Journal of Economic Behavior & Organization
Vol. 44 (2001) 455–465

JOURNAL OF
Economic Behavior
& Organization

www.elsevier.com/locate/econbase

Incentive schemes as a signaling device

Roman Inderst[1]

University of Mannheim, Sonderforschungsbereich 504, 68131 Mannheim, Germany

Received 3 June 1999; received in revised form 31 January 2000; accepted 16 February 2000

Abstract

This paper considers a model of moral hazard where the principal has private information. For instance, in an organizational setting, the firm may be better informed about the profitability of a sales area for which it seeks to employ a new sales representative. We show how this information asymmetry may lead to a game of signaling which can mute incentives. © 2001 Elsevier Science B.V. All rights reserved.

JEL classification: D23; D82

Keywords: Incentives; Signaling; Moral hazard; Adverse selection

1. Introduction

In recent years, numerous contributions have been made to solve the puzzle of why firms' managerial or labor contracts provide rather low incentives compared to the predictions of standard models of moral hazard (see Gibbons, 1990 for a summary). In this paper, we show how the introduction of private information of the principal allows us to derive a new rationale for flat (or low-powered) incentives.

To fix ideas, consider a setting where the principal (e.g. the owner of a firm) has superior information about the profitability of a specific sales area. This is a realistic assumption as he has access to usually unverifiable long-term sales records. As the effort of an employed sales representative is typically not observable, the principal should propose a contract in which the agent participates in the realized sales. Indeed, if both parties are risk neutral and if the profitability of the sales area, i.e. the principal's type, is known to the agent, the inherent problem of moral hazard can be perfectly resolved. The optimal contract basically 'sells' the sales area to the agent in order to elicit first-best effort. The agent's wage thus fully reflects the fortunes of his sales area, while the principal adjusts the base wage (i.e.

E-mail address: r.inderst@ucl.ac.uk (R. Inderst).
[1] URL: http://www.ucl.ac.uk/%7Euctprin.

PII: S0167-2681(00)00142-6

the wage paid in the worst case) to make the agent just indifferent between accepting and rejecting.

If the principal has, however, private information about his type, the employee's evaluation of a proposed incentive scheme depends crucially on his beliefs about the profitability of the sales area. This dependency is strongest if the contract provides high incentives, while it is mitigated if the agent's income is less linked to sales revenues. As a consequence, the first-best contracts signed under complete information are no longer feasible as they do not satisfy incentive compatibility for less profitable principals. Loosely speaking, since an agent who fully participates in the fortunes of the firm is prepared to work for a more profitable principal at a lower base wage, principals with less profitable areas would strictly prefer to mimic this offer. We show how incentive compatibility can be restored at the cost of reducing incentives. In technical terms, the steepness (or power) of the incentive scheme plays the role of a sorting variable. As games of signaling are typically plagued by a multiplicity of equilibria, we impose refinement criteria. We find that the contract proposed by a more profitable principal provides less than first-best incentives to credibly signal the firm's type.

To our knowledge, the explanation of flat or low-powered incentive schemes by private information of the principal is new to the literature. In fact, as noted by Sappington (1991), contract design by an informed principal has altogether received little attention. Among the few contributions are Lutz (1989), Spier (1992), and Jost (1996), who all study different questions such as the use of warranties. Moreover, Maskin and Tirole (1992) have offered a general framework for analyzing mechanism design by an informed principal.

The rest of the paper is organized as follows. Section 2 introduces a simple model of a one-shot signaling game, which is analyzed in Section 3. In Section 4, we allow for renegotiations to discuss the robustness of our results. Section 5 concludes with possible extensions.

2. The model

2.1. Players, payoffs, and the problem of moral hazard

A principal can employ an agent for a single period to produce a real-valued output y. Output depends stochastically on the principal's type $t \in T = \{l, h\}$, which is his private information, and the agent's choice of effort $e \in [0, \bar{e}]$, which is not contractible. Each type $t \in T$ is associated with a profitability index s_t with $s_h > s_l = 0$. The agent has prior beliefs $v = \Pr(t = h)$ with $0 < v < 1$. We specify next a simple stochastic production function. Output may have two realizations $y \in \{0, \bar{y}\}$ with $0 < \bar{y}$, and the production function is additive in effort and profitability with

$$\Pr(y = \bar{y}|e, t) = p_0 + p_1(e + s_t), \tag{1}$$

where $p_1 > 0$. To ensure that no choice of (e, t) leads to a degenerate probability distribution, we require $p_0 > 0$ and $p_0 + p_1(\bar{e} + s_h) < 1$. In what follows, we always obtain an interior choice of effort by assuming additionally $\bar{e} > \bar{y}p_1$. As only the output is contractible, a (deterministic) contract is fully described by a transfer (or wage) α if $y = 0$ and a transfer $\alpha + \beta$ if $y = \bar{y}$. We refer to α as the base (wage) and to β as the premium, which determines the steepness of the wage function $w(y) = \alpha + \beta y/\bar{y}$. We abbreviate a contract by $c = (\alpha, \beta)$ and restrict attention to $c \in C = \Re \times \Re_0^+$, i.e. to nonnegative premia. Reservation values are normalized to zero. Given a wage payment w, an effort e, and an output realization y, the principal realizes the utility $V(w, y) = y - w$ and the agent the utility $U(e, w) = w - \frac{1}{2}e^2$. Note that both parties are (income) risk neutral.

Below we focus on the contracting stage where the principal makes a take-it-or-leave-it offer. We therefore consider already at this point the inherent problem of moral hazard. Given an accepted contract c, the agent must choose $e \in [0, \bar{e}]$ to maximize $\beta(p_0 + p_1(e + s_t)) - \frac{1}{2}e^2$. By the additivity in (1), this choice is unique and independent of the principal's type. It is denoted by the function $\hat{e}(\beta)$, where $\hat{e}(\beta) = \beta p_1$ for $0 \le \beta \le p_1/\bar{e}$ and $\hat{e}(\beta) = \bar{e}$ for $\beta > p_1/\bar{e}$. Given some contract c, we can now derive the principal's 'indirect utility function'. This is given by

$$V_t(c) = -\alpha + (\bar{y} - \beta)(p_0 + p_1\hat{e}(\beta) + p_1 s_t).$$

Similarly, the agent realizes

$$U_t(c) = \alpha + \beta(p_0 + p_1 s_t + p_1\hat{e}(\beta)) - \tfrac{1}{2}\hat{e}(\beta)^2.$$

2.2. *The game of signaling*

We now formalize a two-stage game of signaling where the principal makes a take-it-or-leave-it offer. We represent a behavior strategy for the principal by γ; for each t, $\gamma(\cdot|t)$ is a probability distribution over C. We restrict consideration to distributions $\gamma(\cdot|t)$ with a finite support.[2] We represent a behavior strategy for the agent by $\rho(\cdot)$; for each c, $\rho(\cdot|c)$ is a probability distribution over the set $R = \{0, 1\}$, where 1 denotes acceptance and 0 rejection. After observing a proposal c, the agent updates his prior beliefs v to $\mu(t|c)$, where μ maps C into the simplex Δ_T. Using the derived indirect utility functions from Section 2.1, the principal's expected payoff from proposing c equals

$$\bar{V}_t(c, \rho) = \rho(1|c)V_t(c),$$

while the agent's payoff equals

$$\bar{U}(c, \rho, \mu) = \rho(1|c)\sum_{t \in T}\mu(t|c)U_t(c).$$

An equilibrium is now summarized by the vector of strategies and beliefs $\sigma = (\gamma, \rho, \mu)$ with $\gamma = \{\gamma_t\}_{t \in T}$. We require that in equilibrium strategies are sequentially optimal and beliefs are consistently updated. The set of equilibria is denoted by Σ.[3]

[2] In fact, this comes without loss of generality as it would hold in any equilibrium, since $V_t(c)$ satisfies a standard sorting condition in β.

[3] Formally, an equilibrium is required to be perfect Bayesian, i.e. $\sigma \in \Sigma$ iff

1. $\forall t \in T : \gamma(c|t) > 0$ implies $c \in \arg\max_{c \in C} \bar{V}_t(c, \rho(c))$,
2. $\forall c \in C : \rho(c) \in \arg\max_{r \in [0,1]} r \sum_{t \in T} \mu(t|c)U_t(c)$,
3. $\forall c \in \{c' \in C | \exists t \in T \text{ s.t. } \gamma(c|t) > 0\}$: $\mu(h|c) = v\gamma(c|h)/[v\gamma(c|h) + (1 - v)\gamma(c|l)]$.

For further reference, note that by optimality an agent holding the beliefs $\mu(\cdot|c)$ after observing c will only accept if the following condition of individual rationality holds:

$$\sum_{t\in T}\mu(t|c)U_t(c) = \alpha + \beta\left[p_0 + p_1\left(\sum_{t\in T}\mu(t|c)s_t + \hat{e}(\beta)\right)\right] - \frac{1}{2}\hat{e}(\beta)^2 \geq 0. \qquad \text{(IR)}$$

Finally, in an abuse of notation, we write $\bar{V}_t(\sigma)$ for type t's payoff and $\bar{U}(\sigma)$ for the agent's payoff under $\sigma \in \Sigma$.

3. Analysis of the signaling game

As a benchmark, suppose first that there is complete information about the principal's type. In the unique subgame perfect equilibrium the principal of type t proposes the contract c_t^*, where $\beta^* = \bar{y}$ basically 'sells the firm' and implements first-best effort $e^* = p_1\bar{y}$, while $\alpha_t^* = -\bar{y}(p_0 + p_1 s_t) + \frac{1}{2}(\bar{y}p_1)^2$ is chosen to make the agent indifferent between accepting and rejecting.

By $\alpha_{\mathrm{l}}^* < \alpha_{\mathrm{h}}^*$ these contracts are not incentive compatible if the principal's type is his private information. To restore incentive compatibility, we can now use the premium as a sorting variable. This is possible as the principal's utility satisfies a standard sorting (or single-crossing) condition with respect to the contractual variables. This is expressed in the following lemma.

Lemma 1. *For any $c = (\alpha, \beta)$ and $\hat{c} = (\hat{\alpha}, \hat{\beta})$ with $\hat{\beta} < \beta$ and $V_{\mathrm{l}}(\hat{c}) \geq V_{\mathrm{l}}(c)$, it holds that $V_{\mathrm{h}}(\hat{c}) > V_{\mathrm{h}}(c)$.*

Proof. Substituting $V_{\mathrm{l}}(\hat{c}) \geq V_{\mathrm{l}}(c)$ we obtain

$$V_{\mathrm{h}}(\hat{c}) - V_{\mathrm{h}}(c) = -\hat{\alpha} + (\bar{y} - \hat{\beta})(p_0 + p_1(s_{\mathrm{h}} + \hat{e}(\hat{\beta})) + \alpha - (\bar{y} - \beta)(p_0 + p_1(s_{\mathrm{h}} + \hat{e}(\beta))) \geq s_{\mathrm{h}}(\beta - \hat{\beta}) > 0. \qquad \square$$

Lemma 1 is intuitive. Under a flatter scheme, the principal participates more in the fortunes of the sales area, which is more attractive for the high-type principal. The single-crossing property of Lemma 1 gives rise to the following proposition, which follows immediately from incentive compatibility.

Proposition 1. *In any $\sigma \in \Sigma$ and for any $c \in C$, $\gamma(c|l) > 0$ implies $\gamma(\hat{c}|h) = 0$ for all $\hat{c} \in C$ with $\hat{\beta} > \beta$.*

In words, a high-type principal will not offer a steeper incentive scheme than a less profitable principal. To make more accurate predictions, we must refine Σ as signaling games are typically plagued by a plethora of equilibria.

We first derive the least-cost separating (LCS) allocation. In an equilibrium supporting the LCS allocation, the low type offers his first-best contract c_{l}^*, while the high type's offer

$c_{\mathrm{h}}^{\mathrm{S}}$ is chosen to maximize his utility $V_{\mathrm{h}}(\cdot)$ subject to the agent's participation constraint IR and the low type's incentive compatibility constraint

$$V_{\mathrm{l}}(c_{\mathrm{h}}^{\mathrm{S}}) \leq V_{\mathrm{l}}(c_{\mathrm{l}}^{*}). \tag{IC}$$

The LCS offer $c_{\mathrm{h}}^{\mathrm{S}}$ can be easily calculated.

Lemma 2. *There exists a unique LCS contract $c_{\mathrm{h}}^{\mathrm{S}}$. The premium and the induced effort $e_{\mathrm{h}}^{\mathrm{S}}$ are characterized by*

$$e_{\mathrm{h}}^{\mathrm{S}} = p_1\beta_{\mathrm{h}}^{\mathrm{S}} = e^* + s_{\mathrm{h}} - \sqrt{s_{\mathrm{h}}^2 + 2e^* s_{\mathrm{h}}},$$

where $0 < \beta_{\mathrm{h}}^{\mathrm{S}} < \beta_{\mathrm{h}}^{*} = \bar{y}$. Moreover, there exists an equilibrium $\sigma \in \Sigma$ implementing the LCS allocation, i.e. where $\gamma(c_{\mathrm{l}}^{*}|l) = 1$ and $\gamma(c_{\mathrm{h}}^{\mathrm{S}}|h) = 1$.

Proof. Recall that $c_{\mathrm{h}}^{\mathrm{S}}$ maximizes $V_{\mathrm{h}}(c)$ subject to $V_{\mathrm{l}}(c_{\mathrm{h}}^{\mathrm{S}}) \leq V_{\mathrm{l}}(c_{\mathrm{l}}^{*})$ (IC) and $U_{\mathrm{h}}(c) \geq 0$ (IR). We show first that IC and IR must bind. If $c_{\mathrm{h}}^{\mathrm{S}}$ specifies $\beta_{\mathrm{h}}^{\mathrm{S}} = \bar{y}$, optimality implies $\alpha_{\mathrm{h}}^{\mathrm{S}} = \alpha_{\mathrm{l}}^{*} < \alpha_{\mathrm{h}}^{*}$, such that IC indeed binds. In case $\beta_{\mathrm{h}}^{\mathrm{S}} \neq \bar{y}$, we can use a slack in the IC constraint to move along the agent's indifference curve in order to elicit more or less effort and reap the additional surplus.[4] As IC must therefore bind and as $V_{\mathrm{l}}(c) < V_{\mathrm{l}}(c_{\mathrm{l}}^{*})$ holds strictly for any c with $\beta = 0$ satisfying IR, $\beta_{\mathrm{h}}^{\mathrm{S}} > 0$ follows immediately. Substitution of the binding IC into the payoff function for the high type yields $V_{\mathrm{h}}(c_{\mathrm{h}}^{\mathrm{S}}) = V_{\mathrm{l}}(c_{\mathrm{l}}^{*}) + (\bar{y} - \beta_{\mathrm{h}}^{\mathrm{S}})(s_{\mathrm{h}} - s_{\mathrm{l}})$, such that $\beta_{\mathrm{h}}^{\mathrm{S}} \leq \bar{y}$ follows from optimality. Consider next IR. If the constraint was not binding at a contract specifying $\beta_{\mathrm{h}}^{\mathrm{S}} > 0$, we could use the slack and move marginally along the low type's indifference curve to construct a contract c with a lower premium $\beta < \beta_{\mathrm{h}}^{\mathrm{S}}$, where IR is still satisfied and where $V_{\mathrm{l}}(c) = V_{\mathrm{l}}(c_{\mathrm{h}}^{\mathrm{S}})$. By Lemma 1 it holds that $V_{\mathrm{h}}(c) > V_{\mathrm{h}}(c_{\mathrm{h}}^{\mathrm{S}})$, which contradicts the optimality of leaving IR slack. Having proved $0 < \beta_{\mathrm{h}}^{\mathrm{S}} \leq \bar{y}$ and that both constraints (IR) and (IC) are binding, the construction of $c_{\mathrm{h}}^{\mathrm{S}}$ follows simply from solving the respective equalities. Finally, we can support an equilibrium with $\gamma(c_{\mathrm{l}}^{*}|l) = 1$ and $\gamma(c_{\mathrm{h}}^{\mathrm{S}}|h) = 1$ by specifying that the agent holds pessimistic out-of-equilibrium beliefs ($\mu(c|l) = 1$ for $c \notin \{c_{\mathrm{l}}^{*}, c_{\mathrm{h}}^{\mathrm{S}}\}$). □

By applying a result of Kreps and Sobel (1994), it can be shown that the well-known Intuitive Criterion uniquely selects the LCS allocation. Applied to our two-stage game, where we use the derived 'indirect utility functions' to suppress the inherent problem of moral hazard, the Intuitive Criterion basically implies for any equilibrium $\sigma \in \Sigma$ the following test: there must not exist a contract c with $\gamma(c|t) = 0$ for $t \in T$ such that $\bar{V}_{\mathrm{l}}(c, \hat{\rho}) < \bar{V}_{\mathrm{l}}(\sigma)$ holds for all possible best responses $\hat{\rho}$ to c, while $\bar{V}_{\mathrm{h}}(c, \check{\rho}) > \bar{V}_{\mathrm{h}}(\sigma)$ holds for all best responses $\check{\rho}$ where the agent's beliefs $\check{\mu}(\cdot|c)$ satisfy $\check{\mu}(h|c) = 1$. According to Cho and Kreps (1987), such a contract c constitutes a credible deviation for the high type.

[4] This can be made formal by constructing a function $\hat{\alpha}(\beta)$ with $U_{\mathrm{h}}(\hat{\alpha}(\beta), \beta) = U_{\mathrm{h}}(c_{\mathrm{h}}^{\mathrm{S}})$, which exists from the implicit function theorem given the additivity in α.

Proposition 2. *Any equilibrium* $\sigma \in \Sigma$ *satisfying the Intuitive Criterion implements the LCS allocation with* $\gamma(c_l^*|l) = 1$ *and* $\gamma(c_h^S|h) = 1$.

While the Intuitive Criterion has the advantage of making a clear-cut prediction, it suffers from well-known shortcomings. In particular, it is unsatisfying that the selection is not sensitive to the prior distribution of types. As an extreme implication, observe that for v close to 1 the contract of the high type is much distorted, while for $v = 1$ the game with complete information has a unique equilibrium where the agent chooses first-best effort.

This will be different under the following selection procedure, which will also be useful to obtain more intuition regarding the role of private information (below we have more to say on the choice of refinements). Given a compact subset $\hat{\Sigma}$ of Σ, denote by $M_t(\hat{\Sigma})$ the set of equilibria maximizing the payoff of type t, i.e. $\sigma \in M_t(\hat{\Sigma})$ iff there exists no $\hat{\sigma} \in \hat{\Sigma}$ with $\bar{V}_t(\sigma) < \bar{V}_t(\hat{\sigma})$. Define next the set $M^* = M_l(M_h(\Sigma))$. We will show by construction that this set is non-empty. We call M^* the set of 'lexicographically maximum' equilibria. The following proposition shows that for given prior beliefs v all equilibria in M^* implement a unique allocation. Precisely, there exists a boundary value $v_b \in (0, 1)$ such that all equilibria in M^* implement the LCS allocation if the prior distribution satisfies $v < v_b$, while for $v \geq v_b$ all equilibria in M^* implement a unique pooling allocation. In the latter case, both types offer the contract $c^P(v)$, which depends on the prior distribution v. The base $\alpha^P(v)$ is chosen to reduce the agent's expected payoff to zero, while the premium $\beta^P(v)$ implements the effort level $e^P(v)$ with

$$e^P(v) = p_1\beta^P(v) = p_1[e^* - s_h(1 - v)]. \tag{2}$$

The selection of a unique equilibrium allocation allows us to define a function $\beta_h(v)$ which denotes the high type's premium.

Proposition 3. *There exists a unique value* $v_b \in (0, 1)$ *such that all equilibria* $\sigma \in M^*$ *specify* $\gamma(c_l^*|l) = 1$ *and* $\gamma(c_h^S|h) = 1$ *for* $v < v_b$, *while for* $v \geq v_b$ *it holds that* $\gamma(c^P(v)|l) = \gamma(c^P(v)|h) = 1$. *The premium* $\beta_h(v)$ *offered by the high type is therefore constant for* $v < v_b$ *and strictly increasing for* $v \geq v_b$, *while* $\lim_{v\uparrow v_b} \beta_h(v) < \beta_h(v_b)$.

Proof. Consider first the restriction to equilibria $\sigma \in M_h(\Sigma)$. If $\sigma \in M_h(\Sigma)$ is separating (i.e. there is no c with $\gamma(c|h) > 0$, $\gamma(c|l) > 0$), the uniqueness of c_l^* and c_h^S imply that it must be in pure strategies and implement the LCS allocation. Suppose next that an equilibrium is pooling, i.e. there exists c with $\gamma(c|h) > 0$ and $\gamma(c|l) > 0$. We argue that $\sigma \in M_h(\Sigma)$ implies that both types choose c with probability 1. Note first that by the single-crossing property in Lemma 1 there exists no other contract $c' \neq c$ such that $\gamma(c'|h) > 0$ and $\gamma(c'|l) > 0$. Hence, $0 < \gamma(c|l) < 1$ implies $V_l(c) = V_l(c_l^*)$. In this case, the contract c would also satisfy the constraints of the LCS program, while $V_h(c) < V_h(c_h^S)$ follows from the agent's IR constraint given $\gamma(c|l) > 0$. This contradicts $\sigma \in M_h(\Sigma)$. If the high type randomizes with $0 < \gamma(c|h) < 1$, we can from $\mu(h|c) < v$ construct another equilibrium where both types choose with probability 1 a contract c' yielding $V_h(c') > V_h(c)$. To see this, choose $\beta' = \beta$ and adjust α' upwards until IR is binding for the consistent beliefs $\mu(h|c') = v$. In summary, $\sigma \in M_h(\Sigma)$ implements either the LCS allocation or it holds that both types offer the same (pooling) contract with probability 1.

Consider next a candidate for a pooling equilibrium $\sigma \in M_{\mathrm{h}}(\Sigma)$. An upper boundary for the high type's utility is obtained if IR binds. This allows to derive a unique contract $c^{\mathrm{P}}(v)$, for which the premium is defined in (2). Note that $V_{\mathrm{h}}(c^{\mathrm{P}}(v))$ is continuous and strictly increasing in v. As the utility strictly exceeds $V_{\mathrm{h}}(c_{\mathrm{h}}^{\mathrm{S}})$ at $v = 1$, while it is strictly lower at $v = 0$,[5] there exists a unique value v_{b} where $V_{\mathrm{h}}(c^{\mathrm{P}}(0)) = V_{\mathrm{h}}(c_{\mathrm{h}}^{\mathrm{S}})$. We show next that for $v \geq v_{\mathrm{b}}$ there exists $\sigma \in \Sigma$ which supports the pooling contract $c^{\mathrm{P}}(v)$. If we choose pessimistic out-of-equilibrium beliefs, this holds if $V_{\mathrm{l}}(c^{\mathrm{P}}(v)) \geq V_{\mathrm{l}}(c_{\mathrm{l}}^{*})$. To see that this inequality holds even strictly for $v \geq v_{\mathrm{b}}$, note that otherwise $c^{\mathrm{P}}(v) \neq c_{\mathrm{h}}^{\mathrm{S}}$ would satisfy the constraints of the LCS program, which contradicts the uniqueness of $c_{\mathrm{h}}^{\mathrm{S}}$. We have thus shown that $\sigma \in M_{\mathrm{h}}(\Sigma)$ is non-empty and that any $\sigma \in M_{\mathrm{h}}(\Sigma)$ implements the LCS allocation for $v < v_{\mathrm{b}}$, the pooling contract $c^{\mathrm{P}}(v)$ for $v > v_{\mathrm{b}}$, and either of these allocations for $v = v_{\mathrm{b}}$. As we have also shown that $V_{\mathrm{l}}(c^{\mathrm{P}}(v_{\mathrm{b}})) > V_{\mathrm{l}}(c_{\mathrm{l}}^{*})$, $M^{*} = M_{\mathrm{l}}(M_{\mathrm{h}}(\Sigma))$ removes the ambiguity at $v = v_{\mathrm{b}}$ by selecting the pooling equilibria.

Finally, the assertion regarding $\beta_{\mathrm{h}}(v)$ follows from direct calculations (the jump at $v = v_{\mathrm{b}}$ is intuitive as the high type's gain from separation must be 'compensated' by a higher distortion, i.e. a flatter incentive scheme implying a lower effort). □

The restriction to equilibria in M^{*} provides a useful tool to illustrate the relation between the problem of adverse selection and the steepness of the high type's incentive scheme. As the problem of adverse selection becomes more severe with a decrease in the high type's probability v, the more profitable principal proposes a flatter incentive scheme.

At this point, we should add some more comments regarding the less well-known selection criterion in Proposition 3. We have previously criticized the Intuitive Criterion, which was applied for Proposition 2, as contracts are not sensitive regarding the prior distribution of types. Additionally, the logic of the selection process is highly questionable. More precisely, the 'intuitive' argument underlying this criterion lacks some form of consistency in the following sense. Starting from a given equilibrium, it basically adjusts (out-of-equilibrium) beliefs at a single previously unreached information set. However, if this is supposed to simulate some kind of introspection or forward induction by both the informed and the uninformed party, this cannot be reasonably done without simultaneously adjusting beliefs at other information sets, including some sets on the equilibrium path. This reasoning has induced Mailath et al. (1993) to develop a new criterion, that of 'undefeated equilibria'.[6] While the LCS allocation fails this criterion for $v > v_{\mathrm{b}}$, the selection in Proposition 3 passes this refinement for all values of v.[7] As a final support for the choice in Proposition 3, observe that in contrast to the application of the Intuitive Criterion, the selected allocation is constrained efficient.[8]

[5] The last result follows as $V_{\mathrm{l}}(c^{\mathrm{P}}(0)) < V_{\mathrm{l}}(c_{\mathrm{l}}^{*})$, such that $c^{\mathrm{P}}(0)$ would also satisfy the constraints of the LCS program.

[6] This criterion has recently gained more prominence. For an application, see for instance Taylor (1999).

[7] More precisely, for $v < v_{\mathrm{b}}$, only the characterized allocation can be supported by undefeated equilibria, while for $v \geq v_{\mathrm{b}}$, the selection is not unique. In fact, for $v > v_{\mathrm{b}}$, any pooling equilibrium is undefeated if the high type realizes not less than that in the LCS allocation and if the agent's expected payoff equals zero.

[8] More formally, it satisfies the criterion of interim efficiency proposed by Holmstroem and Myerson (1983).

4. Renegotiations

Observe that the effort implemented by the more profitable principal in the equilibria selected for Propositions 2 and 3 is strictly less than first-best for $v < 1$. It may be argued that contracts where the choice of effort is not efficient will not survive possible renegotiations taking place before the agent actually selects his effort level.

To address this issue, we consider next a model which explicitly allows for renegotiations. However, we still restrict attention to a game of signaling where all offers are made by the informed party, whereas the uninformed side is confined to the passive role of accepting or rejecting. Precisely, we follow Beaudry and Poitevin (1993) and allow (re-)negotiations to proceed over a possibly infinite number of periods. In the period $n = 1$, the principal makes a first proposal. If the agent accepts the offer, the game proceeds into the next round. In $n = 2$, the principal has the possibility to confirm and implement the accepted contract, or to offer a renegotiated proposal. If this is rejected, the game ends and the previous contract is implemented without further renegotiations. If the agent accepts the new offer, the game proceeds into the next round of renegotiations.

Formally, let ω_1^n denote the history of the game until some period n. Recall that in n the principal may make a (new) proposal. Denote by ω_2^n the history of the game including any such new proposal. We denote the respective sets of histories by Ω_i^n for $i = 1, 2$ and $n \geq 0$. As the game starts in $n = 0$, we specify $\Omega_1^1 = \emptyset$. We now extend the principal's strategy space to simplify the exposition of equilibrium strategies. A principal may offer the null-contract $\emptyset$, which ends bargaining and implements the last accepted offer. Denote $C_0 = C \cup \{\emptyset\}$. As in Beaudry and Poitevin (1993) we restrict attention to pure strategies. The principal's strategy is then a sequence of functions $\{\gamma^n\}_{n=1}^{\infty}$ with $\gamma^n(\cdot|t) : \Omega_1^n \to C_0$. Similarly, the strategy of the agent is a sequence $\{\rho^n\}_{n=1}^{\infty}$ with $\rho^n : \Omega_2^n \times C_0 \to R$ and $R = \{0, 1\}$. The agent holds the beliefs $\mu^n : \Omega_2^n \to \Delta_T$ for $i \in \{1, 2\}$. We aggregate $\gamma = \{\gamma^n\}_{n=1}^{\infty}$, $\rho = \{\rho^n\}_{n=1}^{\infty}$, $\mu = \{\mu^n\}_{n=1}^{\infty}$, and $\sigma = (\gamma, \rho, \mu)$. Again we require that strategies are sequentially optimal and that beliefs are consistently updated in an equilibrium.[9] Denote the set of equilibria under renegotiation by Σ^{R}.

Beaudry and Poitevin introduce a new refinement which they call the 'extended divinity criterion'. It follows straightforward from their analysis that this criterion selects a unique allocation in our game. Interestingly, we arrive at the same outcome if we apply the criterion of the 'lexicographically maximum' (M^*) introduced in Section 3. In the selected allocation, a principal of type $t \in T$ implements with probability 1 a unique contract c_t^{R}.[10] The set of contracts $(c_{\mathrm{l}}^{\mathrm{R}}, c_{\mathrm{h}}^{\mathrm{R}})$ is derived as follows. The high type's contract $c_{\mathrm{h}}^{\mathrm{R}}$ maximizes $V_{\mathrm{h}}(c)$ subject to the constraint $vU_{\mathrm{h}}(c) + (1-v)U_{\mathrm{l}}(c) \geq 0$. Given this choice of $c_{\mathrm{h}}^{\mathrm{R}}$, the low type's contract $c_{\mathrm{l}}^{\mathrm{R}}$ maximizes $V_{\mathrm{l}}(c)$ subject to $U_{\mathrm{l}}(c) \geq U_{\mathrm{l}}(c_{\mathrm{h}}^{\mathrm{R}})$.

It is straightforward to verify that by optimality $(1-v)U_{\mathrm{l}}(c_{\mathrm{l}}^{\mathrm{R}}) + vU_{\mathrm{h}}(c_{\mathrm{h}}^{\mathrm{R}}) = 0$, implying $U_{\mathrm{l}}(c_{\mathrm{h}}^{\mathrm{R}}) < 0$ and $U_{\mathrm{h}}(c_{\mathrm{h}}^{\mathrm{R}}) > 0$. Hence, the low type is subsidized under this pair of contracts. If we assume that $s_{\mathrm{h}} < e^*$, the premia under the two contracts are given by

$$\beta_{\mathrm{l}}^{\mathrm{R}} = \bar{y}, \qquad p_1\beta_{\mathrm{h}}^{\mathrm{R}} = e^* - s_{\mathrm{h}}(1-v).$$

[9] For an explicit statement of the equilibrium requirements we refer the reader to Beaudry and Poitevin (1993).

[10] As players are patient as in Beaudry and Poitevin (1993), it is irrelevant when these contracts are implemented.

Hence, the low type's contract implements first-best effort, while the choice of effort is again distorted downwards in the high type's contract. Observe that $0 < \beta_{\rm h}^{\rm R} < \bar{y}$ and that $\beta_{\rm h}^{\rm R}$ is strictly increasing in υ. In analogy to the selection in Proposition 2, the distortion in $\beta_{\rm h}^{\rm R}$ gets more pronounced the higher the prior probability of the low type $1 - \upsilon$.

Proposition 4. *The set of equilibria $M_{\rm h}(\Sigma^{\rm R})$ is non-empty, and in any $\sigma \in M_{\rm h}(\Sigma^{\rm R})$ type $t \in T$ implements $c_t^{\rm R}$.*

Proof. We first construct an equilibrium where t implements $c_t^{\rm R}$. Suppose that both types offer $c_{\rm h}^{\rm R}$ in $n = 1$, which is accepted by the agent. In $n = 2$, the high type offers the null-contract and thus ends the bargaining game. The low type offers $c_{\rm l}^{\rm R}$, which is accepted and implemented in $n = 3$. Note first that the agent's expected utility from accepting in $n = 1$ is just equal to his reservation value. In $n = 2$, the agent is indifferent between $c_{\rm h}^{\rm R}$ and $c_{\rm l}^{\rm R}$ as he knows that he faces the low type. Finally, the strategies for both types of the principal are optimal if we specify that the agent has pessimistic out-of-equilibrium beliefs.

Consider next the set of equilibria $M_{\rm h}(\Sigma^{\rm R})$. Any equilibrium path can be split up into a pooling sequence of offers ending with the acceptance of a contract $c^{\rm P}$, and a sequence of (separating) renegotiations resulting in two possibly different implemented contracts c_t for $t \in T$. Consider first an equilibrium path where $c_{\rm h} = c^{\rm P}$. Optimality for the low type requires that he finally implements a contract $c_{\rm l}$ which maximizes $V_{\rm l}(c)$ subject to $U_{\rm l}(c) \geq U_{\rm l}(c^{\rm P})$. By optimality, this implies $U_{\rm l}(c_{\rm l}) = U_{\rm l}(c^{\rm P})$. Hence, $c^{\rm P} \neq \emptyset$ only satisfies the agent's participation constraint if $\upsilon U_{\rm h}(c^{\rm P}) + (1-\upsilon)U_{\rm l}(c^{\rm P}) \geq 0$. By construction of $c_{\rm h}^{\rm R}$, any contract $c^{\rm P} \neq c_{\rm h}^{\rm R}$ must therefore realize $V_{\rm h}(c^{\rm P}) < V_{\rm h}(c_{\rm h}^{\rm R})$. Hence, any equilibrium $\sigma \in M_{\rm h}(\Sigma^{\rm R})$ where the high type does not renegotiate $c^{\rm P}$ ($c_{\rm h} = c^{\rm P}$) must specify $c^{\rm P} = c_{\rm h}^{\rm R}$.

Suppose next that the high type renegotiates to a contract $c_{\rm h} \neq c^{\rm P}$.[11] We show first that $\sigma \in M_{\rm h}(\Sigma^{\rm R})$ holds iff $c_{\rm h} = c_{\rm h}^{\rm R}$. Consider again the problem to maximize $V_{\rm l}(c)$ subject to $U_{\rm l}(c) \geq U_{\rm l}(c^{\rm P})$. Denote the realized value as a function of $c^{\rm P}$ by $V^{\rm R}(c^{\rm P})$. By the agent's participation constraint (where we can use from a previous argument that $U_{\rm l}(c_{\rm l}) = U_{\rm l}(c^{\rm P})$) and the low type's incentive compatibility constraint, the triple $(c^{\rm P}, c_{\rm l}, c_{\rm h})$ must satisfy: (i) $U_{\rm h}(c_{\rm h}) \geq U_{\rm h}(c^{\rm P})$; (ii) $\upsilon U_{\rm h}(c_{\rm h}) + (1-\upsilon)U_{\rm l}(c^{\rm P}) \geq 0$; (iii) $V^{\rm R}(c^{\rm P}) \geq V_{\rm l}^{\rm R}(c_{\rm h})$. By construction of $V^{\rm R}(\cdot)$, (iii) transforms to $U_{\rm l}(c_{\rm h}) \geq U_{\rm l}(c^{\rm P})$. By (i) and (ii), $c_{\rm h}$ must satisfy $U_{\rm h}(c_{\rm h}) \geq U_{\rm h}(c^{\rm P})$ and $U_{\rm l}(c_{\rm h}) \geq U_{\rm l}(c^{\rm P})$, given the initial pooling offer $c^{\rm P}$. Hence, $c_{\rm h}$ satisfies $\upsilon U_{\rm h}(c_{\rm h}) + (1-\upsilon)U_{\rm l}(c_{\rm h}) \geq 0$ such that $\sigma \in M_{\rm h}(\Sigma^{\rm R})$ iff $c_{\rm h} = c_{\rm h}^{\rm R}$. It remains to show that $c_{\rm h} = c_{\rm h}^{\rm R}$ implies $c_{\rm l} = c_{\rm l}^{\rm R}$. To see this, note first that $\upsilon U_{\rm h}(c_{\rm h}^{\rm R}) + (1-\upsilon)U_{\rm l}(c_{\rm h}^{\rm R}) = 0$ and (ii) implies $U_{\rm l}(c^{\rm P}) \geq U_{\rm l}(c_{\rm h}^{\rm R})$. As $U_{\rm l}(c_{\rm h}^{\rm R}) \geq U_{\rm l}(c^{\rm P})$ follows from (iii), this implies $U_{\rm l}(c_{\rm h}^{\rm R}) = U_{\rm l}(c^{\rm P})$. As by previous arguments $c_{\rm l}$ maximizes $V_{\rm l}(c)$ subject to $U_{\rm l}(c) \geq U_{\rm l}(c^{\rm P}) = U_{\rm l}(c_{\rm h}^{\rm R})$, this implies $c_{\rm l} = c_{\rm l}^{\rm R}$ by the construction of $c_{\rm l}^{\rm R}$. □

Observe that we only apply the one-step selection procedure where we choose the set of equilibria maximizing the high type's utility. This step already yields a unique allocation (recall that this was not the case in Proposition 2 for $\upsilon = \upsilon_{\rm b}$).

[11] Note that this does not imply that $c_{\rm h}$ implements first-best effort. With a switch of support in the agent's beliefs such a proposal may be unprofitable. For more on this issue, see Beaudry and Poitevin (1993).

5. Conclusion

We consider a standard problem of moral hazard with the additional feature that the principal has private information at the time of contracting. This gives rise to a game of signaling. Under various selection criteria we find that more profitable principals (or firms) should offer contracts which provide less than first-best incentives for the agent. By participating more in the fortunes of the firm, this allows more profitable principals to credibly signal their types. Moreover, incentives should be lower if the problem of adverse selection gets more severe as e.g. the prior probability of less profitable types increases. We also discuss the issue of renegotiation and propose a particular game form where these results continue to hold.

In this paper, we only consider two types of principals and restrict consideration to a very simple production function. The extension to a richer set of types is immediate, while we can show that results do not change qualitatively under more general production technologies. [12] A possible extension would be to allow for long-term relationships where production takes place in more than one period. In a long-term relationship, the principal's type is also endogenously revealed over time as output is realized. It would be interesting to know how this is taken into account by equilibrium contracts.

Acknowledgements

Financial support from Deutsche Forschungsgemeinschaft, SFB 504, is gratefully acknowledged. I thank seminar participants at FU Berlin and Humboldt University (Workshop on Corporate Governance) for helpful comments. Comments by an anonymous referee helped to improve the exposition of results.

References

Beaudry, P., Poitevin, M., 1993. Signalling and renegotiation in contractural relationships. Econometrica 61, 745–782.
Cho, I.-K., Kreps, D., 1987. Signalling games and stable equilibria. Quarterly Journal of Economics 102, 1367–1389.
Gibbons, R., 1990. Incentives and careers in organizations. In: Kreps, D., Wallis K. (Eds.), Advances in Economics and Econometrics: Theory and Applications. Proceedings of the Seventh World Congress. Cambridge University Press, Cambridge.
Holmstroem, B., Myerson, R., 1983. Efficient and durable decision rules with incomplete information. Econometrica 51, 1799–1819.
Inderst, R., 1998. Signaling with incentive schemes. SFB 504 Working Paper 98-36.
Jost, P.-J., 1996. On the role of commitment in a principal–agent relationship with an informed principal. Journal of Economic Theory 68, 510–530.
Kreps, D., Sobel, J., 1994. Signalling. In: Aumann, Hart (Eds.), Handbook of Game Theory. Elsevier, Amsterdam.
Lutz, N., 1989. Warranties as signals under consumer moral hazard. Rand Journal of Economics 20, 239–255.
Mailath, G., Okuno-Fujiwara, M., Postlewaite, A., 1993. Belief-based refinements in signalling games. Journal of Economic Theory 60, 241–276.

[12] For more on this, see Inderst (1998).

Maskin, E., Tirole, J., 1992. The principal-agent relationship with an informed principal, II: Common values. Econometrica 60, 1–42.
Sappington, D., 1991. Incentives in principal–agent relationships. Journal of Economic Perspectives 5, 45–66.
Spier, K., 1992. Incomplete contracts and signalling. Rand Journal of Economics 23, 432–443.
Taylor, C., 1999. Time-on-the-market as a sign of quality. Review of Economic Studies 66, 555–578.