MANAGERIAL INCENTIVES BASED
ON INSIDER INFORMATION**

B. Caillaud and B. Jullien*

N° 9128

* C.N.R.S. and C.E.P.R.E.M.A.P, Paris

** The revision of this paper has benefitted from J. Tirole's insightful comments. Financial support from the Commissariat Général au Plan is gratefully acknowledged.
MANAGERIAL INCENTIVES BASED ON INSIDER INFORMATION

Abstract

In a moral hazard situation, we model the fact that the agent may get private signals about the final outcome of his effort (insider information), before the actual and public realization of this outcome. The optimal contract we analyze is equivalent to allowing the agent to realize an option on the final profits before the realization of these profits: it makes the agent bet on the final outcome. Such a contract, which effectively uses messages about signals, will be optimal if the pure moral hazard contract (without messages) involves a large cost for the principal because output allows a poor inference on actions, and if bad actions (for the principal) are not "more informative" about output than good actions (actions that the principal wants to implement) in the sense of Blackwell. The theory provides a justification of the widespread use of stock options (or phantom stock options) in managerial compensation schemes, as opposed to compensation schemes that only rely on salary, bonus and (restricted) stock plans.

J.E.L. Classification Number: 026, 512

Key words: managerial compensation, moral hazard, revelation, Blackwell informativeness of signals, options.

INCITATIONS MANAGERIALES ET INFORMATION D'INITIE

Résumé

En situation d'aléa moral, nous formalisons le fait qu'un agent puisse obtenir de l'information privée sur le résultat final de ses efforts (information d'initié), avant que ce résultat ne soit effectivement réalisé et connu publiquement. Le contrat optimal que nous analysons est équivalent au droit d'exercer une option sur les profits futurs de l'entreprise: ce contrat fait parier l'agent sur ses résultats. Un tel contrat, qui repose sur la révélation de l'information d'initié est optimal si le contrat d'aléa moral pur (sans révélation) est très coûteux pour le principal parce que les résultats permettent une inférence de mauvaise qualité sur les actions prises, et si de mauvaises actions (pour le principal) ne sont pas plus informatives que les bonnes actions (que le principal veut induire), au sens de Blackwell. La théorie offre une justification pour l'usage répandu d'options dans les compensations managériales, en sus des salaires, bonus et participations usuelles.

Nomenclature J.E.L. : 026, 512

Mots clés: Compensations managériales, aléa moral, révélations, contenu informatif au sens de Blackwell, options.
1. INTRODUCTION

There is a lot of evidence that managerial compensation schemes are much more complex than simple fixed salaries. Compensation packages are usually composed of a fixed salary, a bonus contingent on accounting indices of the firm's performances, and a wide variety of financial instruments related to the market value of the stocks of the firm (see Smith-Watts [1982] and Abowd [1990] for example). One of the major applications of the theory of moral hazard has been to provide an explanation of these facts, that does not rely on considerations about tax advantages. Management requires effort from the manager in the form of actions or decisions that are costly to him and that shareholders have not the opportunity or the time to observe. Complex compensation packages are thus designed to provide good work incentives to the manager without submitting him to too high a risk in total revenue: they make managerial compensation positively correlated with different measures of the performances of the firm, either accounting measures or market valuation. The theory therefore explains the use of the three components mentionned above, although some empirical studies suggest that the magnitude of the phenomenon is quite limited (See Jensen-Murphy [1990]).

The theory, however, does not propose any explanation of the use of different financial instruments. These instruments can be classified in two categories. The first type of instruments consists basically in giving the manager a specific stake in the firm, according to a time schedule, usually in the form of shares of the total stock of the firm that cannot be traded before a specified and quite remote date (restricted stocks, stock appreciation rights,...). The second type of instruments leaves the manager
some discretion in choosing the precise compensation package: they take the form of options on the stock of the firm, or warrants, with trade restrictions and a maturity date before which the manager should exercise the option rights if desired. Standard moral hazard models cannot distinguish between these two classes of instruments, and cannot offer a rationale for leaving the manager some discretion. The usual explanation for the use of stock options is either that they may allow non linear compensation schemes (they induce piecewise linear schemes) or that their tax treatment may be more advantageous.

In this paper, we offer an alternative explanation, in terms of incentives in a hidden knowledge model. Stock options have the distinctive role of extracting the manager's insider information on the firm's performances: they constitute a limited and controlled way of allowing insider trading so as to improve work incentives and risk sharing. The use of stock options in managerial compensation packages is proved to be valuable, compared to simple stock plans, when the efforts and activities characterizing good management also endow the manager with "better" insider information about the firm's performances.

It should be rather obvious that managers acquire insider information while running the firm, i.e. information about the future performances of the firm that never get publicly known. When completing a project, it is very likely that the manager gets familiar with the nature of the project, its quality and reliability, the difficulties in its realization and its likely cost. He may spend time studying market surveys or run tests on prototypes in R&D stages. When running efficiently the firm, he may decide some auditing of the organization and supervision of his employees. Indeed
information gathering is a major part of managerial activity. Insider information allows the manager a more accurate prediction on what profits and performances will be, long before actual measures become publicly available. The different signals the manager receives are not observable to the shareholders so that managerial compensation cannot be made directly contingent on them. In other words, the standard results about the redundancy of signals compared to sufficient statistics on the management quality do not apply (Holmström [1979], Shavell [1979]) and cannot characterize when stock options should be used. The use of stock options is related to the desirability of inducing revelation of the manager's information, i.e. of making his compensation contingent on his announcements or messages about the true signals he has obtained.

The optimality of using a managerial contract that relies on messages about insider information depends on whether it allows the shareholders to provide better work incentives at a lower cost than a standard (moral hazard) contract. Messages will be used in an optimal contract if insider information allows the manager a more accurate prediction of final performance when he has undertaken an action preferred by shareholders rather than when he has undertaken an action that shareholders want to deter. In this case, shareholders can design a bet on final performances: this increases the risk of the manager's compensation for a given level of performance, but it relaxes the incentive constraint since after a bad action the bet is worse for the manager; than after a good action therefore it is possible to reduce the overall premium for good performances. An obvious illustration of this trade-off can be readily given. Suppose good management is perfectly informative, i.e. allows the
manager to learn the future profits of the firm before they are actually and publicly known, whereas bad management means playing golf all the time and thereby being uninformed about the health of the firm. Making the manager bet on future profits at an intermediate date is a way to introduce some risk in the compensation of the lazy manager but none in the compensation of a hard-working manager who already knows future profits. Being lazy becomes a less attractive decision for the manager, and there is less need to pay the manager according to the firm's performances.

We show that the effective use of messages in the optimal managerial contract depends on the cost of inducing the desired action without messages, measured by the relative magnitude of likelihood ratios for the good and the bad actions. If this (pure moral hazard) cost is high, it is optimal to try to elicit revelation of insider information; the critical level depends on the comparison in the sense of Blackwell of insider information following a good and a bad action of management. The better the informativeness after a good action compared to a bad action, the lower the critical level, i.e. the more likely it is that messages should be used. In a continuous framework, one can even show that messages will be used except if (locally) bad actions are more informative than better actions.

These findings allow some sharp modification of the nature of optimal managerial contracts. Shareholders should reward the manager not only on the basis of the realization of performances but also on the prediction the manager is able to make on the performance of the firm, before it is publicly known. Of course good performances must be well paid, but the reward must be higher if the manager is himself very confident about the quality of the work he has done. Such a bet-contract will be optimal if,
the harder the manager works, the more he gets acquainted with signals about the future performances of the firm. Note as an example that most supervision activities clearly satisfy the condition. They consist in efforts devoted to acquire relevant information on employees, on their need and on their capacities. Our analysis is therefore of particular interest for this type of situation, and predicts that managers or top executives in charge of monitoring or supervision activities should hold stock options in their compensation package. On the other hand, if gathering insider information can only be made by neglecting productive activities and engaging in wasteful occupation, then bad management actions are more informative than good management actions and our paper proves that stock options should not be used in an optimal managerial contract.

The paper is organized as follows. Section 2 presents the model in a discrete setting and derive some preliminary results that simplify the type of signal technology that we study later. The link with Holmström [1979] or Shavell [1979] results about sufficient statistics is explained in details. Section 3 gives the precise conditions for optimality of the use of options or messages about insider information in a two-action, two-signal, two-output model. Section 4 extends the analysis to a general setting with a continuum of performances, a continuum of actions, and a finite number of signals describing insider information. Section 5 makes the link between our formal model and the practice of financial instruments in managerial compensation contracts.

2. A GENERAL FRAMEWORK

We consider a Principal-Agent relationship under moral hazard where
the agent's unobservable action, called effort and denoted \( a \in A \), generates two stochastic outcomes\( ^{(1)} \):

. a public outcome \( x \in X \) which the principal and the agent and any other party can observe perfectly and costlessly at some date after action \( a \) has been undertaken

. and a signal \( s \in S \) which the agent privately observes after he has undertaken action \( a \) but before the public outcome is observed.

In the paper, we take this formal setting to model the relationships between a (group of) shareholder(s) and a manager, whose management decisions and efforts cannot be directly monitored by the shareholder (or the Board of Directors taken as a group of representatives of the shareholders) and who gets insider information. But our framework is general enough so that it could model many other types of Principal-Agent relationship. In the relationship between an investor and an entrepreneur, the later often gets private signals from his actions before the date of maturity of the debt, i.e. before it is publicly known that he can repay the debt or that he must go bankrupt. In public contracts, the executant receives information about the realization date and final cost overrun during the course of the task. Many examples could be proposed.

Let us present the formal details of the model. \( A \) is a compact of \( \mathbb{R} \) (possibly a finite set). \( S \) and \( X \) are finite sets. Action \( a \) determines the joint distribution of \( (x,s) \), which we summarize by \( p(s|x,a) \) and \( p(x|a) \), respectively the probability of \( s \) conditional on \( (x,a) \) and the marginal probability of \( x \) conditional on \( a \). And we will assume that these probabilities are strictly positive for all \( s, x \) and \( a \), to avoid shifting support problems.
In this setting, a contract determines the agent's salary $t \in \mathbb{R}$ as a function of the final (verifiable) outcome $x$, and of a possible message $m \in M$ reported by the agent in the course of the relationship: $t = T(x,m)$ (where $M$ has at least the cardinality of $A \times S$ a priori). The timing is summarized in Diagram 1.

Diagram 1

```
<table>
<thead>
<tr>
<th>Contract</th>
<th>agent chooses</th>
<th>agent gets</th>
<th>agent reports</th>
<th>$x \in X$</th>
<th>salary paid</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T(.)$ signed</td>
<td>$a \in A$</td>
<td>signal $s \in S$</td>
<td>$m \in M$</td>
<td>occurs</td>
<td>$T(x,m)$</td>
</tr>
</tbody>
</table>
```

The principal is assumed to be risk neutral and cares about net profits $x-t$. The agent is risk averse and cares about revenue and effort; his VNM utility function is $U(t) - C(a)$, $U$ is strictly increasing, weakly concave, twice continuously differentiable and unbounded on $\mathbb{R}$. His reservation utility is normalized at 0.

The above timing shows that a contract $T(.) : X \times M \rightarrow \mathbb{R}$ can be considered as a menu of output-contingent payment schedules (i.e. of bonuses) in which the agent chooses in the course of the relationship, before final output is known. Hence, some discretion is left to the manager in the choice of his compensation schedule, everything is not settled at the contracting date. Thus we can reinterpret such a contract as a general form of option.

The role of communication should not surprise the reader since the mechanism design literature has shown that communication is essential in
presence of private information. Now it is immediate that there is no need to allow communication or messages before signal $s$ is realized or after output $x$ is publicly known. Messages before the realization of $s$ could only bear on what action the agent has taken. They would be useless, just as in a standard moral hazard problem messages are useless. Messages after public realization of output are useless because at that point the agent would just pick the message that maximizes the amount of transfers given that output is known, and not the expected utility of a transfer schedule where the signal $s$ can help predict the distribution of final output.

The Revelation Principle (see e.g. Myerson [1979]) simplifies the problem by asserting that, in our specific framework, the principal can restrict attention to contracts that induce truthful reporting of the agent's private information\(^2\).

Using this, we define a cost minimizing contract implementing action $a^*$.

**Definition 1**: The cost minimizing contract implementing action $a^* \in A$ is a solution of program ($P$):

$$(P) \quad \min_{u(.): X \times S \to \mathbb{R}} \mathbb{E}[U^{-1}(u(x,s)) | a^*]$$

s.t.

1. $\mathbb{E}[u(x,s) | a^*] - C(a^*) \geq 0$ \hspace{1cm} (1)
2. $\mathbb{E}[u(x,s) | a^*] - C(a^*) \geq \mathbb{E}[u(x,m(s)) | a] - C(a)$ \hspace{1cm} (2)

for any $a \in A$ and $m(.): S \to S$.

Note that, as is common in the literature, we use indirect utilities $u(x,s) = U(T(x,s))$ to characterize a contract. (1) is the usual participation or individual rationality constraint. (2) is composed of
truthful revelation constraints when action $a^*$ is taken, and of incentive constraints in the choice of action given that after action $a^*$, the agent's equilibrium strategy of announcement is to tell the truth.

Suppose we neglect the presence of the signal. We would be left with a standard moral hazard problem, and a cost minimizing contract (without messages) would be a function $\nu(.) : X \rightarrow \mathbb{R}$ that satisfies (1) and (2) and minimizes expected transfers among all contracts without messages that satisfy (1) and (2). Given that $S$ is finite, if action $a^*$ can be implemented without messages, i.e. if there exists a cost-minimizing contract without messages that implements $a^*$, then (P) has a solution for $a^*$, i.e. $a^*$ can be implemented by a cost minimizing contract with messages (although possibly degenerate). In what follows we will focus on actions that can be implemented without messages and we will investigate whether the solution of (P) (which therefore exists) actually depends on messages or not. The question is then: can messages improve the cost of implementing an action, and how do they help?

Before answering these questions we will simplify the analysis on two grounds. We will show that there is an equivalence relationship between signal technologies that allows us to restrict attention to a subclass of problems. And we will exhibit a sufficient condition that guarantees that messages can be omitted in a cost-minimizing contract.

Fix the set $X$, $A$ and the marginal distribution of $x$ given $a$, $p(x|a)$. Consider a signal technology as the pair $(S, p(.|x,a))$, composed of the (finite) set of signals and the conditional probabilities of appearance of each signal. We say that $(S, p(.|x,a))$ is equivalent to $(\Sigma, \pi(.|x,a))$ if and only if:
\(\forall \alpha \in A, \exists \psi(\cdot, \alpha) : S \rightarrow \Sigma\) one-to-one, such that
\[
\forall s \in S, p(s | x, \alpha) = \pi(\psi(s, \alpha) | x, \alpha).
\] (3)

Two signal technologies are equivalent if and only if they differ from the agent's perspective only by the labelling of signals, since then, knowing the action \(a\) he has taken, the agent can rename the signals \(s\) as \(\psi(s, a)\).

The agent's perspective, i.e. assuming the action is known, is the right perspective to adopt as illustrated by the following Proposition.

**Proposition 1**: Given \(X, A\) and \(p(z | a)\), if two finite signal technologies \((S, p)\) and \((\Sigma, \pi)\) are equivalent, and \(u(x, \sigma)\) is a cost-minimizing contract implementing action \(a^*\) under \((\Sigma, \pi)\), then for all \(\psi : S \times A \rightarrow \Sigma\) satisfying (3), \(u(x, \psi(s, a^*))\) is a cost-minimizing contract implementing \(a^*\), and the value of the program \((P)\) is the same under both technologies.

**Proof**: Take program \((P)\). For \(g(\cdot) = U^{-1}(\cdot)\) or the identity mapping,
\[
\mathbb{E}[g(u(x, \psi(s, a^*)) | a^*)] = \mathbb{E}[g(u(x, \sigma)) | a^*].
\]

So \(u(x, \psi(s, a^*))\) satisfies (1) and the expected cost of this contract is the same as the expected cost of \(u(x, \sigma)\). Now \(\forall m(\cdot) : S \rightarrow S, \forall \alpha, \exists \mu(\cdot) : \Sigma \rightarrow \Sigma\) such that \(\mu(\cdot) = \psi(m(\psi^{-1}(\cdot, a)), a^*)\). Therefore:
\[
\mathbb{E}[u(x, \psi(m(s), a^*)) | a] = \sum_{x, s} u(x, \psi(m(s), a^*)) p(s | x, a) p(x | a)
\]
\[
= \sum_{x, s} u(c, \psi(m(\psi^{-1}(\psi(s, a), a)), a^*)) \pi(\psi(s, a) | x, a) p(x | a)
\]
\[
= \sum_{x, \sigma} u(c, \psi(m(\psi^{-1}(\sigma, a)), a^*)) \pi(\sigma | x, a) p(x | a)
\]
\[
= \mathbb{E}[u(x, \mu(\sigma)) | a],
\]

and the set of incentive compatibility conditions is also satisfied.

Q.E.D.
Obviously if (3) were true with $\psi(\cdot, a) = \psi''(\cdot)$ for all $a$, both the principal and the agent would know that one signal is a relabelling of the other, and then the cost-minimizing contract would only involve a relabelling. Proposition 1 yields a stronger result since it says that, even though the principal does not observe the action and therefore cannot be certain of the relabelling involved, she can compute the precise relabelling corresponding to the equilibrium action and it is enough to ensure incentive compatibility.

Our second preliminary result is related to the well-known conditions on sufficient statistics. Suppose that the signal $s$ is publicly observable. We know from Holmström [1979] or Shavell [1979] that if $p(s|z, a)$ is independent of $a$, then a cost-minimizing contract implementing $a$ may only be based on output and not on the signal, since output is a sufficient statistics on $a$. A fortiori, if the signal is not observable and output is a sufficient statistics on $a$, the revelation constraints impose some restriction on the set of admissible contracts so that the principal could not achieve a strictly lower cost of implementing $a$ using messages. Hence if $p(s|z, a)$ is independent of $a$, a cost-minimizing contract implementing action $a^*$ does not need to rely on messages. The difference with Holmström is that this condition is not necessary for messages to be useless (not even generically), since the revelation constraints impose a further reduction of the possible role of the signal. So, the sufficient statistics condition determines only a small class of problems where the use of options is not justified to compensate the manager.

Note in passing that Proposition 1 allows a slight generalization of
the sufficient statistics condition in our framework. The idea is that the condition may just hold from the agent perspective: if output is a sufficient statistics for a transformation of the signal conditional on the equilibrium action, then this signal, and therefore a message based on this signal, is useless in a cost minimizing contract.

Proposition 2: Given $X, A, p(x|a)$, if the signal technology $(s, p(.|x,a))$ is equivalent to a signal technology $(\Sigma, \pi(.|x))$ where output is a sufficient statistics on the action, then a cost-minimizing contract does not have to rely on messages.

The proof is a straightforward application of Proposition 1 and is therefore omitted.

3. THE OPTIMALITY OF OPTIONS IN THE SIMPLEST DISCRETE MODEL.

General discrete models of moral hazard are typically quite complicated and involve many difficulties from which we want to abstract. This section provides the economic intuition behind the result on the optimality of the use of options in a very simple two-action, two-output and two-signal model. Formally we take:

$A = \{a_B, a_G\}$

$X = \{x_L, x_H\}$ with $x_L < x_H$

$S = \{\tilde{s}, s\}$

We continue to assume full support distributions. $a_G$ is defined as the good action, and the basic moral hazard conflict of interests is summarized by the fact that $a_G$ dominates $a_B$ in the sense of first order stochastic
dominance, but \( a_g \) is costlier than \( a_b \) for the agent. \( s \) is a signal on output after an action has been taken, and \( \bar{s} \) is normalized to be the good signal, i.e. \( \bar{s} \) is more favorable than \( s \) after any action, in the sense of Milgrom [1981](3). These assumptions are formalized in Al:

\[
\text{Al : } \forall x \in \{x_L, x_G\}, \forall s \in \{\bar{s}, s\}, \forall a \in \{a_B, a_G\},
\]

i) \( p(s|x, a) > 0 \) and \( p(x|a) > 0 \);

ii) \( p(x_H|a_G) > p(x_H|a_B) \);

iii) \( C(a_G) > C(a_B) = 0 \);

iv) \( p(x_H|\bar{s}, a) > p(x_H|s, a) \).

**Lemma 1 :** given Al), Aliv) is equivalent to \( p(s|x_H, a) > p(s|x_L, a) \).

**Proof :**

\[
p(x_H|\bar{s}, a) > p(x_H|s, a) \iff \\
p(s|x_H, a)p(x_H|a) > [1-p(s|x_H, a)]p(x_H|a) \\
\frac{1}{p(s|a)} > \frac{1}{[1-p(s|x_H, a)]p(x_H|a)} \\
1 + \frac{p(s|x_L, a)p(x_H|a)}{p(s|x_H, a)p(x_H|a)} > 1 + \frac{[1-p(s|x_L, a)]p(x_L|a)}{p(s|x_H, a)p(x_H|a)} \\
p(s|x_L, a) < p(s|x_H, a) \iff p(s|x_L, a) < p(s|x_H, a)
\]

Q.E.D.

Assumption Aliv) guarantees that \( \bar{s} \) is a good prediction for the occurrence of high output \( x_H \), hence the name "good signal". The comparison of signals will however rely more on the informational content of signals.
than on the "good news" aspect. It will be crucial to compare the informational content of the signal once \( a_g \) has been taken to the informational content of the signal once \( a_b \) has been chosen. The term "informational content" must be understood in the sense of Blackwell [1951], as reformulated by Gjesdal [1982]. For a formal statement, we denote \( p(.|x,a) \) the line vector of probabilities of \( s \) conditional on \( x \) and \( a \), and \( p(.|.a) \) the matrix obtained by piling up the line vectors \( p(.|x,a) \) for all \( x \):

\[
p(.|x,a) = (p(s|x,a),p(g|x,a)) \text{ and } p(.|.a) = \begin{bmatrix} p(.|x_H,a) \\ p(.|x_L,a) \end{bmatrix}.
\]

Recall also that \( M = (m_{ij})_{ij} \) is a Markov \( I \times I \)-matrix, if and only if:

\[
\forall i \in \{1, \ldots , I\}, \forall j \in \{1, \ldots , I\}, m_{ij} > 0 \text{ and } \sum_{j=1}^{I} m_{ij} = 1.
\]

**Definition 1:** Given two actions \((a,a') \in A^2\), the signal \( s \) after \( a \) has been chosen is said to be a "more informative" signal on output \( x \) than the signal \( s \) after \( a' \) if and only if the distribution of \( s \) conditional on \((x,a')\) is a stochastic transformed of the signal of \( s \) conditional on \((x,a)\), i.e. if and only if there exists \( M(a,a') \), a Markov \(|S| \times |S|\) matrix such that:

\[
p(.|.a') = p(.|.a)M(a,a').
\]

In our two-signal, two-outcome, two-action setting, it is convenient to define some coefficients, which we call Blackwell coefficients and we define as:

\[
b(s,a) = \frac{p(s|x_H,a)}{p(s|x_L,a)}.
\]

The usefulness of these Blackwell coefficients is embodied in the
characterization of the relationship "to be more informative than".

**Proposition 3**: Under A1, \( s \) after \( a \) is a strictly more informative signal on output than \( s \) after \( a' \) if and only if:

\[
b(\tilde{s},a) > b(\tilde{s},a') \text{ and } b(\tilde{s},a) < b(\tilde{s},a').
\]  

(6)

**Proof**: \( s \) after \( a \) is strictly more informative than \( s \) after \( a' \) if and only if \( \exists (\alpha,\beta), \ 0 < \alpha < 1, 0 < \beta < 1 \), such that,

\[
\begin{bmatrix}
p(s|\bar{x}_H,a) p(\bar{x}_H|a') & p(s|\bar{x}_H,a) p(\bar{x}_H|a) \\
p(s|\bar{x}_L,a) p(\bar{x}_L|a') & p(s|\bar{x}_L,a) p(\bar{x}_L|a)
\end{bmatrix} = \begin{bmatrix} \alpha & 1-\alpha \\
1-\beta & \beta \end{bmatrix}.
\]  

(7)

Given assumption A1, \( p(.|L,a) \) is invertible for all \( a \in \{a_0, a_1\} \), so that (7) is equivalent to:

\[
\alpha = \frac{p(s|\bar{x}_L,a)p(\bar{x}_H|a')-p(s|\bar{x}_H,a)p(\bar{x}_L|a')} {p(s|\bar{x}_L,a)p(\bar{x}_H|a)-p(s|\bar{x}_H,a)p(\bar{x}_L|a)}. \]

\[
\beta = \frac{p(s|\bar{x}_H,a)p(\bar{x}_L|a')-p(s|\bar{x}_L,a)p(\bar{x}_H|a')} {p(s|\bar{x}_L,a)p(\bar{x}_H|a)-p(s|\bar{x}_H,a)p(\bar{x}_L|a)}. \]

The denominator is equal to \( p(s|\bar{x}_H,a)-p(s|\bar{x}_L,a) \) and therefore is (strictly) positive. The existence of \( (\alpha,\beta) \) in \( ]0,1[^2 \) is thus equivalent to:

\[
b(\tilde{s},a) > b(\tilde{s},a') \quad (8a)
\]

\[
b(\tilde{s},a') > b(\tilde{s},a) \quad (8b)
\]

\[
b(\tilde{s},a) > b(\tilde{s},a') \quad (8c)
\]

\[
b(\tilde{s},a') < b(\tilde{s},a') \quad (8d)
\]
Now, A1 implies that $0 < b(s,a) < 1 < b(\bar{s},a)$ for all $a$, so (8) is equivalent to (8c)-(8d), i.e. to (6).

Q.E.D.

Proposition 3 shows that the informativeness of $s$ after $a$ is positively related to the difference between the two Blackwell coefficients corresponding to $\bar{s}$ and $s$. To confirm this intuition, note first that when $s$ and $x$ tend to be independently distributed after some action $a$, $b(s,a)$ tends toward 1 for all $s$ which says that the dispersion between the two Blackwell coefficients is minimal. At the opposite, suppose that after some action $a$, $s$ tends to be perfectly correlated with $x$. Then $b(\bar{s},a)$ tends to $\infty$ and $b(s,a)$ tends to 0 and the dispersion between the two Blackwell coefficients tends to be very large.

We are now in a position to state our main theorem. For that, we use the notation:

$$r(a) = \frac{p(x_H|a)}{p(x_L|a)} = \frac{p(x_H|a)}{1 - p(x_H|a)}.$$

**Theorem 1:** i) Given A1, there exists a critical level $e^* \in [0,1]$, such that messages are effectively used in the cost minimizing contract implementing $a_g$ if and only if $\frac{r(a_g)}{r(a_b)} > e^*$.

ii) Moreover, $e^*$ is characterized as follows:

iia) if $b(\bar{s},a_g) > b(\bar{s},a_b)$, then $e^* = \frac{b(s,a_g)}{b(s,a_b)}$;

iib) if $b(\bar{s},a_g) \leq b(\bar{s},a_b)$, $b(s,a_g) < b(s,a_b)$, then

$$\frac{b(s,a_g)}{b(\bar{s},a_b)} \leq e^* = \frac{b(s,a_g)}{p(s|x_L,a_g)} \leq \frac{b(s,a_g)}{p(s|x_H,a_g) - p(s|x_L,a_g)} < \frac{b(s,a_g)}{b(s,a_b)};$$

iic) if $b(\bar{s},a_g) < b(\bar{s},a_b)$ and $b(s,a_g) > b(s,a_b)$, then $e^* = 1$. 
The proof of Theorem 1 is relegated in Appendix. The intuition however can be readily explained. The theorem is composed of two points. Part 1) states that messages are useful whenever the ratio \( \frac{r(a_b)}{r(a_G)} \) is larger than a critical level. Note first that this ratio takes values in \([0,1]\). In the pure moral hazard contract (without messages), the ratio \( \frac{r(a_b)}{r(a_G)} \) measures how costly it is to implement action \( a_G \). When this ratio is high, it is hard to draw much confidence on the action taken just using final output for inference. So the principal must introduce a large risk in the payment to the agent, which has a cost. The informativeness of the signal, however, is independent of this ratio, so that inference using messages, i.e. making the agent bet on his output, becomes more and more attractive compared to using only output, so as to overcome the cost of revelation.

Part ii) of the theorem characterizes the critical level \( e' \). A first intuitive result from Part iic) is that when the signal is more informative about output after \( a_B \) than after \( a_G \), then making the agent predict his output will not help identifying an agent who took the good action. In this case then, messages should not be used.

However, when the distribution of the signal after \( a_G \) dominates the distribution after \( a_B \) (a special case of iia)), \( e' \) will be quite low. It means that, except if output itself is very informative about the choice of actions, messages will be used in the cost minimizing contract inducing the (informative) action \( a_G \), since it is then easy to design a bet on final output which is fair if well informed (after \( a_G \)), but not profitable if not well informed (after \( a_B \)).
Messages can be used despite output is not so poorly informative about action and $a_0$ is not more informative than $a_B$ (Blackwell domination is not a total ordering). For these situations where the distributions cannot be ordered, signal revelation may still be useful and induce $e^* < 1$, even relatively low. Suppose for example that after $a_0$, one signal is almost perfectly informative on output, while it is not the case after $a_B$. Then $e^*$ tends to 0 and unsurprisingly messages should always be used.

Let us try to give an heuristic description of the result of Theorem 1. A contract based on the observation of $z$ and on a report in $\{s, \bar{s}\}$ will be identified to four numbers, $(u, k, z, y)$ defined by:

$$
\begin{align*}
U(t(x_L, s)) &= u + k \\
U(t(x_H, \bar{s})) &= u + k - z \\
U(t(x_L, \bar{s})) &= u - y \\
U(t(x_H, s)) &= u + k - z
\end{align*}
$$

$z$ is the premium in utility terms for a success ($z = x_H$) and a good prediction of success ($s = s$), i.e. for correctly predicting a success. $y$ is the premium for correctly predicting a failure ($z = x_L, s = \bar{s}$). $k$ is the overall premium for success ($z = x_H$).

The revelation principle implies that after $a_0$ has been taken, it must be optimal for the agent to reveal the true signal he received. Revelation constraints can be written as:

$$
\frac{p(x_H | s, a_0)}{p(x_L | s, a_0)} z \geq y \geq \frac{p(x_H | \bar{s}, a_0)}{p(x_L | \bar{s}, a_0)} z .
$$

Hence a cost minimizing contract is such that either $y = z = 0$, in which case it is the simple moral hazard contract, or $y > 0$ and $z > 0$, in which case it relies on messages in the form of two premia for correct prediction of final output.

Given truthful revelation of $s$, the agent faces after $a_0$ the lottery
shown on Diagram 2:

\[ p(x_H | a_g) \quad p(s | x_H, a_g) \quad u+k \]
\[ p(x_L | a_g) \quad p(g | x_H, a_g) \quad u+k-z \]
\[ p(s | x_L, a_g) \quad u-y \]
\[ p(g | x_L, a_g) \quad u \]

Diagram 2

Messages make the agent bet on his final performance on the basis of the information he acquired through the signal. A revelation contract will then reward the agent for correctly guessing the final output \((y, z > 0)\); but a contract tends also to reward the agent for a good output \((k \geq 0, \text{ at least in the pure moral hazard contract})\). Given the lottery in Diagram 1, it is clear that a contract will rely on messages if this allows to trade off the premia for correct guesses against the premium for good outcome, i.e. if increases in \(y\) and \(z\) beyond \(0\) serve to decrease \(k\) substantially compared to a contract without messages. Messages will then introduce a (presumably small) risk due to the agent betting on the outcome, but will help to reduce the overall risk due to the reward for good output. So the use of messages appears as a substitute to output-based rewards, in the set of instruments of the principal. In general both messages and output can be used to determine the agent's reward, but messages may be disregarded since they involve a cost of revelation (risk in betting).
4. GENERAL RESULTS WITH A CONTINUUM OF ACTIONS

The conclusion of Theorem 1 could be rephrased as follows. Unless the informativeness of signals is smaller under \( q_0 \) than under \( q_8 \), the cost minimizing contract that implements \( q_0 \) should use messages when the cost of incentives is high enough in terms of output-based rewards. Turning to a model with a continuum of actions and assuming MLRP, one knows that the principal prefers higher actions and the agent prefers lower actions. In a cost minimizing contract implementing an action \( q_0 \), incentives must be given so that the agent does not want to deviate to \( q_0 - da \), i.e. locally. The local analog of \( \frac{r(a_8)}{r(a_0)} \), i.e. \( \frac{r(a_0 - da)}{r(a_0)} \) tends to 1 when \( da \) tends to 0, so that a direct intuition suggests, following Theorem 1, that if the informativeness of \( s \) after action \( a \) is not locally decreasing in a neighborhood of \( q_0 \), messages should be used in the cost minimizing contract implementing \( q_0 \). This section proves that this intuition is actually valid.

We modify the general framework of section 2 taking \( X \) to be an interval of \( \mathbb{R} \), and \( f(x|a) \) is the density of \( x \) conditional on \( a \). \( p(s_i|x,a) \) and \( f(x|a) \) are assumed \( C^2 \) in \( a \). Higher actions imply higher marginal disutility for the agent: \( C(.) \) is increasing convex. The usual MLRP (See Milgrom [1981]) holds, so that \( \frac{\partial}{\partial x}(f(x|a)) > 0 \).

We also need a local notion of decreasing informativeness. According to Blackwell's condition, the signal will be more informative at \( a \) than at \( a' \), if there exists some Markov matrix \( M(a,a') \) such that for all \( x \),

\[ p(.|x,a) M(a,a') = p(.|x,a') \]

If we write this condition as a first order approximation we obtain:

\[ p(.|x,a)(M(a,a') - Id) = p(.|x,a') - p(.|x,a) + o(a' - a) \]

Set \( M(a,a) = Id \), then we obtain:
We denote \( B(a) = \frac{\partial}{\partial a'} M(a, a') \bigg|_{a'} = a \). The condition is then
\[
\left\{ \begin{array}{l}
p(.lx,a)B(a) = p_a(.lx,a) \\
\text{Id} + B(a) \varepsilon \text{ is Markov for } \varepsilon > 0 \text{ small enough, so that } b_{ij}(a) \geq 0 \text{ for } j \neq i, \\
\sum_{j} b_{ij}(a) = \frac{\partial}{\partial a'} \left( \sum_{i} m_{i,j}(a,a') \right) = 0.
\end{array} \right.
\]

Following this line of reasoning we define decreasing informativeness at some action \( a_0 \) as follows:

**Definition 2 (4)**: Informativeness is locally decreasing at \( a_0 \) if there exists some \((n \times n)\) matrix \( B(a_0) = \{b_{ij}(a_0)\} \) such that:

i) \( \forall x \in X, p_a(.lx,a_0) = p(.lx,a_0)B(a_0) \);

ii) \( \forall i \sum_{j=1}^{n} b_{ij}(a_0) = 0, \quad b_{ij}(a_0) \geq 0 \text{ if } j \neq i. \)

In order to exploit the continuous structure we have to use the first order approach. There are now several well-known sets of conditions that ensure that the first order approach is valid for the standard moral hazard model without messages (see e.g. Rogerson (1985), Jewitt (1988)).

Instead of assuming that the first order approach works, we will concentrate on cost-minimizing contracts implementing an action uniquely. More precisely we will concentrate on regular actions:

**Definition 3**: An action \( a_0 \in A \) is said to be regular if it belongs to the interior of \( A \), and if in the simple moral hazard setting, there exists a cost minimizing contract \( u_0(.) \) implementing \( a_0 \) and such that:
\[
\{a_0\} = \operatorname{Argmax}_{\alpha \in A} \mathbb{E}[u_0(x) | a] - C(a).
\]
A regular action has the nice feature that the cost minimizing contract that implements it satisfies the first order condition: \( \forall x \in X, \)
\[
(U^{-1})'(u_0(x)) = \lambda_0 + \mu_0 \frac{f^a(x|a_0)}{f(x|a_0)}
\]
with \( \mu_0 > 0 \), whenever \( a_0 \) is not the first best action, as proved in the following lemma.

**Lemma 2**: For any regular action \( a_0 \) implemented by \( u_0(.) \) in the simple moral hazard setting and which is not a first best action, there exists \( \lambda_0 \) and \( \mu_0 \), \( \mu_0 \) positive such that (9) holds.

**Proof**: Consider \( \delta(.) : X \rightarrow \mathbb{R} \) integrable with respect to \( f(.|a_0)dx \) satisfying:
\[
\int \delta(x)f(x|a_0)dx = 0, \tag{10}
\]
\[
\int \delta(x)f^a(x|a_0)dx = 0. \tag{11}
\]
Then, there exists \( \epsilon_\delta > 0 \), such that \( \forall \epsilon \leq \epsilon_\delta \),
\[
\{a_0\} = \text{Argmax}_{a \in A} \int [u_0(x) + \epsilon \delta(x)]f(x|a)dx - C(a)
\]
by (11) and the fact that \( a_0 \) is regular. Moreover \( u_0(.) + \epsilon \delta(.) \) is clearly individually rational by (10), so that it constitutes a feasible contract implementing \( a_0 \). Now the cost of this contract for the principal is:
\[
\gamma_\delta(\epsilon) = E[U^{-1}(u_0(x) + \epsilon \delta(x))|a_0].
\]
Since \( u_0(.) \) is the cost-minimizing contract, it must be that there does not exist a \( \delta(.) \) satisfying (10) and (11) such that \( \frac{d}{d\epsilon} \gamma_\delta(\epsilon)|_{\epsilon = 0} < 0 \), i.e. such that:
\[
\int (U^{-1})'(u_0(x))\delta(x)f(x|a_0) \; dx < 0.
\]
So \( \{\delta(.) = 0\} \) must be solution to the program:
Min $\int (U^{-1})'(u_0(x))\delta(x)f(x|a_0)dx$

s.t. (10) and (11)

In this program, the minimand and the constraints are linear in $\delta(.)$, so that the condition for $\{\delta(.)=0\}$ to be solution is that there exist $(\lambda_0,\mu_0)\in \mathbb{R}^2$ such that: for almost all $x$,

$$(U^{-1})(u_0(x))f(x|a_0) = \lambda_0 f(x|a_0) + \mu_0 f_a(x|a_0),$$

i.e. (9). It remains to show that $\mu_0 > 0$. If $\mu_0 = 0$, then $u_0(.)$ would be constant and then $a_0$ would be a first best action, which is ruled out. If $\mu_0 < 0$, then by MLRP, $u_0(.)$ would be strictly decreasing in $x$ (a.e). So a derivative of $u_0(.)$ exists almost everywhere and:

$$\frac{d}{da} \left[ \int u_0(x)f(x|a)dx - C(a) \right] = - \frac{du_0}{dx} \int f(x|a)dx - C'(a)$$

$$= \int \frac{du_0}{dx} \ F_a(x|a)dx - C'(a).$$

MLRP implies first order stochastic dominance so that the expression above would be strictly negative for all $a$ and $a_0$ should be the minimal action $\inf A$, thus not regular. Therefore, $\mu_0 > 0$.

Q.E.D.

Lemma 2 does not contradict Rogerson [1987] findings, since it relies on the regularity property of an action, which is an endogenous property of the moral hazard problem. Since our focus is on the modification of the moral hazard setting, this approach is justified in our paper. It should be noted that almost all implementable actions in a moral hazard problem are regular. The point of course is that the optimal action could precisely be non-regular, as exemplified in the classical criticism of Mirrless [1976] (See the diagram in Grossman-Hart [1983]). The theorem that follows applies
to regular actions. So every assumption that guarantees that the first order approach works for the optimal action will guarantee that our theorem applies to the action that is optimal in the moral hazard setting. It guarantees that a necessary and sufficient condition for messages to be strictly valuable in our framework is that informativeness be not locally decreasing at the implemented regular action.

**Theorem 2**: Under MLRP, consider a regular action $a_0$ that is not a first best action. The cost minimizing contract implementing $a_0$ uses effectively message communication if and only if informativeness is not locally decreasing at $a_0$.

**Proof**: a) Suppose that the cost minimizing contract does not use messages and therefore is $u_0(x)$ implementing $a_0$. Now consider a small deviation consisting of a contract that leads to utility payoff $u_0(x) + \epsilon \delta(x,s)$, with $\epsilon > 0$. Suppose that $\delta(\ldots)$ is chosen such that:

$$
\forall(s_i,s_j) \in S^2, \int \delta(x,s_i)f(x|s_i,a_0)dx > \int \delta(x,s_j)f(x|s_j,a_0)dx,
$$

$$
\sum_i \int \delta(x,s_i) \frac{\partial}{\partial a} (p(s_i|x,a)f(x|a))|_{a=a_0} dx = 0,
$$

$$
\sum_i \int \delta(x,s_i) p(s_i|x,a_0)f(x|a)dx = 0.
$$

Since $u_0(.)$ is individually rational, (14) implies that the deviation is individually rational. (12) implies that the deviation still induces truthful reporting of the signal after $a_0$. Consider now the expected utility of the agent after action $a$ when using an optimal announcement strategy:

$$
\int u_0(x)f(x|a)dx + \epsilon \max_m \left\{ \sum_i \int \delta(x,m(s_i))p(s_i|x,a)f(x|a)dx \right\}.
$$

As a function of $a$, this expression is $C^2$ for any $\epsilon > 0$, and has a local
extremum at \( a = a_0 \) by (13). Now since \( a_0 \) is regular, one can find \( \epsilon_0 \) positive small enough such that \( a_0 \) is a global maximum.

So the deviation \( u_0(.) + \epsilon \delta(.,.) \) is feasible for any \( \epsilon < \epsilon_0 \). If \( u_0(.) \) is the cost-minimizing contract, it must be that for no \( \delta(.,.) \) satisfying (12)-(13)-(14), \( \frac{d}{d\epsilon} \left\{ E \left[ U^{-1}(u_0(x) + \epsilon \delta(x,s)) \operatorname{I} \right] \right\}_{\epsilon=0} < 0 \), i.e.:

\[
\sum_i (U^{-1})'(u_0(x))\delta(x,s_i)f(x|a_0)p(s_i|x,a_0)dx < 0
\]

So, it is necessary that:

\[
0 = \operatorname{Min} \sum_i (U^{-1})'(u_0(x))\delta(x,s_i)p(s_i|x,a_0)f(x|a_0)dx \quad (P1)
\]

s.t. \( \delta(.,.) \) satisfies (12)-(13)-(14).

All constraints and the objective function of this program are linear with respect to \( \delta(.,.) \), so that it is necessary that there exist \( (\lambda,\mu) \in \mathbb{R}^2 \) and \( \beta_{ij}, j \neq i, i,j \in \{1,\ldots,n\} \) positive such that, for all \( s_i \in S \) and almost all \( x \in X \):

\[
(U^{-1})'(u_0(x))p(s_i|x,a_0)f(x|a_0) = \lambda p(s_i|x,a_0)f(x|a_0) + \mu[p_a(s_i|x,a_0)f(x|a_0) + \sum_{j \neq i} \beta_{ij}p(s_i|x,a_0)f(x|a_0)]
\]

\[
- \sum_{j \neq i} \beta_{ij}p(s_j|x,a_0)f(x|a_0).
\]

(15)

\( \lambda \) corresponds to (14), \( \mu \) to (13), and \( \beta_{ij}p(s_i|x,a_0) \) corresponds to (12) for \( s_i \) preferred to \( s_j \) given \( s_i \). Summing (15) over all \( s_i \in S \) for a given \( x \), yields:

\[
(U^{-1})'(u_0(x))f(x|a_0) = \lambda f(x|a_0) + \mu f_a(x|a_0).
\]

(16)

Now given Lemma 2 and MLRP, it is then necessary that \( \lambda = \lambda_0 \) and \( \mu = \mu_0 > 0 \). Using (16) in (15), one gets:

\[
\mu_0 p_a(s_i|x,a_0) + \sum_{j \neq i} \beta_{ij}p(s_i|x,a_0) = \sum_{j \neq i} \beta_{ij}p(s_j|x,a_0).
\]

(17)

Defining \( b_{h\ell} = \beta_{h\ell}/\mu_0 \geq 0 \),

\[
b_{hh} = - \sum_{j \neq h} \beta_{hj}/\mu_0.
\]
one obtains a matrix $B$, with non negative off-diagonal elements, such that

$$\sum_{\ell} b_{\ell,\ell} = 0$$

and

$$p_{a}(l|x,a_0) = p(l|x,a_0).$$

(18)

Hence for $u_0(.)$ to be a cost minimizing contract implementing $a_0$ regular, it is necessary that informativeness be locally decreasing in $a_0$.

b) Conversely, suppose the cost minimizing contract relies effectively on messages. Then, there exists $u_1(.,.) : X \times S \to \mathbb{R}$ integrable, satisfying individual rationality ((1)) and incentive compatibility ((2)) such that:

$$\mathbb{E}[U^{-1}(u_0(x))|a_0] > \mathbb{E}[U^{-1}(u_1(x,s))|a_0].$$

(19)

Since $U^{-1}$ is convex, denoting $\delta(x,s) = u_1(x,s) - u_0(x)$, one has:

$$\mathbb{E}[U^{-1}(u_1(x,s))|a_0] > \mathbb{E}[U^{-1}(u_0(x))|a_0] + \mathbb{E}[(U^{-1})'(u_0(x))\delta(x,s)|a_0].$$

With (19) this implies:

$$\mathbb{E}[(U^{-1})'(u_0(x))\delta(x,s)|a_0] < 0.$$  

(20)

Moreover it is trivial that $\delta(.,.)$ satisfies (12)-(13)-(14) (the individual rationality constraint can be taken to be binding for $u_1(.)$). So if messages are strictly valuable, the program (P1) of part a) of the proof does not yield a value of 0. Now given this program is linear in $\delta$, the condition (18) is necessary and sufficient for (P1) to yield zero value. Hence if messages are valuable, (18) cannot hold for any matrix $B$ with positive off diagonal elements and columns summing to 0. Informativeness cannot be locally decreasing at $a_0$.

Q.E.D.

Theorem 2 gives a strong characterization of situations where the use of messages is strictly valuable in cost-minimizing contracts. Let us provide some intuition. In our framework, given an action $a_0$ regular, the
agent would be always tempted to decrease his action if he only cared about
the cost \( C(a) \). Incentive compatibility indicates precisely that the
distribution of rewards contingent on output would then be less favorable,
so that the natural propensity of the agent be counterbalanced. In the
presence of intermediate signals, if smaller actions generate also better
information at the interim stage on future output, there is no way to
design a bet for the agent that is fair after \( a_0 \) but unprofitable after
\( a < a_0 \). Since in this last case, the agent would have a more precise
information partition. If however smaller actions are not more informative,
it means that some realizations of the signal allow a better prediction on
output after \( a_0 \) than after another \( a < a_0 \). It is then possible to design a
bet, where the reward schedule is highly volatile after those informative
signals, and which is less profitable if \( a < a_0 \) has been taken than if \( a_0 \)
has been taken. Part of the burden of incentive compatibility can be put on
the premium for good prediction of output, which allows ultimately a global
reduction in the premium for good performance, i.e. on the slope \( \frac{du}{dx} \).

The condition that informativeness is non locally decreasing is not
very intuitive. So we will try to illustrate it in the remaining of this
section. Suppose that the signal can only take two values: \( S = \{ s, g \} \).
Informativeness is locally decreasing at \( a_0 \) if and only if,

\[
\exists (\alpha, \beta) \in \mathbb{R}^2, \forall x \in X, \delta_x p(s|x, a_0) + (\alpha + \beta) p(s|x, a_0) = \beta. \tag{21}
\]

Considering (21) for all \( a \) and \( x \), it is easy to obtain a family of signal
distributions for which informativeness is everywhere decreasing: for
example any function \( p(s|x, a) = \lambda(x) \exp(-\alpha a) \) for \( \alpha > 0 \), and \( \lambda(.) \) such that
\( p(s|x, a) \in [0,1] \) for all \( x \in X \), satisfies this property.

Perhaps more interestingly, it is also possible to find two-signal
technologies where informativeness is never decreasing. Suppose effort $a$ determines output on one hand and on the other hand the probability $\pi(a)$ of discovering that $x > x^*$, for some critical level $x^*$ that can be viewed as a threshold of success. $a$ is then perfectly correlated with the possibility of getting informed about success or failure before the actual realization of output. Then $p(\tilde{s}|x,a) = \pi(a)1_{\{x > x^*\}}$ and $\partial_a p(\tilde{s}|x,a) = \pi'(a)1_{\{x > x^*\}}$.

Supposing that $\pi(.)$ is strictly increasing, then it is easy to show that (21) cannot hold for $x \in X$ as long as $x^*$ lies in the interior of $X$. Informativeness is never decreasing for any action and our theorem shows that a cost minimizing contract implementing a regular action should then rely on the use of messages about the signal.

5. ILLUSTRATION IN TERMS OF MANAGERIAL COMPENSATION

Coming back to our simple discrete example, we want now to give a complete illustration of the results obtained in Theorem 1 in terms of managerial compensation schemes. Consider an entrepreneur at date 0 who is selling the firm to the public while designing a managerial contract so as to maximize the firm's value. The entrepreneur foresees that at date 1 the manager that will be hired will take an action $a \in \{a_0, a_1\}$ and will get at date 2 a private signal about the firm's final value. At date 3 the firm's output will be publicly realized for a value $x$ and the firm will be worth nothing more and will be liquidated.

A managerial contract can be designed as follows. At date 0, the entrepreneur writes a corporate charter where the manager's compensation package is described and issues (a normalized number of) one share, of which $(1-\alpha)$ are sold to the public. The managerial compensation package is
composed of:

**a fixed salary**: $\omega$ to be paid at date 3

**a restricted stock plan**: $\alpha$ shares are given to the manager with a clause of restriction of trade until date 3. Alternatively the manager is credited with $\alpha$ shares but is not given the actual shares; he is merely paid at the end of period 3 the cash value of these shares. This is called a phantom stock plan (see Smith-Watts [1982] for a very clear discussion of managerial compensation).

**a stock option** on the firm’s stocks: technically, the firm issues a warrant, i.e. a promise to issue $\beta$ new shares against a cash payment of $p$, if the manager actually decides to exercise this option before date 3. Moreover this warrant is submitted to trade restrictions as well as the $\beta$ shares into which it can be converted. Again, an alternative is to use cash payments that mimics the working of the warrant, which is known as stock appreciation rights (See Smith-Watts [1982], p 142).

It should be noted that the use of these forms of managerial compensation is widely spread, and detailed evidence can be found in Smith-Watts [1982], Abowd [1990] and Jensen-Murphy [1990] among others. There are however very few attempts to explain the differences in terms of incentives between these financial instruments. Most explanations focus on tax advantage differences. Here we provide a theory that points out the equivalence between incentive contracts that rely on messages and the use of stock options in complement to restricted stock plans in managerial compensation.

Given that a warrant is equivalent to an American call option on the stocks of the firm (See Copeland-Weston [1988], p 473-476), it is easy to
compute the manager's conditional payoffs. Suppose that \((\omega, \alpha, p, \beta)\) are such that:

\[
\begin{align*}
&u(\omega+\alpha(x_H-\omega)) = u(x_H, \bar{s}) = u + k - z \\
&u(\omega+\alpha(x_L-\omega)) = u(x_L, \bar{s}) = u \\
&u(\omega + \frac{\alpha + \beta}{1 + \beta} (x_H + p - \omega)) = u(x_H, \bar{s}) = u + k \\
&u(\omega + \frac{\alpha + \beta}{1 + \beta} (x_L + p - \omega)) = u(x_L, \bar{s}) = u - y
\end{align*}
\]

then the compensation package is equivalent to the cost minimizing contract implementing \(a_g\), as described in section 3. The revelation constraints when \(a_g\) is taken are equivalent to the manager exercising the stock option rights if and only if he got insider information \(\bar{s}\) (given that he undertook \(a_g\)). The incentives constraints make it profitable for the manager to undertake \(a_g\).

The manager is prevented from trading any of the financial instruments in his compensation package, so at any date, all potential traders can only exchange shares on the basis of the same public (outside) information. Assuming the exercise of the option right is public information, it is easy to compute the value of one share of the firm at any date:

<table>
<thead>
<tr>
<th></th>
<th>the warrant has not been exercised</th>
<th>the warrant has been exercised</th>
</tr>
</thead>
<tbody>
<tr>
<td>after (x) is known</td>
<td>(x-\omega)</td>
<td>(\frac{x-\omega+p}{1+\beta})</td>
</tr>
<tr>
<td>before (x) is known after maturity</td>
<td>(E[x-\omega</td>
<td>a_g, \bar{s}])</td>
</tr>
<tr>
<td>before date 2</td>
<td>(p(\bar{s}</td>
<td>a_g)E\left[\frac{x-\omega+p}{1+\beta}</td>
</tr>
</tbody>
</table>

In particular, it is simple computation to check that the value of \((1-\alpha)\) share at date 0 is precisely equal to the expected value of gross profits for action \(a_g\), \(E[x|a_g]\), minus the expected payments to the manager.
So the compensation package with trade restriction is equivalent to the contract described in section 3, and is one way for the initial entrepreneur to achieve maximal profits of going public with a manager running the firm.

It is interesting to examine whether trade restrictions are actually needed to achieve the optimum for the entrepreneur. First note that relaxing trade restrictions cannot improve the entrepreneur's profit. Why so? Because allowing trade after the signature of the compensation contract is formally equivalent to allowing renegotiation of the contract. In the paper we assumed that renegotiation could be prevented. Fudenberg-Tirole [1990] have shown that renegotiation in moral hazard models leads to complicated mixed equilibria that are always (weakly) worse than the full commitment solution.

The question is then: is there an equilibrium of the game without trade restrictions which leads to the same profits for the initial entrepreneur? The answer is obviously no because facing the same initial contract, the manager could take \( a_\beta \) and sell all \( \alpha \) shares initially hold (even at price \( E[z - \omega | a_\beta] \)) thereby gaining insurance. This would yield ex-ante utility larger than what he would get in the full commitment situation and would make the deviation on \( a_\beta \) strictly profitable (recall that the optimal contract make both the individual rationality and incentive compatibility contraints binding). Similarly, in situation where the manager is willing to exercise the option after \( (a_\beta, \xi) \) (which may occur for some parameter values as it appears in the proof of Theorem 1 in the Appendix), he would prefer then to sell the \( \beta \) shares acquired via the warrant (even at price \( E[\frac{z+p-\omega}{1+\beta} | a_\beta, \xi] \)) thereby gaining insurance, and the
deviation on $a_b$ would be profitable.

More interestingly, suppose shares are not tradable, whether acquired via the warrant or given initially at date 0. But suppose the warrant is tradable. Would the manager want to sell it at some point, thereby modifying the expected profit for the entrepreneur from proposing the compensation contract? It is easy to see in the Appendix that when $p(x^n | s, a_n) = p(x^n | s, a_b) = p(x^n | s, a_n) < p(x^n | s, a_n)$, the manager is indifferent between exercising the option or not after $(s, a_n)$ or $(s, a_b)$ or $(s, a_b)$ in the cost-minimizing contract implementing $a_n$:

$$E[u(w + \alpha(x - w))|a_n] = E\left[u(w - p + \alpha + \beta(x + p - w))|a_n\right],$$

which in turns implies that the value of the warrant is positive for a risk neutral investor even though he will never get signal $s$. So, the manager would gain by selling the warrant and choosing action $a_n$, thereby upsetting the incentive properties of the compensation package.

Summarizing, there exist parameter values for which trade restrictions are essential to guarantee that the entrepreneur extracts the maximal profit from his decision of selling the firm to the public with the right corporate charter.

The optimal corporate charter determines the precise compensation package under which a manager will be hired. Under the conditions of Theorem 1, this package leaves some discretion to the manager, namely to exercise his option rights on basis of his private, or insider, information about the firm's future profits. This discretionary power is crucial in the model we have proposed and explains why, aside from tax advantages, restricted stock plans and/or stock options (with trade restriction) are used in some companies or for some specific executives and not in other
cases. This can be viewed as a limited and controlled possibility of insider trading, restricted to the exercise of an option with the firm. It could be interesting to pursue this route in an applied study to analyze precisely the type of executives and of firms where stock options are used in relation to the informational content of the decisions and activities attached to the job.
FOOTNOTES

(1) To the usual moral hazard framework (see e.g. Grossman-Hart [1983]), our model adds the presence of a private signal for the agent. For similar models, see Antle [1982] & [1984], Christensen [1981].

(2) To be precise, when looking for a cost-minimizing contract that implements a given action $a^*$, one can restrict attention to contracts that induce truthful reporting of $(s,a^*)$. So one can focus on a contract $T'(x,s) = T(x,s,a^*)$ that induces truthful revelation of $s$ if action $a^*$ is indeed taken, i.e. on the equilibrium path. When considering deviations however, the agent can think of taking another action than $a^*$ and another announcement strategy than truth-telling.

(3) This last part, is composed of a normalization, say $\hat{s}$ after $a_0$, and a simplifying assumption (after $a_B$). Other cases are derived using the equivalence principle of Proposition 1.

(4) Notice that this definition is weaker than the condition: in a neighborhood of $a$, information is decreasing. Indeed it allows the relation $p(.|z,a)M(a,a',x) = p(.|z,a')$ provided that $\frac{\partial}{\partial x} M(a,a',x) |_{a=0} = 0$.\]
REFERENCES


APPENDIX

PROOF OF THEOREM 1

We shall use the notation:
\[ u(x_L, s) = u ; u(x_L, \bar{s}) = u-y ; u(x_H, s) = u+k-z ; u(x_H, \bar{s}) = u+k \]

In what follows we shall extensively use the relation:
\[ \frac{p(x_H | s,a)}{p(x_L | s,a)} = b(s,a)r(a). \]

Direct computation shows that the agent tells the truth after \( a_g \) if and only if:
\[ b(\bar{s},a_g)r(a_g)z > y > b(s,a_g)r(a_g)z \]  \( \text{(R)} \)

Similarly:
- If \( b(\bar{s},a_g)r(a_g)z > y > b(s,a_g)r(a_g)z \), the agent tells the truth after \( a_g \).
- If \( b(s,a_g)r(a_g)z < y \), the agent always announces \( s \) after \( a_g \).
- If \( b(s,a_g)r(a_g)z > y \), the agent always announces \( \bar{s} \) after \( a_g \).

Given the revelation constraint, it is clear that at the optimum \( y = z = 0 \), or \( y > 0 \) and \( z > 0 \), in which case messages are effectively used.

Claim: 1) If \( b(\bar{s},a_g)r(a_g) > b(s,a_g)r(a_g) \), the optimal contract induces truthful revelation of \( s \) after \( a_g \).

2) If \( b(\bar{s},a_g)r(a_g) < b(s,a_g)r(a_g) \), the agent announces \( s \) after \( a_g \) and \( y = b(s,a_g)r(a_g)z \).

Proof: Case 1). Suppose that \( y > b(\bar{s},a_g)r(a_g)z \), so that \( s \) is announced after \( a_g \). Consider the deviation \( dy < 0, du = p(\bar{s}|x_L,a_g)dy, d(u+k) = dz = 0 \). This preserves the agent's utility and truthful revelation after \( a_g \). The expected utility after \( a_g \) is \( u+p(x_H|a_g)(k-s) \); it decreases so that the incentives are maintained. Finally the principal's cost decreases since \( U^{-1} \)
is convex (see diagram 2, section 3, which shows that the deviation is a mean preserving decrease in risk): a contradiction.

A similar argument holds for the case $y < b(g,a_b)r(a_b)z$.

Case 2). In this case (R) implies that after $a_b$, $s$ is always announced. Now if $y > b(g,a_0)r(a_0)z$ the deviation considered in case a) is still feasible and profitable and the contract cannot be optimal.

Q.E.D.

It follows from the claim, that the set of constraints that must be verified reduces to:

1 participation constraint

1 incentive constraint whose form depends on whether $b(s,a_b)r(a_b)$ is larger or smaller than $b(s,a_0)r(a_0)$.

2 revelation constraints of the form: $q_1z \geq y \geq q_2z$, where the precise values of $q_1$ and $q_2$ depend on the comparison between the various values taken by $b(s,a)r(a)$ for $s = s, \bar{s}$ and $a = a_0, a_b$.

The question we want to address is: is the pure moral hazard contract $(u^0,k^0,0,0)$ optimal or not? If, at the optimum, all revelation constraints as strict inequalities, $(u^0,k^0,0,0)$ must be the solution, implying $q_1z = y = q_2z = 0$: a contradiction. It follows that at the optimum, one revelation constraint is verified as an equality. So $(u^0,k^0,0,0)$ is not solution if and only if there exists $q \in \{q_1, q_2\}$ such that it is not solution of the programme obtained by replacing the two revelation constraints $q_1z \geq y \geq q_2z$ by the constraint:

$$y = qz, \ y \geq 0.$$  

The programme to be solved is then:
Max \[ \begin{array}{c}
- p(x_{H}, \bar{s}|a_{0}) U^{-1}(u+k) - p(x_{H}, s|a_{0}) U^{-1}(u+k-z) \\
- p(x_{L}, \bar{s}|a_{0}) U^{-1}(u-y) - p(x_{L}, s|a_{0}) U^{-1}(u) 
\end{array} \]

s.t.
\[ y = qz \quad \text{(R)} \]
\[ y > 0 \quad \text{(P)} \]
\[ u + k p(x_{H}|a_{0}) - z p(x_{H}, \bar{s}|a_{0}) - y p(x_{L}, \bar{s}|a_{0}) - C(a_{0}) \geq 0 \quad \text{(IR)} \]
\[ u + k p(x_{H}|a_{0}) - z p(x_{H}, \bar{s}|a_{0}) - y p(x_{L}, \bar{s}|a_{0}) - C(a_{0}) \geq u + k p(x_{H}|a_{0}) - z p(x_{H}, \bar{s}|a_{0}) - y p(x_{L}, \bar{s}|a_{0}) \quad \text{(IC)} 
\]

\((u^{0}, k^{0}, 0, 0)\), the pure moral hazard contract, is solution if and only if there exist \((\lambda^{0}, \mu^{0}, \rho^{0})\) non negative and \(v^{0}\) such that \((u^{0}, k^{0}, 0, 0)\) satisfies the first order conditions:

\[
\begin{array}{c}
- p(x_{H}|a_{0})(\phi^{0} + \Delta^{0}) - p(x_{L}|a_{0})\phi^{0} + \lambda^{0} = 0 \\
- p(x_{H}|a_{0})(\phi^{0} + \Delta^{0}) + \lambda^{0} p(x_{H}|a_{0}) + \mu^{0}[p(x_{H}|a_{0}) - p(x_{H}|a_{0})] = 0 \\
p(x_{H}, s|a_{0})(\phi^{0} + \Delta^{0}) - \lambda^{0} p(x_{H}, s|a_{0}) - \mu^{0}[p(x_{H}, s|a_{0}) - p(x_{H}, s|a_{0})] + qv^{0} = 0 \\
p(x_{L}, \bar{s}|a_{0})\phi^{0} - \lambda^{0} p(x_{L}, \bar{s}|a_{0}) - \mu^{0}[p(x_{L}, \bar{s}|a_{0}) - p(x_{L}, \bar{s}|a_{0})] + v^{0} + \rho^{0} = 0
\end{array}
\]

where we use the notation:
\[ \phi^{0} = (U^{-1})'(u_{0}) > 0, \]
\[ \Delta^{0} = (U^{-1})'(u_{0} + k_{0}) - (U^{-1})'(u_{0}) > 0. \]

Solving for \((\lambda^{0}, \mu^{0}, \rho^{0}, v^{0})\), one gets:
\[
\begin{aligned}
\lambda^0 &= \phi^0 + p(x_H|a_g) \Delta^0 > 0 \\
\mu^0 &= \frac{p(x_H|a_g)}{p(x_H|a_g) - p(x_H|a_b)} p(x_L|a_g) \Delta^0 > 0 \\
q^0 &= p(x_L|a_g) \Delta^0 \left[ p(x_H, s|a_g) - \frac{p(x_H, s|a_g) - p(x_H, s|a_b)}{p(x_H|a_g) - p(x_H|a_b)} p(x_H|a_g) \right] \\
\rho^0 &= p(x_L, \tilde{s}|a_g) p(x_H|a_g) \Delta^0 + \frac{p(x_L, \tilde{s}|a_g) - p(x_H, \tilde{s}|a_b)}{p(x_H|a_g) - p(x_H|a_b)} p(x_H|a_g) p(x_L|a_g) \Delta^0 \\
&\quad - \frac{p(x_L|a_g) \Delta^0}{q} \left[ p(x_H, s|a_g) - \frac{p(x_H, s|a_g) - p(x_H, s|a_b)}{p(x_H|a_g) - p(x_H|a_b)} p(x_H|a_g) \right]
\end{aligned}
\]

Now, \((u^0, k^0, 0, 0)\) is not solution if and only if \(\rho^0 < 0\).

\[
\iff q(p(\tilde{s}|x_L, a_g) - p(\tilde{s}|x_L, a_b)) < r(a_b)(p(\tilde{s}|x_H, a_g) - p(\tilde{s}|x_H, a_b)).
\]

Let us now consider each case according to the different values of \(q\).

**Case 1 of the Claim :**

\[
\frac{r(a_b)}{r(a_g)} \geq \frac{b(s, a_g)}{b(s, a_b)}
\]

**Case 1A/** : \(q = q_1 = \inf \{b(\tilde{s}, a_b)r(a_g), b(s, a_g)r(a_b)\}\)

1A1/ Suppose : \(b(s, a_g)r(a_g) \geq b(s, a_b)r(a_b) = q_1\).

\((u^0, k^0, 0, 0)\) is not solution if and only if \(\rho^0 < 0\), i.e. :

\[
b(s, a_g) [p(\tilde{s}|x_L, a_g) - p(\tilde{s}|x_L, a_b)] < [p(\tilde{s}|x_H, a_g) - p(\tilde{s}|x_H, a_b)]
\]

\[
\iff p(\tilde{s}|x_H, a_g)p(\tilde{s}|x_L, a_g) < p(\tilde{s}|x_H, a_g)p(\tilde{s}|x_L, a_b)
\]

\[
\iff b(s, a_g) > b(\tilde{s}, a_b)
\]

If this condition is verified, we have \(\frac{b(s, a_g)}{b(\tilde{s}, a_b)} > 1 > \frac{r(a_b)}{r(a_g)}\) automatically.

1AII/ Suppose : \(b(s, a_g)r(a_g) > b(s, a_b)r(a_b) = q_1\).

\[
\rho^0 < 0 \iff r(a_b) \left\{ 1 - \frac{p(\tilde{s}|x_H, a_g)}{p(\tilde{s}|x_H, a_b)} \right\} > r(a_g) \left\{ 1 - \frac{p(\tilde{s}|x_L, a_g)}{p(\tilde{s}|x_L, a_b)} \right\}
\]

which, given \(r(a_g) > r(a_b)\), implies \(b(s, a_g)r(a_g) > b(\tilde{s}, a_b)r(a_b)\). This last inequality is not compatible with the former inequality. So, this case is empty.
Case 1B/ : \( q = q_2 = \text{Sup}\{b(s,a_0)r(a_0), b(s,a_8)r(a_8)\} \)

1BI/ Suppose : \( b(s,a_0)r(a_0) \leq b(s,a_8)r(a_8) = q_2 \).

\( \rho^0 < 0 \iff b(s,a_0) > b(s,a_8) \), which is verified given the assumption above.

1BII/ Suppose : \( b(s,a_0)r(a_0) \leq b(s,a_8)r(a_8) = q_2 \).

\( \rho^0 < 0 \iff r(a_8)b(s,a_0)\{p(s|x_L,a_0) - p(s|x_L,a_8)\} < r(a_8)\{p(s|x_H,a_0) - p(s|x_H,a_8)\} \)

Case 2 of the Claim : \( \frac{b(s,a_0)}{b(s,a_8)} > \frac{r(a_0)}{r(a_8)} \).

In this case we know that \( q = b(s,a_0)r(a_0) \). The incentive constraint is :

\( u + kp(x_H|a_0) - sp(x_H,s|a_0) - yp(x_L,s|a_0) - C(a_0) \geq u + (k-\varepsilon)p(x_H|a_0) \).

The same method yields :

\( \rho^0 < 0 \iff \frac{r(a_0)}{r(a_8)} > \frac{b(s,a_0)}{b(s,a_8)} \).

Summary

Messages will be used if and only if one of the following sets of conditions holds :

\begin{itemize}
  \item either 1AI/ : \( \frac{r(a_0)}{r(a_8)} > \frac{b(s,a_0)}{b(s,a_8)} \) and \( b(s,a_0) > b(s,a_8) \)
  \item or 1BI/ : \( \frac{r(a_0)}{r(a_8)} > \frac{b(s,a_8)}{b(s,a_0)} \)
  \item or 1BII/ : \( \frac{b(s,a_0)}{b(s,a_8)} > \frac{r(a_0)}{r(a_8)} \) and \( b(s,a_0) > b(s,a_8) \)
  \item or 2/ : \( \frac{b(s,a_0)}{b(s,a_8)} > \frac{r(a_0)}{r(a_8)} \)
\end{itemize}

Final characterization

We restate the above conditions according to the cases of Theorem 1.

a) Suppose that \( b(s,a_0) > b(s,a_8) \). Then 1AI/ and 2/ provide a range :
while 1BI/ and 1BII/ are included in this range.

b) Suppose now that \( b(s, a) \leq b(\bar{s}, a_b) \) and \( b(g, a) < b(g, a_b) \). 1AI and 2 are empty cases. From 1BII/, \( e^* \leq \frac{b(s, a)}{b(\bar{s}, a_b)} \). Moreover one can show that:

\[
p(s\mid x_h,a) = b(s,a) \frac{b(s,a) - 1}{b(s,a) - b(g,a)}
\]

Both are increasing function of \( b(\bar{s},a) \) and of \( b(g,a) \). So one has:

\[
p(s\mid x_h,a) - p(s\mid x_h,a_b) = 0
\]

\[
p(s\mid x_L,a) - p(s\mid x_L,a_b) > 0
\]

One can easily verify that:

\[
b(s,a) \frac{p(s\mid x_L,a) - p(s\mid x_L,a_b)}{p(s\mid x_H,a) - p(s\mid x_H,a_b)} \geq \frac{b(s,a)}{b(\bar{s},a_b)}
\]

Therefore \( e^* = b(s,a) \frac{p(s\mid x_L,a) - p(s\mid x_L,a_b)}{p(s\mid x_H,a) - p(s\mid x_H,a_b)} \).

c) Finally, suppose that \( b(s, a) \leq b(\bar{s}, a_b) \) and \( b(g, a) \geq b(g, a_b), 1AI/, 1BII/ and 2/ are empty cases. Moreover, if \( p(s\mid x_H,a) - p(s\mid x_H,a_b) > 0 \) then:

\[
b(s,a) \leq b(\bar{s},a_b) \leq b(g,a) \frac{p(s\mid x_L,a) - p(s\mid x_L,a_b)}{p(s\mid x_H,a) - p(s\mid x_H,a_b)}
\]

and 1BII is empty.

If \( p(s\mid x_H,a) - p(s\mid x_H,a_b) < 0 \), case 1BII would require:

\[
\frac{r(a_b)}{r(a)} < b(s,a) \frac{p(s\mid x_L,a) - p(s\mid x_L,a_b)}{p(s\mid x_H,a) - p(s\mid x_H,a_b)} \leq \frac{b(\bar{s},a_b)}{b(s,a)} \leq \frac{r(a_b)}{r(a)}
\]

So 1BII is also empty.

Q.E.D.