w(x) \geq \underline{w}^j$. The relationship is similar to the one introduced in Example 3, where two outcomes are possible, i.e., $x \in \{x_B, x_G\}$.

Given that principals are identical, it is necessarily the case that they obtain the same level of profits at equilibrium; we denote it by $\hat{\pi}$. Therefore, the equilibrium contracts (which are necessarily Pareto optimal) are not governed by the agents' PC, but they will be the contracts that maximize the agents' utility subject to the principal's obtaining $\hat{\pi}$. If $n_P < n_A$, then there will be n_P relationships and $\hat{\pi}$ is the maximum benefit that a principal can obtain by contracting with agent a^{n_P} , or with agent a^{n_P+1} (which is the richest agent that does not sign any contract).⁵⁷ Even though agents are the long side of the market, those with an initial wealth higher than \underline{w}^{n_P} obtain rents and, in fact, they sign a contract that is more efficient than the principal-agent contract, in the sense that effort is closer to the first-best. The rents and the efficiency of the contract signed by agent a^j do not depend on the absolute value of \underline{w}^j but on the relative value of \underline{w}^j compared to \underline{w}^{n_P} . Similarly, if $n_P > n_A$, then there will be n_A relationships, $\hat{\pi} = 0$, and all the rents will go to the agents.

As principals compete for the wealthier agents, they are compelled to offer better contracts in order to attract them. These agents obtain higher utility, the limited liability constraint is less stringent and hence the effort level approaches the first-best. The effect of competition on the power of incentives and the efficiency of the relationship has already been pointed out by Barros and Macho-Stadler (1988), in a situation where two principals compete for a good agent.⁵⁸

The analysis of Dam and Pérez-Castrillo (2006) also indicates that a larger inequality in the distribution of agent wealth leads to more efficient relationships. In their framework, a public authority that would like to distribute some money which could serve as collateral in tenancy

⁵⁷Stable outcomes are typically not unique. For example, in the current situation, $\hat{\pi}$ can be any number in the interval whose lower bound is given by the benefits that a principal obtains by hiring agent a^{n_P} and the upper bound is the benefits she obtains by hiring agent a^{n_P+1} .

⁵⁸The effect of competition on the efficiency of the incentive contracts is also the main objective of Dam (2015). Edmans, Gabaix, and Landlier (2009) and Baranchuk, MacDonald, and Yang (2011) study the implications of the assignment of managerial talent to firm size. Also, Hongy, Serfes, and Thiele (2012) study a market with heterogeneous entrepreneurs and venture capital firms. They show that the entry of new venture capital firms has a "ripple effect" throughout the entire market: all start-ups receive more capital in exchange for less equity and the relationships are more efficient.

relations may need to induce inequality among the tenants. If it distributed a small amount to every tenant, then the relative differences in initial wealth would not change and the landowners would appropriate the additional amount distributed. On the other hand, an unequal distribution of the money among a few tenants improves the efficiency and the agents appropriate more than the additional money they receive.

4.4 Competition among mission-oriented and profit-oriented firms

In many markets, principals are heterogeneous not in terms of productivity or costs but in terms of the importance that they give to their mission. Indeed, many public bureaucracies and private nonprofit organizations give more weight to their mission than to profits. Also, some private profit-oriented firms give some weight to an objective other than profits (for example, the use of clean technologies or the development of the community). Similarly, as we discussed in subsection 3.2, the main heterogeneity among agents (workers) may be due not to their ability or risk aversion, but to their intrinsic motivation to work for certain types of firm. In that subsection, we characterized the optimal principal-agent contract for an agent whose intrinsic motivation to work for the firm is $I(x, \tau)$, which depends on the outcome x and the type of the principal τ (see equation (13)).

To discuss the role of matching the mission preferences of principals and agents, we present a model similar to Besley and Ghatak (2005). We consider a market with two types of principals and two types of agents. The types of all the participants are perfectly observable. In case of success, a profit-oriented principal receives a monetary payoff of $x_G^0 > 0$. The payoff $x_G^M > 0$ that a mission-oriented principal receives in case of success may have a nonpecuniary component. Similarly, there are agents who only care about the monetary reward (we will refer to them as type-0 agents) whereas there are mission-oriented agents who receive an intrinsic motivation of $I^M \equiv I(x_G^M, \tau = M) > 0$ if they work for a mission-oriented firm. To simplify the number of cases, we assume that $I^M \leq x_G^M \leq 2x_G^0$, that is, the agent's intrinsic motivation is not larger than the firm's payoff. Also, we assume that the number of mission-oriented agents is the same as the number of mission-oriented principals.

At the equilibrium matching, there is segregation, in the sense that mission-oriented agents work for mission-oriented principals whereas type-0 agents work for profit-oriented principals. The matching is assortative because it raises organizational productivity.

Even though the nature of the matching is the same irrespective of

the number of principals and agents in the profit-oriented sector, the agents' bonuses and the principals' profits in both sectors are affected by those numbers. Suppose first that there is full employment in the profit-oriented sector (that is, the number of type-0 agents is lower than the number of type-0 firms). Then, the equilibrium bonuses and the optimal effort levels in this market for the two types of agents are

$$\begin{aligned} w^{0*}(x_G^0) &= x_G^0 \text{ and } e^{0*} = x_G^0 \\ w^{M*}(x_G^M) &= x_G^0 - I^M \text{ and } e^{M*} = x_G^0. \end{aligned}$$

Thus, competition for the type-0 agents drives the expected payoff of type-0 principals to zero. The utility that the mission-oriented agents obtain is set by what they could obtain by switching to the profit-oriented sector. The mission-oriented principals benefit from the agents' intrinsic motivation through a reduction in the salary they need to attract them.

Second, if there is unemployment in the profit-oriented sector (that is, the number of type-0 agents is higher than the number of profit-oriented firms) then the supply of motivated agents is determined by their unemployment payoff. The bonuses and optimal levels of effort are

$$\begin{aligned} w^{0*}(x_G^0) &= \frac{1}{2}x_G^0 \text{ and } e^{0*} = \frac{1}{2}x_G^0 \\ w^{M*}(x_G^M) &= \frac{1}{2}(x_G^M - I^M) \text{ and } e^{M*} = \frac{1}{2}(x_G^M + I^M). \end{aligned}$$

In this case, the existence of the market does not influence the levels of the bonuses. It only provides information on the nature of the matching between principals and agents.

References

- [1] Akerlof, G.A. (1982) "Labor contracts as a partial gift exchange," *Quarterly Journal of Economics* 97(4), 543-569.
- [2] Aetna Insurance Co. (Aetna) (1867) *Aetna Guide to Fire Insurance for the Representatives of the Aetna Insurance Co.* Hartford: Aetna.
- [3] Alonso-Paulí, E. and Pérez-Castrillo, D. (2012) "Codes of Best Practice in Competitive Markets for Managers," *Economic Theory* 49, 113-141.
- [4] Aron, D.J. and P. Olivella (1994) "Bonus and penalty schemes as equilibrium incentive devices, with application to manufacturing systems," *Journal of Law, Economics and Organization* 10 (1), 1-34.
- [5] Arrow, K.J. (1963) "Uncertainty and the Welfare Economics of Medical Care," *The American Economic Review* 53(5), 941-973.

- [6] Arrow, K.J. (1968) “The Economics of Moral Hazard: Further Comment,” *The American Economic Review* 58(3), 537-539.
- [7] Arrow, K.J. (1971) “Insurance, Risk and Resource Allocation,” in *Essays in the Theory of Risk Bearing*. Chicago: Markham.
- [8] Auster, S. (2013) “Asymmetric Awareness and Moral Hazard,” *Games and Economic Behavior* 82, 503-521.
- [9] Baranchuk, N., MacDonald, G. and Yang, J. (2011) “The Economics of Super Managers,” *Review of Financial Studies* 24 (10), 3321-3368.
- [10] Barros, F. and Macho-Stadler, I. (1988) “Competition for Managers and Market Efficiency,” *Journal of Economics and Management Strategy* 7, 89-103.
- [11] Bartling, B. (2011) “Relative Performance or Team Evaluation? Optimal Contracts for Other-Regarding Agents,” *Journal of Economic Behavior and Organization* 79 (3), 183-193.
- [12] Becker, G.S. (1973) “A theory of Matching: Part I,” *Journal of Political Economy* 81, 813-846.
- [13] Bénabou, R. and Tirole, J. (2003) “Intrinsic and extrinsic motivation,” *Review of Economics Studies* 70, 489-520.
- [14] Besley, T. and Ghatak, M. (2005) “Competition and Incentives with Motivated Agents,” *The American Economic Review* 95, 617-36.
- [15] Bhattacharyya, S. and Lafontaine, F. (1995) “Double-Sided Moral Hazard and the Nature of Share Contracts,” *RAND Journal of Economics* 26, 761-781.
- [16] Bolton, B. and Dewatripont, M. (2005) *Contract Theory*, MIT Press.
- [17] Bolton, G.E. and Ockenfels, A. (2000) “ERC: A Theory of Equity, Reciprocity, and Competition,” *American Economic Review* 90, 166-193.
- [18] Carroll, G. (2015) “Robustness and Linear Contracts,” *American Economic Review* 105 (2), 536-563.
- [19] Chaigneau, P. (2014) “Risk Aversion, Prudence and Compensation,” *The European Journal of Finance* 21 (15), 1357-1373.
- [20] Chakraborty, A. and Citanna, A. (2005) “Occupational Choice, Incentives and Wealth Distribution,” *Journal of Economic Theory* 122, 206-224.
- [21] Dam, K. (2015) “Job Assignment, Market Power and Managerial Incentives,” *The Quarterly Review of Economics and Finance* 57, 222-233.
- [22] Dam, K. and Pérez-Castrillo, D. (2006) “The Principal Agent Market,” *Frontiers in Economic Theory, Berkeley Electronics* 1 (1).

- [23] Dam, K. and Ruiz-Pérez, D. (2012) “On the Existence of Share Contracts under Limited Liability,” *Economics Letters* 117, 552-555.
- [24] De la Rosa, L.E. (2011) “Overconfidence and Moral Hazard,” *Games and Economic Behavior* 73, 429-451.
- [25] De Meza, D. and Webb, D.C. (2007) “Incentive design under loss aversion,” *Journal of the European Economic Association* 5 (1), 66-92.
- [26] Demougin, D. and Fluet C. (2006) “Group vs. Individual Performance Pay When Workers Are Envious,” in *An Economic Perspective on Entrepreneurial Decision-Making*, D. Demougin and C. Schade (eds.), Duncker & Humblot, Berlin.
- [27] Demski, J., Sappington, D. and Spiller, P. (1988) “Incentive Schemes with Multiple Agents and Bankruptcy Constraints,” *Journal of Economic Theory* 44, 156-167.
- [28] Dur, R. and Glazer, A. (2008) “Optimal Incentive Contracts when Workers Envy their Bosses,” *Journal of Law Economics and Organization* 24 (1), 120-137.
- [29] Dufwenberg, M. and Kirchsteiger, G. (2004) “A Theory of Sequential Reciprocity,” *Games and Economic Behavior* 47, 268-298.
- [30] Edmans, A., Gabaix, X. and Landier, A. (2009) “A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium,” *Review of Financial Studies* 22 (12), 4881-4917.
- [31] Englmaier, F. (2005) “Moral Hazard, Contracts and Social Preferences: A Survey,” in *Psychology, Rationality and Economic Behavior*, Palgrave Macmillan UK, 125-139.
- [32] Englmaier, F. and Wambach, A. (2010) “Optimal Incentive Contracts under Inequity Aversion,” *Games and Economic Behavior* 69 (2), 312-328.
- [33] Englmaier, F. and Leider, S. (2012) “Contractual and Organizational Structure with Reciprocal Agents” *American Economic Journal: Microeconomics* 4 (2), 146-183.
- [34] Eswaran, M. and Kotwal, A. (1985) “A Theory of Two-Tier Labor Markets in Agrarian Economies,” *The American Economic Review* 75(1), 162-177.
- [35] Falk A. and Fischbacher, U. (2006) “A theory of reciprocity,” *Games and Economic Behavior* 54, 293-315.
- [36] Fehr, E. and Schmidt, K.M. (1999) “A Theory of Fairness, Competition, and Cooperation,” *Quarterly Journal of Economics* 114, 817-868.
- [37] Feltham, G.A. and Xie, J. (1994) “Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations,” *The Ac-*

- counting Review* 69 (3), 429-453.
- [38] Fleckinger, P. and Roux, N. (2012) "Collective versus Relative Incentives: the Agency Perspective", mimeo Paris School of Economics.
 - [39] Gale, D. and Shapley, L.S. (1962) "College Admissions and the Stability of Marriage," *American Mathematical Monthly* 69, 9-14.
 - [40] Gervais, S., Heaton, J.B., and Odean, T. (2011), "Overconfidence, Compensation Contracts, and Capital Budgeting," *The Journal of Finance* 66 (5), 1735-1777.
 - [41] Ghatak, M. and Karaivanov, A. (2014) "Contractual Structure in Agriculture with Endogenous Matching," *Journal of Development Economics* 110, 239-249.
 - [42] Ghatak, M. and Pandey, P. (2000) "Contract Choice in Agriculture with Joint Moral Hazard in Effort and Risk," *Journal of Development Economics* 63, 303-326.
 - [43] Goel, A.M. and A.V. Thakor (2006) "Optimal Contracts when Agents Envy each other," Working Paper.
 - [44] Grossman, S.J. and Hart, O.D. (1983) "An Analysis of the Principal-Agent Problem," *Econometrica* 51 (1), 7-46.
 - [45] Guo, B. (forthcoming) "Manager Replacement, Employee Protest and Corporate Control," *Economics of Governance*.
 - [46] Harris, M. and Raviv, A. (1979) "Optimal Incentive Contracts with Imperfect Information," *Journal of Economic Theory* 20, 231-259.
 - [47] Herweg, F., Müller, D. and Weinschenk, P. (2010) "Binary Payment Schemes: Moral Hazard and Loss Aversion," *American Economic Review* 100 (5), 2451-2477.
 - [48] Holmström, B. (1979) "Moral Hazard and Observability," *Bell Journal of Economics* 10, 74-91.
 - [49] Holmström, B. (1982) "Moral Hazard in Teams," *Bell Journal of Economics* 13, 324-340.
 - [50] Holmström, B. and Milgrom, P. (1987) "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica* 55 (2), 303-328.
 - [51] Holmström, B. and Milgrom, P. (1991) "Multitask Principal-Agent Analyses: Incentive Contracts, Assets Ownership, and Job Design," *Journal of Law, Economics, and Organization* 7 (Sup.), 24-52.
 - [52] Hongy, S., Serfes, K. and Thiele, V. (2012) "The Market for Venture Capital: Entry, Competition, and the Survival of Start-Up Companies," SSRN working paper.
 - [53] Innes, R.D. (1990) "Limited Liabilities and Incentive Contracting

- with ex ante Action Choices,” *Journal of Economic Theory* 52, 45-67.
- [54] Itoh, H. (2004) “Moral Hazard and Other-Regarding Preferences,” *Japanese Economic Review* 55 (1), 18-45.
- [55] James, H.S. Jr. (2005) “Why did you do that? An economic examination of the effect of extrinsic compensation on intrinsic motivation and performance,” *Journal of Economic Psychology* 26, 549-566.
- [56] Jewitt, I. (1988) “Justifying the First-Order Approach to Principal-Agent Problems,” *Econometrica* 56 (5), 1177-1190.
- [57] Jewitt, I., Kadan, O. and Swinkels, J.M. (2008) “Moral Hazard with Bounded Payments,” *Journal of Economic Theory* 143, 59-82.
- [58] Kahneman, D. and Tversky, A. (1984) “Choices, Values, and Frames,” *American Psychologist* 39 (4), 341-350.
- [59] Kaneko, M. (1982) “The Central Assignment Game and the Assignment Markets,” *Journal of Mathematical Economics* 10, 205-32.
- [60] Kaya, A. and Vereshchagina, G. (2015) “Moral Hazard and Sorting in a Market for Partnerships,” *Economic Theory* 60, 73-121.
- [61] Kim, S.K. (1997) “Limited Liability and Bonus Contracts,” *Journal of Economics and Management Strategy* 6, 899-913.
- [62] Kirkegaard, R. (2014) “A Unifying Approach to Incentive Compatibility in Moral Hazard Problems,” University of Guelph working paper.
- [63] Kirkegaard, R. (2015) “Contracting with Private Rewards,” University of Guelph working paper.
- [64] Köszegi, B. (2014) “Behavioral Contract Theory,” *Journal of Economic Literature* 52 (4), 1075-1118.
- [65] Köszegi, B. and Rabin, M. (2006) “A Model of Reference-Dependent Preferences,” *Quarterly Journal of Economics* 121 (4), 1133-1166.
- [66] Köszegi, B. and Rabin, M. (2007) “Reference-Dependent Risk Preferences,” *American Economic Review* 97 (4), 1047-73.
- [67] Kreps, D.M. (1997) “Intrinsic motivation and extrinsic incentives,” *American Economic Review* 87 (2), 359-364.
- [68] Laffont, J.J. and Martimort, D. (2002) *The theory of incentives: the principal-agent model*, Princeton University Press, Princeton.
- [69] Legros, P. and Newman, A. (2002) “Monotone Matching in Perfect and Imperfect Worlds,” *Review of Economic Studies* 69, 925-942.
- [70] Legros, P. and Newman, A. (2007) “Beauty is a Beast, Frog is a Prince: Assortative Matching with Nontransferabilities,” *Econo-*

- metrica* 75, 1073-102.
- [71] Lewis, T. (1980) “Bonuses and Penalties in Incentive Contracting,” *The Bell Journal of Economics* 11, 291-301.
 - [72] Li, F. and Ueda, M. (2009) “Why Do Reputable Agents Work for Safer Firms?” *Finance Research Letters* 6, 2-12.
 - [73] Macho-Stadler, I. and Pérez-Castrillo, D. (1997) *An introduction to the Economics of Information: incentives and contracts*, Oxford University Press, Oxford.
 - [74] Macho-Stadler, I. and Pérez-Castrillo, D. (2014) “Copyright Licensing under Asymmetric Information,” *Review of Economic Research on Copyright Issues* 11 (2), 1-26.
 - [75] Macho-Stadler, I., Pérez-Castrillo, D. and Porteiro, N. (2014) “Co-existence of Long-term and Short-term Contracts,” *Games and Economic Behavior* 86, 145-164.
 - [76] Mas-Colell, A., Whinston, M.D. and Green, J.R. (1985) *Microeconomic Theory*, Oxford University Press, New York.
 - [77] Matthews, S.A. (2001) “Renegotiating Moral Hazard Contracts Under Limited Liability and Monotonicity,” *Journal of Economic Theory* 97, 1-29.
 - [78] Menezes, C., Geiss, C. and Tressler, J. (1980) “Increasing Downside Risk,” *American Economic Review* 70 (5), 921-932.
 - [79] Milgrom, P.R. (1981) “Good News and Bad News: Representation Theorems and Applications,” *The Bell Journal of Economics* 12, 380-391.
 - [80] Mirrlees, J. (1975) “The Theory of Moral Hazard and Unobservable Behavior-Part I,” Oxford University working paper.
 - [81] Mirrlees, J. (1976) “The Optimal Structure of Incentives and Authority within an Organization,” *The Bell Journal of Economics* 7, 105-131.
 - [82] Mirrlees, J. (1999) “The Theory of Moral Hazard and Unobservable Behavior-Part I,” *Review of Economic Studies* 66 (1), 3-21.
 - [83] Mookherjee (1984) “Optimal incentive schemes with many agents,” *Review of Economic Studies* 51 (3), 433-446.
 - [84] Murdock, K. (2002) “Intrinsic motivation and optimal incentive contracts,” *Rand Journal of Economics* 33 (4), 650-671.
 - [85] Oyer, P. (2000) “A Theory of Sales Quotas with Limited Liability and Rent Sharing,” *Journal of Labor Economics* 18, 405-426.
 - [86] Park, E.-S. (1995) “Incentive Contracting Under Limited Liability,” *Journal of Economics and Management Strategy* 4, 477-490.
 - [87] Pauly, M.V. (1968) “The Economics of Moral Hazard: Comment,” *The American Economic Review* 58 (3), 531-537.
 - [88] Poblete, J. and Spulber, D. (2012) “The Form of Incentive Con-

- tracts: Agency with Moral Hazard, Risk Neutrality, and Limited Liability,” *Rand Journal of Economics* 43, 215-234.
- [89] Quérrou, N., Soubeyran, A. and Soubeyran, R. (2015) “Moral Hazard and Capability,” working paper, Université de Montpellier.
- [90] Rabin, M. (1993) “Incorporating Fairness into Game Theory and Economics,” *American Economic Review* 83 (5), 1281-1302.
- [91] Rabin, M. (2002) “A Perspective on Psychology and Economics,” *European Economic Review* 46, 657-685.
- [92] Rey-Biel, P. (2008) “Inequity Aversion and Team Incentives,” *The Scandinavian Journal of Economics* 110 (2), 297-320.
- [93] Rogerson, W.R. (1985) “The First Order Approach to Principal-agents Problems,” *Econometrica* 53, 1357-1367.
- [94] Romano, R.E. (1994) “Double Moral Hazard and Resale Price Maintenance,” *RAND Journal of Economics* 25, 455-466.
- [95] Ross, S. (1973) “The Economic Theory of Agency: The Principal’s Problem,” *American Economic Review* 63, 134-139.
- [96] Roth, A. and Sotomayor, M. (1990) *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, Cambridge University Press, New York and Melbourne, Econometric Society Monographs.
- [97] Salanie, B. (1997) *The Economics of Contracts*, MIT Press.
- [98] Santos-Pinto, L. (2008) “Positive Self-image and Incentives in Organisations,” *The Economic Journal* 118 (531), 1315-1332.
- [99] Sappington, D. (1983) “Limited Liability Contracts Between Principal and Agent,” *Journal of Economic Theory* 29, 1-21.
- [100] Segal, U. and Sobel, J. (2007) “Tit for Tat: Foundations of Preferences for Reciprocity in Strategic Settings,” *Journal of Economic Theory* 136 (1), 197-216.
- [101] Serfes, K. (2005) “Risk Sharing vs. Incentives: Contract Design under Two-sided Heterogeneity,” *Economics Letters* 88, 343-349.
- [102] Serfes, K. (2008) “Endogeneous Matching in a Market with Heterogeneous Principals and agents,” *International Journal of Game Theory* 36, 587-619.
- [103] Shapley, L.I. and Shubik, M. (1972) “The Assignment Game I: The Core,” *International Journal of Game Theory* 1, 111-130.
- [104] Shavell, S. (1979a) “On Moral Hazard and Insurance,” *Quarterly Journal of Economics* 93, 541-562.
- [105] Shavell, S. (1979b) “Risk Sharing and Incentives in the Principal and Agent Relationship,” *The Bell Journal of Economics* 10, 55-73.
- [106] Sinclair-Desgagné, B. and Spaeter, S. (2013) “The Prudent Principal,” GATE working paper.

- [107] Sobel, J. (2005) “Interdependent Preferences and Reciprocity,” *Journal of Economic Literature* XLIII, 392-436.
- [108] Spence, M. and Zeckhauser, R. (1971) “Insurance, Information, and Individual Action,” *American Economic Review* 61, 380-387.
- [109] Stiglitz, J.E. (1974) “Incentives and Risk Sharing in Sharecropping,” *Review of Economic Studies* 41, 219-256.
- [110] von Thadden, E.L. and Zhao, X. (2012) “Incentives for Unaware Agents,” *Review of Economic Studies* 79 (3), 1151-1174.
- [111] von Thadden, E.L. and Zhao, X. (2014) “Multi-task agency with unawareness,” *Theory and Decision* 77 (2), 197-222.
- [112] Zeckhauser, R. (1970) “Medical Insurance: A Case Study of the Trade-Off Between Risk Spreading and Appropriate Incentives,” *Journal of Economic Theory* 2, 10-26.