# Hidden Action and Incentives

Bernard Caillaud
Ecole Ponts et Chaussée, Paris, and CNRS

Benjamin E. Hermalin
University of California, Berkeley

A common economic occurrence is the following: Two parties, *principal* and *agent*, are in a situation—typically of their choosing—in which actions by the agent impose an externality on the principal. Not surprisingly, the principal will want to influence the agent's actions. This influence will often take the form of a contract that has the principal compensating the agent contingent on either his actions or the consequences of his actions. Table 1 lists some examples of situations like this. Note that, in many of these examples, the principal is buying a good or service from the agent. That is, many buyer-seller relationships naturally fit into the principal-agent framework. This note covers the basic tools and results of agency theory in this context.

**Table 1:** Examples of Moral-Hazard Problems

| Principal | Agent | Problem | Solution |
|---|---|---|---|
| Employer | Employee | Induce employee to take actions that increase employer's profits, but which he finds personally costly. | Base employee's compensation on employer's profits. |
| Plaintiff | Attorney | Induce attorney to expend costly effort to increase plaintiff's chances of prevailing at trial. | Make attorney's fee contingent on damages awarded plaintiff. |
| Homeowner | Contractor | Induce contractor to complete work (e.g., remodel kitchen) on time. | Give contractor bonus for completing job on time. |
| Landlord | Tenant | Induce tenant to make investments (e.g., in time or money) that preserve or enhance property's value to the landlord. | Pay the tenant a fraction of the increased value (e.g., share-cropping contract). Alternatively, make tenant post deposit to be forfeited if value declines too much. |

To an extent, the principal-agent problem finds its root in the early literature on insurance. There, the concern was that someone who insures an asset might then fail to maintain the asset properly (e.g., park his car in a bad neighborhood). Typically, such behavior was either unobservable by the insurance company or too difficult to contract against directly; hence, the insurance contract could not be directly contingent on such behavior. But because this behavior—known as *moral hazard*—imposes an externality on the insurance company (in this case, a negative one), insurance companies were eager to develop contracts that guarded against it. So, for example, many insurance contracts have *deductibles*—the first $k$ dollars of damage must be paid by the insured rather than the insurance company. Since the insured now has $\$k$ at risk, he'll think twice about parking in a bad neighborhood. That is, the insurance contract is designed to mitigate the externality that the agent—the insured—imposes on the principal—the insurance company. Although principal-agent analysis is more general than this, the name "moral hazard" has stuck and, so, the types of problems considered here are often referred to as moral-hazard problems. A more descriptive name, which is also used in the literature, is *hidden-action problems*.[1]

moral hazard

hidden action

As we've already suggested, the principal-agent model with hidden action has been applied to many questions in economics and other social sciences. Not surprisingly, we will focus on well-known results. Nonetheless, we feel there's some value added to this. First, we will establish some notation and standard reasoning. Second, we will focus on examples, drawn from industrial organization and the theory of the firm, of interest in the study of strategy. Finally, we offer our personal opinions on the achievements and weaknesses of the moral-hazard model.

# 1 The Moral-Hazard Setting

Let us first give a general picture of the situation we wish to analyze in depth.

1. Two players are in an economic relationship characterized by the following two features: First, the actions of one player, *the agent*, affect the well-being of the other player, *the principal*. Second, the players can agree *ex ante* to a reward schedule by which the principal pays the agent.[2] The reward schedule represents an *enforceable* contract (i.e., if there is a dispute about whether a player has lived up to the terms of the contract, then a court or similar body can adjudicate the dispute).

agent

principal

2. The agent's action is *hidden*; that is, he knows what action he has taken but the principal does not directly observe his action. [Although we will consider, as a benchmark, the situation in which the action can be contracted on directly.] Moreover, the agent has complete discretion in choosing his action from some set of feasible actions.[3]

[1] Other surveys include Hart and Holmström (1987), to whom we are greatly indebted. The books by Salanié (1997) and Macho-Stadler and Pérez-Castrillo (1997) also include coverage of this material.

[2] Although we typically think in terms *positive* payments, in many applications payments could be *negative*; that is, the principal fines or otherwise punishes the agent.

[3] Typically, this set is assumed to be exogenous to the relationship. One could imagine,

3. The actions determine, usually stochastically, some *performance measures.* In many models, these are identical to the benefits received by the principal, although in some contexts the two are distinct. The reward schedule is a function of (at least some) of these performance variables. In particular, the reward schedule can be a function of the *verifiable* performance measures (recall that information is verifiable if it can be observed perfectly—without error—by third parties).

4. The *structure* of the situation is common knowledge between the players.

For example, consider a salesperson who has discretion over the amount of time or effort he expends promoting his company's products. Much of these actions are unobservable by his company. The company can, however, measure in a verifiable way the number of orders or revenue he generates. Because these measures are, presumably, correlated with his actions (i.e., the harder he works, the more sales he generates *on average*), it may make sense for the company to base his pay on his sales—put him on commission—to induce him to expend the appropriate level of effort.

Here, we will also be imposing some additional structure on the situation:

- The players are symmetrically informed at the time they agree to a reward schedule.
- Bargaining is take-it-or-leave-it: The principal proposes a contract (reward schedule), which the agent either accepts or rejects. If he rejects it, the game ends and the players receive their *reservation utilities* (their expected utilities from pursuing their next best alternatives). If he accepts, then both parties are bound by the contract.

RESERVATION UTILITIES

- Contracts cannot be renegotiated.
- Once the contract has been agreed to, the only player to take further actions is the agent.
- The game is played once. In particular, there is only period in which the agent takes actions and the agent completes his actions before any performance measures are realized.

All of these are common assumptions and, indeed, might be taken to constitute part of the "standard" principal-agent model.

The link between actions and performance can be seen as follows. Performance is a random variable and its probability distribution depends on the actions taken by the agent. So, for instance, a salesperson's efforts could increase his average (expected) sales, but he still faces upside risk (e.g., an economic boom in his sales region) and downside risk (e.g., introduction of a rival product). Because the performance measure is only stochastically related to the

however, situations in which the principal had some control over the set *ex ante* (e.g., she decided what tools the agent would have available to work with).

action, it is generally impossible to perfectly infer the action from the realization of the performance measure. That is, the performance measure does not, generally, reveal the agent's action—it remains "hidden" despite observing the performance measure.

The link between actions and performance can also be viewed in a indirect way in terms of a *state-space model*. Performance is a function of the agent's actions and of the *state of nature*; that is, a parameter (scalar or vector) that describe the economic environment (e.g., the economic conditions in the salesperson's territory). In this view, the agent takes his action before knowing the state of nature. Typically, we assume that the state of nature is not observable to the principal. If it she could observe it, then she could perfectly infer the agent's action by inverting from realized performance. In this model, it is not important whether the agent later observes the state of nature or not, given he could deduce it from his observation of his performance and his knowledge of his actions.

STATE-SPACE MODEL

There is a strong assumption of physical causality in this setting, namely that actions by the agent determine performances. Moreover, the process is viewed as a static production process: There are neither dynamics nor feedback. In particular, the contract governs one period of production and the game between principal and agent encompasses only this period. In addition, when choosing his actions, the agent's information is identical to the principal's. Specifically, he can*not* adjust his actions as the performance measures are realized. The sequentiality between actions and performance is strict: First actions are completed and, only then, is performance realized. This strict sequentiality is quite restrictive, but relaxing the model to allow for less rigid a timing introduces dynamic issues that are far more complex to solve. We will simply mention some of them in passing here.

# 2 Basic Two-Action Model

We start with the simplest principal-agent model. Admittedly, it is so simple that a number of the issues one would like to understand about contracting under moral hazard disappear. On the other hand, many issues remain and, for pedagogical purposes at least, it is a good place to start.[4]

## 2.1 The two-action model

Consider a salesperson, who will be the agent in this model and who works for a manufacturer, the principal. The manufacturer's problem is to design incentives for the salesperson to expend effort promoting the manufacturer's product to consumers.

[4]But the pedagogical value of this section should not lead us to forget caution. And caution is indeed necessary as the model oversimplifies reality to a point that it delivers conclusions that have no match in a more general framework. One could say that the two-action model is tailored so as to fit with naive intuition and to lead to the desired results without allowing us to see fully the (implicit) assumptions on which we are relying.

Let $x$ denote the level of sales that the salesperson reaches within the period under consideration. This level of sales depends upon lots of demand parameters that are beyond the salesperson's control; but, critically, they also depend upon the salesperson's efforts—the more effort the salesperson expends, the more consumers will buy in expectation. Specifically, suppose that when the salesperson does not expend effort, sales are distributed according to distribution function $F_0(\cdot)$ on $\mathbb{R}_+$.[5] When he does expend effort, sales are distributed $F_1(\cdot)$. Observe effort, here, is a binary choice. Consistent with the story we've told so far, we want sales to be greater, in expectation, if the salesperson has become informed; that is, we assume[6]

$$\mathbb{E}_1[x] \equiv \int_0^{+\infty} x dF_0(x) > \int_0^{+\infty} x dF_1(x) dx \equiv \mathbb{E}_0[x]. \tag{1}$$

Having the salesperson expend effort is sales-enhancing, but it is also costly for the salesperson. Expending effort causes him disutility $C$ compared to no effort. The salesperson has discretion: He can incur a personal cost of $C$ and boosts sales by choosing action $a = 1$ (expending effort), or he can expend no effort, action $a = 0$, which causes him no disutility but does not stimulate demand either. Like most individuals, the salesperson is sensitive to variations of his revenue, so that his preferences over income exhibit risk aversion. We also assume that his utility exhibits additive separability in money and action. Specifically, let his utility be

$$U(s, x, a) = u(s) - aC,$$

where $s$ is a payment from the manufacturer and $u(\cdot)$ is strictly increasing and concave (he prefers more money to less and is risk averse). Of course, the salesperson could also simply choose not to work for the manufacturer. This would yield him an expected level of utility equal to $U_R$. The quantity $U_R$ is the salesperson's *reservation utility*. RESERVATION UTILITY

The manufacturer is a large risk-neutral company that cares about the sales realized on the local market net of the salesperson's remuneration or share of the sales. Hence, its preferences are captured by the utility (profit) function:

$$W(s, x, a) = x - s.$$

Assume the manufacturer's size yields it all the bargaining power in its negotiations with the salesperson.

Suppose, as a benchmark, that the manufacturer could observe and establish whether the salesperson had expended effort. We will refer to this benchmark as the *full or perfect information case*. Then the manufacturer could use a contract that is contingent on the salesperson's effort, $a$. Moreover, because FULL-INFORMATION CASE

[5] Treating the set of "possible" sales as $[0, \infty)$ is without loss of generality, since the boundedness of sales can be captured by assuming $F(x \geq \bar{x}) = 0$ for some $\bar{x} < \infty$.

[6] Observe that this notation covers both the case in which $F_a(\cdot)$ is a differentiable distribution function or a discrete distribution function.

the salesperson is risk averse, while the manufacturer is risk neutral, it is most efficient for the manufacturer to absorb all risk. Hence, in this benchmark case, the salesperson's compensation would *not* depend on the realization of sales, but only on the salesperson's effort. The contract would then be of the form:

$$s = \begin{cases} s_0 \text{ if } a = 0 \\ s_1 \text{ if } a = 1 \end{cases} .$$

If the manufacturer wants the salesperson to expend effort (i.e., choose $a = 1$), then it must choose $s_0$ and $s_1$ to satisfy two conditions. First, conditional on accepting the contract, the salesperson must prefer to invest; that is,

$$u(s_1) - C \geq u(s_0) . \tag{IC}$$

A constraint like this is known as an *incentive compatibility* constraint (conventionally abbreviated IC): Taking the desired action must maximize the agent's expected utility. Second, conditional on the fact that he will be induced to expend effort, he must prefer to sign the contract than to forgo employment with the manufacturer; that is,

INCENTIVE COMPATIBILITY

$$u(s_1) - C \geq U_R. \tag{IR}$$

A constraint like this is known as an *individual rationality* constraint (conventionally abbreviated IR). The IR constraint is also referred to as a *participation constraint*. Moreover, in selecting $s_0$ and $s_1$, the manufacturer wants to maximize its profits conditional on gaining acceptance of the contract and inducing $a = 1$. That is, it wishes to solve

INDIVIDUAL RATIONALITY

PARTICIPATION CONSTRAINT

$$\max_{s_0, s_1} \mathbb{E}_1[x] - s_1$$

subject to the constraints (IC) and (IR). If we postulate that

$$\begin{aligned} u(s) &< U_R \text{ and} \\ u(s) - C &= U_R \end{aligned} \tag{2}$$

both have solutions within the domain of $u(\cdot)$, then the solution to the manufacturer's problem is straightforward: $s_1$ solves (2) and $s_0$ is a solution to $u(s) < U_R$. It is readily seen that this solution satisfies the constraints. Moreover, since $u(\cdot)$ is strictly increasing, there is no smaller payment that the manufacturer could give the salesperson and still have him accept the contract; that is, the manufacturer cannot increase profits relative to paying $s_1$ by making a smaller payment. This contract is known as a *forcing contract*. For future reference, let $s_1^F$ be the solution to (2). Observe that $s_1^F = u^{-1}(U_R + C)$, where $u^{-1}(\cdot)$ is the inverse function corresponding to $u(\cdot)$.

FORCING CONTRACT

**Technical Aside**

The solution to the manufacturer's maximization problem depends on the domain and range of the utility function $u(\cdot)$. Let $\mathcal{D}$, an interval in $\mathbb{R}$, be the domain and $\mathcal{R}$ be the range. Let $\underline{s}$ be $\inf \mathcal{D}$ (i.e., the greatest lower bound of $\mathcal{D}$) and let $\bar{s}$ be $\sup \mathcal{D}$ (i.e., the least upper bound of $\mathcal{D}$). As shorthand for $\lim_{s \downarrow \underline{s}} u(s)$ and $\lim_{s \uparrow \bar{s}} u(s)$, we'll write $u(\underline{s})$ and $u(\bar{s})$ respectively. If $u(s) - C < U_R$ for all $s \leq \bar{s}$, then *no* contract exists that satisfies (IR). In this case, the best the manufacturer could hope to do is implement $a = 0$. Similarly, if $u(\bar{s}) - C < u(\underline{s})$, then *no* contract exists that satisfies (IC). The manufacturer would have to be satisfied with implementing $a = 0$. Hence, $a = 1$ can be implemented if and only if $u(\bar{s}) - C > \max\{U_R, u(\underline{s})\}$. Assuming this condition is met, a solution is $s_0 \downarrow \underline{s}$ and $s_1$ solving

$$u(s) - C \geq \max\{U_R, u(\underline{s})\}.$$

Generally, conditions are imposed on $u(\cdot)$ such that a solution exists to $u(s) < U_R$ and (2). Henceforth, we will assume that these conditions have, indeed, been imposed. For an example of an analysis that considers bounds on $\mathcal{D}$ that are more binding, see Sappington (1983).

Another option for the manufacturer is, of course, not to bother inducing the salesperson to expend effort promoting the product. There are many contracts that would accomplish this goal, although the most "natural" is perhaps a *non-contingent* contract: $s_0 = s_1$. Given that the manufacturer doesn't seek to induce investment, there is no IC constraint—the salesperson inherently prefers not to invest—and the only constraint is the IR constraint:

$$u(s_0) \geq U_R.$$

The expected-profit-maximizing (cost-minimizing) payment is then the smallest payment satisfying this expression. Since $u(\cdot)$ is increasing, this entails $s_0 = u^{-1}(U_R)$. We will refer to *this* value of $s_0$ as $s_0^F$.

The manufacturer's expected profit conditional on inducing $a$ under the *optimal* contract for inducing $a$ is $\mathbb{E}_a[x] - s_a^F$. The manufacturer will, thus, prefer to induce $a = 1$ if

$$\mathbb{E}_1[x] - s_1^F > \mathbb{E}_0[x] - s_0^F.$$

In what follows, we will assume that this condition is met: That is, in our benchmark case of verifiable action, the manufacturer prefers to induce effort than not to.

Observe the steps taken in solving this benchmark case: First, for each possible action we solved for the optimal contract that induces that action. Then we calculated the expected profits for each possible action assuming the optimal contract. The action that is induced is, then, the one that yields the largest expected profit. This two-step process for solving for the optimal contract is frequently used in contract theory, as we will see.

## 2.2 The optimal incentives contract

Now, we return to the case of interest: The salesperson's (agent's) action is hidden. Consequently, the manufacturer cannot make its payment contingent on whether the salesperson expends effort. The one remaining verifiable variable is performance, as reflected by realized sales, $x$. A contract, then, is function mapping sales into compensation for the salesperson: $s = S(x)$. Facing such a contract, the salesperson then freely chooses the action that maximizes his expected utility. Consequently, the salesperson chooses action $a = 1$ if and only if:[7]

$$\mathbb{E}_1 [u(S(x))] - C \geq \mathbb{E}_0 [u(S(x))] \,. \tag{IC$'$}$$

If this inequality is violated, the salesperson will simply choose not to get informed. Observe that this is the incentive compatibility (IC) constraint in this case.

The game we analyze is in fact a simple Stackelberg game, where the manufacturer is the first mover—it chooses the payment schedule—to which it is committed; and the salesperson is the second mover—choosing his action in response to the payment schedule. The solution to

$$\max_a \mathbb{E}_a [u(S(x))] - aC$$

(with ties going to the manufacturer—see footnote 7) gives the salesperson's equilibrium choice of action by the agent as a *function* of the payment function $S(\cdot)$. Solving this contracting problem then requires us to understand what kind of contract the manufacturer could and will offer.

Observe first that if she were to offer the fixed-payment contract $S(x) = s_0^F$ for all $x$, then, as above, the agent would accept the contract and not bother acquiring information. Among all contracts that induce the agent to choose action $a = 0$ in equilibrium, this is clearly the cheapest one for the manufacturer. The fixed-payment contract set at $s_1^F$ will, however, no longer work given the hidden-action problem: Since the salesperson gains $s_1^F$ whatever his efforts, he will choose the action that has lesser cost for him, $a = 0$. It is in fact immediate that any fixed-payment contract, which would be optimal if the only concern were efficient risk-sharing, will induce an agent to choose his least costly action. Given that it is desirable, at least in the benchmark full-information case, for the salesperson to expend effort selling the product, it seems plausible that the manufacturer will try to induce effort even though—as we've just seen—that must entail the *in*efficient (relative to the first best) allocation of risk to the salesperson.

[7] We assume, when indifferent among a group of actions, that the agent chooses from that group the action that the principal prefers. This assumption, although often troubling to those new to agency theory, is not truly a problem. Recall that the agency problem is a game. Consistent with game theory, we're looking for an equilibrium of this game; i.e., a situation in which players are playing mutual best responses and in which they correctly anticipate the best responses of their opponents. Were the agent to behave differently when indifferent, then we wouldn't have an equilibrium because the principal would vary her strategy—offer a different contract—so as to break this indifference. Moreover, it can be shown that in many models the *only* equilibrium has the property that the agent chooses among his best responses (the actions among which he is indifferent given the contract) the one most preferred by the principal.

We now face two separate questions. First, conditional on the manufacturer wanting to cause the salesperson to expend effort, what is the optimal—least-expected-cost—contract for the manufacturer to offer? Second, are the manufacturer's expected profits greater doing this than not inducing the salesperson to expend effort (i.e., greater than the expected profits from offering the fixed-payment contract $S(x) = s_0^F$)?

As in the benchmark case, not only must the contract give the salesperson an incentive to acquire information (i.e., meet the IC constraint), it must also be individually rational:

$$\mathbb{E}_1 [u(S(x))] - C \geq U_R. \tag{IR$'$}$$

The optimal contract is then the solution to the following program:

$$\max_{S(\cdot)} \mathbb{E}_1 [x - S(x)] \tag{3}$$
$$\text{subject to (IC}') \text{ and (IR}').$$

The next few sections will consider the solution to (3) under a number of different assumptions about the distribution functions $F_a(\cdot)$.

Two assumptions on $u(\cdot)$, in addition to those already given, that will be common to these analyses are:

1. The domain of $u(\cdot)$ is $(\underline{s}, \infty)$, $\underline{s} \geq -\infty$.[8]
2. $\lim_{s \downarrow \underline{s}} u(s) = -\infty$ and $\lim_{s \uparrow \infty} u(s) = \infty$.

An example of a function satisfying *all* the assumptions on $u(\cdot)$ is $\ln(\cdot)$ with $\underline{s} = 0$. For the most part, these assumptions are for convenience and are more restrictive than we need (for instance, many of our results will also hold if $u(s) = \sqrt{x}$, although this fails the second assumption). Some consequences of these and our earlier assumptions are

- $u(\cdot)$ is invertible (a consequence of its strict monotonicity). Let $u^{-1}(\cdot)$ denote its inverse.
- The domain of $u^{-1}(\cdot)$ is $\mathbb{R}$ (a consequence of the last two assumptions).
- $u^{-1}(\cdot)$ is continuous, strictly increasing, and convex (a consequence of the continuity of $u(\cdot)$, its concavity, and that it is strictly increasing).

## 2.3 Two-outcome model

Imagine that there are only two possible realizations of sales, high and low, denoted respectively by $x_H$ and $x_L$, with $x_H > x_L$. Assume that

$$F_a(x_L) = k - qa,$$

[8] Since its domain is an open interval and it is concave, $u(\cdot)$ is continuous everywhere on its domain (see van Tiel, 1984, p. 5).

where $q \in (0,1]$ and $k \in [q,1]$ are known constants.

A contract is $s_H = S(x_H)$ and $s_L = S(x_L)$. We can, thus, write program (3) as

$$\max_{s_H, s_L} (1+q-k)(x_H - s_H) + (k-q)(x_L - s_L)$$

subject to

$$(1+q-k)\,u(s_H) + (k-q)\,u(s_L) - C \geq (1-k)\,u(s_H) + ku(s_L) \qquad \text{(IC)}$$

and

$$(1+q-k)\,u(s_H) + (k-q)\,u(s_L) - C \geq U_R \qquad \text{(IR)}$$

We could solve this problem mechanically using the usual techniques for maximizing a function subject to constraints, but it is far easier, here, to use a little intuition. To begin, we need to determine which constraints are binding. Is IC binding? Well, suppose it were not. Then the problem would simply be one of optimal risk sharing, because, by supposition, the incentive problem no longer binds. But we know optimal risk sharing entails $s_H = s_L$; that is, a fixed-payment contract.[9] As we saw above, however, a fixed-payment contract can*not* satisfy IC:

$$u(s) - C < u(s).$$

Hence, IC must be binding.

What about IR? Is it binding? Suppose it were not (i.e., it were a strict inequality) and let $s_L^*$ and $s_H^*$ be the optimal contract. Then there must exist an $\varepsilon > 0$ such that

$$(1+q-k)\left[u(s_H^*) - \varepsilon\right] + (k-q)\left[u(s_L^*) - \varepsilon\right] - C \geq U_R.$$

Let $\tilde{s}_n = u^{-1}\left[u(s_n^*) - \varepsilon\right]$, $n \in \{L, H\}$. Clearly, $\tilde{s}_n < s_n^*$ for both $n$, so that the $\{\tilde{s}_n\}$ contract costs the manufacturer less than the $\{s_n^*\}$ contract; or, equivalently, the $\{\tilde{s}_n\}$ contract yields the manufacturer greater expected profits than the $\{s_n^*\}$ contract. Moreover, the $\{\tilde{s}_n\}$ contract satisfies IC:

$$\begin{aligned}
& (1+q-k)\,u(\tilde{s}_H) + (k-q)\,u(\tilde{s}_L) - C \\
\geq\; & (1+q-k)\left[u(s_H^*) - \varepsilon\right] + (k-q)\left[u(s_L^*) - \varepsilon\right] - C \\
=\; & (1+q-k)\,u(s_H^*) + (k-q)\,u(s_L^*) - C - \varepsilon \\
\geq\; & (1-k)\,u(s_H^*) + ku(s_L^*) - \varepsilon \text{ (since } \{s_n^*\} \text{ must satisfy IC)} \\
=\; & (1-k)\,u(\tilde{s}_H) + ku(\tilde{s}_L).
\end{aligned}$$

[9] The proof is straightforward if $u(\cdot)$ is differentiable: Let $\lambda$ be the Lagrange multiplier on (IR). The first-order conditions with respect to $s_L$ and $s_H$ are

$$k - q - \lambda(k-q)\,u'(s_L) = 0$$

and

$$1 + q - k - \lambda(1+q-k)\,u'(s_H) = 0,$$

respectively. Solving, it is clear that $s_L = s_H$. The proof when $u(\cdot)$ is not (everywhere) differentiable is only slightly harder and is left to the reader.

But this means that $\{\tilde{s}_n\}$ satisfies *both* constraints *and* yields greater expected profits, which contradicts the optimality of $\{s_n^*\}$. Therefore, by contradiction, we may conclude that IR is also binding at the optimal contract for inducing $a = 1$.

We're now in a situation where the two constraints must bind at the optimal contract. But, given we have only two unknown variables, $s_H$ and $s_L$, this means we can solve for the optimal contract merely by solving the constraints. Doing so yields

$$\hat{s}_H = u^{-1}\left(U_R + \frac{k}{q}C\right) \text{ and } \hat{s}_L = u^{-1}\left(U_R - \frac{1-k}{q}C\right). \tag{4}$$

Observe that the payments *vary* with the state (as we knew they must because fixed payments fail the IC constraint).

Recall that *were a* verifiable, the contract would be $S(x) = s_1^F = u^{-1}(U_R + C)$. Rewriting (4) we see that

$$\hat{s}_H = u^{-1}\left(u\left(s_1^F\right) + \frac{k-q}{q}C\right) \text{ and } \hat{s}_L = u^{-1}\left(u\left(s_1^F\right) - \frac{1+q-k}{q}C\right);$$

that is, one payment is *above* the payment under full information, while the other is *below* the payment under full information. Moreover, the *expected* payment to the salesperson is greater than $s_1^F$:

$$\begin{aligned}
& (1+q-k)\,\hat{s}_H + (k-q)\,\hat{s}_L \\
= \; & (1+q-k)\,u^{-1}\left(u\left(s_1^F\right) + \frac{k-q}{q}C\right) \\
& + (k-q)\,u^{-1}\left(u\left(s_1^F\right) - \frac{1+q-k}{q}C\right) \\
\geq \; & u^{-1}\left[\begin{array}{c} (1+q-k)\left(u\left(s_1^F\right) + \frac{k-q}{q}C\right) \\ + (k-q)\left(u\left(s_1^F\right) - \frac{1+q-k}{q}C\right) \end{array}\right] \\
= \; & u^{-1}\left[u\left(s_1^F\right)\right] = s_1^F;
\end{aligned} \tag{5}$$

where the inequality follows from Jensen's inequality.[10] Provided the agent is strictly risk averse and $q < k$, the above inequality is strict: Inducing the agent to choose $a = 1$ costs strictly more in expectation when the principal cannot verify the agent's action.

Before proceeding, it is worth considering why the manufacturer (principal) suffers from its inability to verify the salesperson's (agent's) action (i.e.,

[10] Jensen's inequality for convex functions states that if $g(\cdot)$ is convex function on an interval of $\mathbb{R}$ (and equal to zero not on that interval), then $\mathbb{E}\{g(X)\} \geq g(\mathbb{E}X)$, where $X$ is a random variable and $\mathbb{E}$ is the expectations operator with respect to $X$ (see, e.g., van Tiel, 1984, p. 11, for a proof). If $g(\cdot)$ is strictly convex and the distribution of $X$ is not degenerate (i.e., does not concentrate all mass on one point), then the inequality is strict. For *concave* functions, the inequalities are reversed.

from the existence of a hidden-action problem). *Ceteris paribus*, the salesperson prefers $a = 0$ to $a = 1$, because expending effort is personally costly to him. Hence, when the manufacturer wishes to induce $a = 1$, its interests and the salesperson's are not aligned. To align their interests, the manufacturer must offer the salesperson incentives to choose $a = 1$. The problem is that the manufacturer cannot directly tie these incentives to the variable in which it is interested, namely the action itself. Rather, it must tie these incentives to sales, which are imperfectly correlated with action (provided $q < k$). These incentives, therefore, expose the agent to risk. We know, relative to the first best, that this is inefficient. Someone must bear the cost of this inefficiency. Because the bargaining game always yields the salesperson the same expected utility (i.e., IR is always binding), the cost of this inefficiency must, thus, be borne by the manufacturer.

Another way to view this last point is that because the agent is exposed to risk, which he dislikes, he must be compensated. This compensation takes the form of a higher *expected* payment.

To begin to appreciate the importance of the hidden-action problem, observe that

$$\begin{aligned} \lim_{q \uparrow k} (1 + q - k)\, \hat{s}_H + (k - q)\, \hat{s}_L &= \lim_{q \uparrow k} \hat{s}_H \\ &= u^{-1}\left[u\left(s_1^F\right)\right] \\ &= s_1^F. \end{aligned}$$

Hence, when $q = k$, there is effectively no hidden-action problem: Low sales, $x_L$, constitute proof that the salesperson failed to invest, because $\Pr\{x = x_L | a = 1\} = 0$ in that case. The manufacturer is, thus, free to punish the salesperson for low sales in whatever manner it sees fit; thereby deterring $a = 0$. But because there is no risk when $a = 1$, the manufacturer does not have to compensate the salesperson for bearing risk and can, thus, satisfy the IR constraint paying the same compensation as under full information. When $q = k$, we have what is known as a *shifting support*.[11] We will consider shifting supports in greater depth later. shifting support

[11]The *support* of distribution $G$ over random variable $X$, sometimes denoted $\text{supp}\{X\}$, is the set of $x$'s such that for all $\varepsilon > 0$,

$$G(x) - G(x - \varepsilon) > 0.$$

Loosely speaking, it is the set of $x$'s that have positive probability of occurring. For example, if $X$ is the outcome of the roll of a die, then $5 \in \text{supp}\{X\}$, since

$$G(5) - G(5 - \varepsilon) \geq \frac{1}{6} \quad \forall \varepsilon > 0.$$

Whereas $5.5 \notin \text{supp}\{X\}$, since

$$G(5.5) - G(5.5 - \varepsilon) = 0 \text{ if } \varepsilon \leq .5.$$

Likewise, if $X$ is uniformly distributed on $[0, 1]$, then any $x \in (0, 1]$ is in $\text{supp}\{X\}$, since

$$G(x) - G(x - \varepsilon) = \min\{x, \varepsilon\}.$$

To see the importance of the salesperson's risk aversion, note that *were* the salesperson risk neutral, then the inequality in (5) would, instead, be an equality and the expected wage paid the salesperson would equal the wage paid under full information. Given that the manufacturer is risk *neutral* by assumption, it would be indifferent between an expected wage of $s_1^F$ and paying $s_1^F$ with certainty: There would be no loss, relative to full information, of overcoming the hidden-action problem by basing compensation on sales. It is important to note, however, that assuming a risk-neutral agent does *not* obviate the need to pay contingent compensation (e.g., we still need $s_H > s_L$)—as can be seen by checking the IC constraint; agent risk neutrality only means that the principal suffers no loss from the fact that the agent's action is hidden.

We can also analyze this version of the model graphically. A graphical treatment is facilitated by switching from compensation space to utility space; that is, rather than put $s_L$ and $s_H$ on the axes, we put $u_L \equiv u(s_L)$ and $u_H \equiv u(s_H)$ on the axes. With this change of variables, program (3) becomes:

$$\max_{u_L, u_H} (1+q-k)\left(x_H - u^{-1}(u_H)\right) + (k-q)\left(x_L - u^{-1}(u_L)\right)$$

subject to

$$(1+q-k)\,u_H + (k-q)\,u_L - C \geq (1-k)\,u_H + k u_L \text{ and} \qquad (\text{IC}'')$$

$$(1+q-k)\,u_H + (k-q)\,u_L - C \geq U_R \qquad (\text{IR}'')$$

Observe that, in this space, the salesperson's indifference curves are straight lines, with lines farther from the origin corresponding to greater expected utility. The manufacturer's iso-expected-profit curves are concave relative to the origin, with curves closer to the origin corresponding to greater expected profit. Figure 1 illustrates. Note that in Figure 1 the salesperson's indifference curves and the manufacturer's iso-expected-profit curves are tangent only at the 45° line, a well-known result from the insurance literature.[12] This shows, graphically, why efficiency (in a first-best sense) requires that the agent not bear risk.

We can re-express (IC″) as

$$u_H \geq u_L + \frac{C}{q}. \qquad (6)$$

Hence, the set of contracts that are incentive compatible lie on or above a line above, but parallel, to the 45° line. Graphically, we now see that an incentive-compatible contract requires that we abandon non-contingent contracts. Figure 2 shows the space of incentive-compatible contracts.

[12] **Proof:** Let $\phi(\cdot) = u^{-1}(\cdot)$. Then the MRS for the manufacturer is

$$-\frac{(k-q)\,\phi'(u_L)}{(1+q-k)\,\phi'(u_H)};$$

whereas the MRS for the salesperson is

$$-\frac{(k-q)}{(1+q-k)}.$$

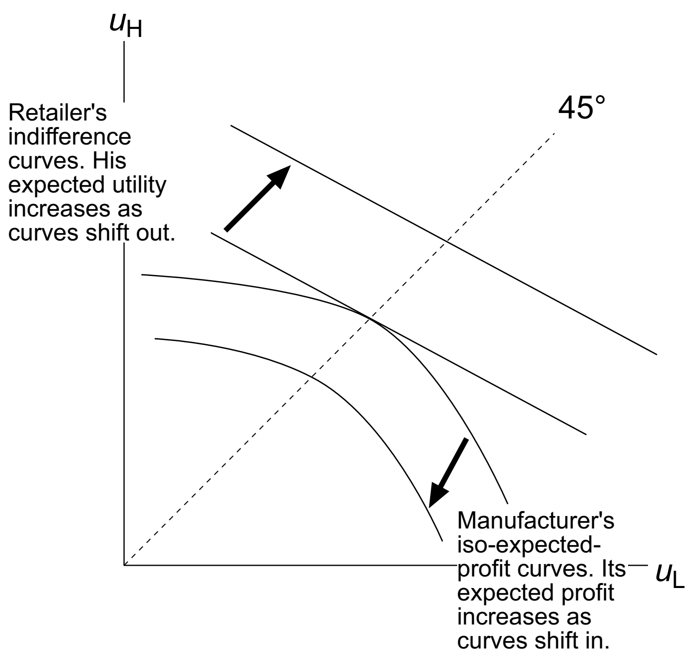


**Figure 1:** Indifference curves in utility space for the manufacturer (principal) and salesperson (agent).

The set of individually rational contracts are those that lie on or above the line defined by (IR$''$). This is also illustrated in Figure 2. The intersection of these two regions then constitutes the set of feasible contracts for inducing the salesperson to choose $a = 1$. Observe that the lowest iso-expected-profit curve—corresponding to the largest expected profit—that intersects this set is the one that passes through the "corner" of the set—consistent with our earlier conclusion that *both* constraints are binding at the optimal contract.

Lastly, let's consider the variable $q$. We can interpret $q$ as representing the correlation—or, more accurately, the informativeness—of sales to the action taken.[13] At first glance, it might seem odd to be worried about the informa-

Since $\phi(\cdot)$ is strictly convex, $\phi'(u) = \phi'(v)$ only if $u = v$; that is, the MRS's of the appropriate iso-curves can be tangent only on the 45° line.

[13]To see this, in a loose sense, suppose the salesperson is playing the mixed strategy in which he chooses $a = 1$ with probability $\beta \in (0, 1)$. Then, *ex post*, the probability that he chose $a = 1$ conditional on $x = x_H$ is

$$\frac{(1 + q - k)\,\beta}{(1 + q - k)\,\beta + (1 - k)\,(1 - \beta)}$$

by Bayes Theorem. It is readily seen that this probability is increasing in $q$; moreover, it reduces to $\beta$—nothing has been learned—if $q = 0$. Similarly, the posterior probability of

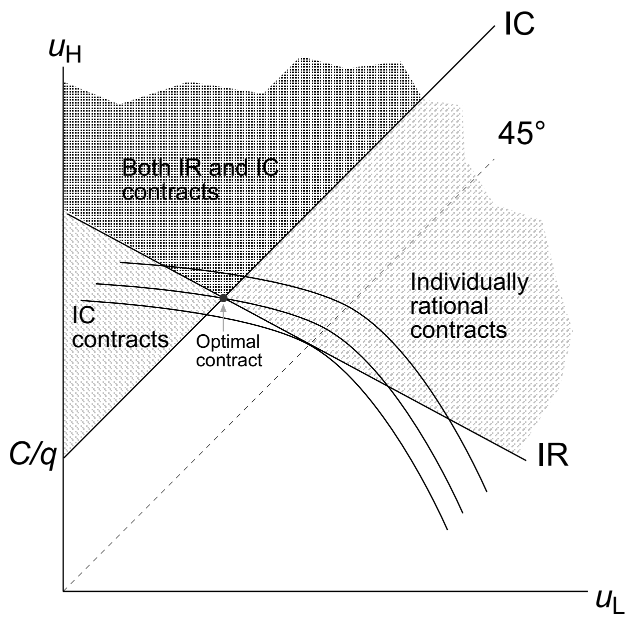


**Figure 2:** The set of feasible contracts.

tiveness of sales since, in equilibrium, the principal can accurately predict the agent's choice of action from the structure of the game and her knowledge of the contract. But that's not the point: The principal is forced to design a contract that pays the agent based on performance measures that are informative about the variable upon which she would truly like to contract, namely his action. The more informative these performance measures are—loosely, the more correlated they are with action—the closer the principal is getting to the ideal of contracting on the agent's action.

In light of this discussion it wouldn't be surprising if the manufacturer's expected profit under the optimal contract for inducing $a = 1$ increases as $q$ increases. Clearly its expected *revenue*,

$$(1 + q - k)\, x_H + (k - q)\, x_L,$$

is increasing in $q$. Hence, it is sufficient to show merely that its expected *cost*,

$$(1 + q - k)\, \hat{s}_H + (k - q)\, \hat{s}_L,$$

$a = 1$ conditional on $x = x_L$ is

$$\frac{(k - q)\, \beta}{(k - q)\, \beta + k\,(1 - \beta)}.$$

It is readily seen that this is decreasing in $q$; moreover, it reduces to $\beta$—nothing has been learned—if $q = 0$.

is non-increasing in $q$. To do so, it is convenient to work in terms of utility (i.e., $u_H$ and $u_L$) rather than directly with compensation. Let $q_1$ and $q_2$ be two distinct values of $q$, with $q_1 < q_2$. Let $\left\{u_n^1\right\}$ be the optimal contract (expressed in utility terms) when $q = q_1$.

**Lemma 1** *There exist $a \in (0,1)$ and $b \in (0,1)$ such that*

$$\begin{aligned} a(1+q_2-k)+b(k-q_2) &= 1+q_1-k; \\ (1-a)(1+q_2-k)+(1-b)(k-q_2) &= k-q_1; \\ a(1-k)+bk &= 1-k; \text{ and} \\ (1-a)(1-k)+(1-b)k &= k. \end{aligned}$$

---

**Proof:** Let

$$a = \frac{q_1 k + q_2(1-k)}{q_2} \text{ and } b = \frac{(1-k)(q_2-q_1)}{q_2}$$

Simple, albeit tedious, algebra confirms that $a$ and $b$ solve the above equations. Clearly, since $q_2 > q_1$, $a \in (0,1)$. To see that $b$ is also, observe that

$$(1-k)(q_2-q_1) < q_2 - q_1 < q_2.$$

■

---

In light of this result, define

$$\begin{aligned} \tilde{u}_H &= a u_H^1 + (1-a) u_L^1 \text{ and} \\ \tilde{u}_L &= b u_H^1 + (1-b) u_L^1, \end{aligned}$$

where $a$ and $b$ are defined by Lemma 1. Let us now see that $\{\tilde{u}_n\}$ satisfies both the IR and IC constraints when $q = q_2$. Observe first that

$$\begin{aligned} & (1+q_2-k)\tilde{u}_H + (k-q_2)\tilde{u}_L \\ = \; & [a(1+q_2-k)+b(k-q_2)]u_H^1 + [(1-a)(1+q_2-k)+(1-b)(k-q_2)]u_L^1 \\ = \; & (1+q_1-k)u_H^1 + (k-q_1)u_L^1 \end{aligned}$$

and

$$\begin{aligned} & (1-k)\tilde{u}_H + k\tilde{u}_L \\ = \; & [a(1-k)+bk]u_H^1 + [(1-a)(1-k)+(1-b)k]u_L^1 \\ = \; & (1-k)u_H^1 + k u_L^1. \end{aligned}$$

Since $\left\{u_n^1\right\}$ satisfies IR and IC when $q = q_1$, it follows, by transitivity, that $\{\tilde{u}_n\}$ solves IR and IC when $q = q_2$.[14] We need, now, simply show that

$$\begin{aligned} & (1+q_2-k)u^{-1}(\tilde{u}_H) + (k-q_2)u^{-1}(\tilde{u}_L) \\ \leq \; & (1+q_1-k)u^{-1}\left(u_H^1\right) + (k-q_1)u^{-1}\left(u_L^1\right). \end{aligned}$$

[14]Note, although it isn't necessary for our analysis, that since we know in this setting that the contract that solves both IR and IC is the optimal contract for inducing $a = 1$, we've just shown that $\{\tilde{u}_n\}$ is the optimal contract when $q = q_2$.

To do this, observe that

$$\begin{aligned} u^{-1}\left(\tilde{u}_H\right) &\leq au^{-1}\left(u_H^1\right)+(1-a)\,u^{-1}\left(u_L^1\right) \text{ and} \\ u^{-1}\left(\tilde{u}_L\right) &\leq bu^{-1}\left(u_H^1\right)+(1-b)\,u^{-1}\left(u_L^1\right), \end{aligned}$$

by Jensen's inequality (recall $u^{-1}(\cdot)$ is convex). Hence we have

$$\begin{aligned} & (1+q_2-k)\,u^{-1}\left(\tilde{u}_H\right)+(k-q_2)\,u^{-1}\left(\tilde{u}_L\right) \\ \leq\; & (1+q_2-k)\left[au^{-1}\left(u_H^1\right)+(1-a)\,u^{-1}\left(u_L^1\right)\right] \\ & +(k-q_2)\left[bu^{-1}\left(u_H^1\right)+(1-b)\,u^{-1}\left(u_L^1\right)\right] \\ =\; & \left[a\,(1-k)+bk\right]u^{-1}\left(u_H^1\right)+\left[(1-a)\,(1-k)+(1-b)\,k\right]u^{-1}\left(u_L^1\right) \\ =\; & (1-q_1-k)\,u^{-1}\left(u_H^1\right)+(k-q_1)\,u^{-1}\left(u_L^1\right) \end{aligned}$$

—the manufacturer's expected cost is no greater when $q=q_2$ as when $q=q_1$—as was to be shown.

**Summary:** Although we've considered the simplest of agency models in this section, there are, nevertheless, some general lessons that come from this. First, the optimal contract for inducing an action other than the action that the agent finds least costly requires a contract that is fully contingent on the performance measure. This is a consequence of the action being unobservable to the principal, not the agent's risk aversion. When, however, the agent is risk averse, then the principal's expected cost of solving the hidden-action problem is greater than it would be in the benchmark full-information case: Exposing the agent to risk is inefficient (relative to the first best) and the cost of this inefficiency is borne by the principal. The size of this cost depends on how good an approximation the performance measure is for the variable upon which the principal really desires to contract, the agent's action. The better an approximation (statistic) it is, the lower is the principal's expected cost. If, as here, that shift also raises expected revenue, then a more accurate approximation means greater expected profits. It is also worth pointing out that one result, which might seem as though it should be general, is not: Namely, the result that compensation is increasing with performance (e.g., $\hat{s}_H > \hat{s}_L$). Although this is true when there are only two possible realizations of the performance measure (as we've proved), this result does not hold generally when there are more than two possible realizations.

## 2.4 Multiple-outcomes model

Now we assume that there are multiple possible outcomes, including, possibly, an infinite number. Without loss of generality, we may assume the set of possible sales levels is $(0,\infty)$, given that impossible levels in this range can be assigned zero probability. We will also assume, henceforth, that $u(\cdot)$ exhibits strict risk aversion (i.e., is strictly concave).

Recall that our problem is to solve program (3), on page 9. In this context, we can rewrite the problem as

$$\max_{S(\cdot)} \int_0^\infty (x - S(x))\, dF_1(x)$$

subject to

$$\begin{aligned} \int_0^\infty u[S(x)]\, dF_1(x) - C &\geq U_R \text{ and} \quad (7) \\ \int_0^\infty u[S(x)]\, dF_1(x) - C &\geq \int_0^\infty u[S(x)]\, dF_0(x), \quad (8) \end{aligned}$$

which are the IR and IC constraints, respectively. In what follows, we assume that there is a well-defined density function, $f_a(\cdot)$, associated with $F_a(\cdot)$ for both $a$. For instance, if $F_a(\cdot)$ is differentiable everywhere, then $f_a(\cdot) = F'_a(\cdot)$ and the $\int dF_a(x)$ notation could be replaced with $\int f_a(x)\, dx$. Alternatively, the possible outcomes could be discrete, $x = x_1, \dots, x_N$, in which case

$$f_a(\hat{x}) = F_a(\hat{x}) - \lim_{x \uparrow \hat{x}} F_a(x)$$

and the $\int dF_a(x)$ notation could be replaced with $\sum_{n=1}^N f_a(x_n)$.

We solve the above program using standard Kuhn-Tucker techniques. Let $\mu$ be the (non-negative) Lagrange multiplier on the incentive constraint and let $\lambda$ be the (non-negative) Lagrange multiplier on the individual rationality constraint. The Lagrangian of the problem is, thus,

$$\begin{aligned} \mathcal{L}(S(\cdot), \lambda, \mu) &= \int_0^{+\infty} [x - S(x)]\, dF_1(x) + \lambda \left( \int_0^{+\infty} u(S(x)) dF_1(x) - C \right) \\ &\quad + \mu \left( \int_0^{+\infty} u(S(x)) dF_1(x) - \int_0^{+\infty} u(S(x)) dF_0(x) - C \right). \end{aligned}$$

The necessary first-order conditions are $\lambda \geq 0$, $\mu \geq 0$, (7), (8),

$$u'[S(x)] \left( \lambda + \mu \left[ 1 - \frac{f_0(x)}{f_1(x)} \right] \right) - 1 = 0, \quad (9)$$

$$\lambda > 0 \Rightarrow \int_0^{+\infty} u(S(x)) dF_1(x) = C, \text{ and}$$

$$\mu > 0 \Rightarrow \int_0^{+\infty} u(S(x)) dF_1(x) - \int_0^{+\infty} u(S(x)) dF_0(x) = C.$$

From our previous reasoning, we already know that the IC constraint is binding. To see this again, observe that if it were not (i.e., $\mu = 0$), then (9) would reduce to $u'[S(x)] = 1/\lambda$ for all $x$; that is, a fixed payment.[15] But we

[15] Note that we've again established the result that, *absent* an incentive problem, a risk-neutral player should absorb all the risk when trading with a risk-averse player.

know a fixed-payment contract is *not* incentive compatible. It is also immediate that the participation constraint must be satisfied as an equality; otherwise, the manufacturer could reduce the payment schedule, thereby increasing her profits, in a manner that preserved the incentive constraint (i.e., replace $S^*(x)$ with $\tilde{S}(x)$, where $\tilde{S}(x) = u^{-1}(u[S^*(x)] - \varepsilon)$).

Note that the necessary conditions above are also sufficient given the assumed concavity of $u(\cdot)$. At every point where it is maximized with respect to $s$, $\left(\lambda + \mu\left[1 - \frac{f_0(x)}{f_1(x)}\right]\right)$ must be positive—observe it equals

$$\frac{1}{u'(s)} > 0$$

—so the second derivative with respect to $s$ must, therefore, be negative.

The least-cost contract inducing $a = 1$ therefore corresponds to a payment schedule $S(\cdot)$ that varies with the level of sales in a non-trivial way given by (9). That expression might look complicated, but its interpretation is central to the model and easy to follow. Observe, in particular, that because $u'(\cdot)$ is a decreasing function ($u(\cdot)$, recall, is strictly concave), $S(x)$ is positively correlated with

$$\lambda + \mu\left[1 - \frac{f_0(x)}{f_1(x)}\right];$$

that is, the larger (smaller) is this term, the larger (smaller) is $S(x)$.

The reward for a given level of sales $x$ depends upon the likelihood ratio

$$r(x) \equiv \frac{f_0(x)}{f_1(x)}$$

of the probability that sales are $x$ when action $a = 0$ is taken relative to that probability when action $a = 1$ is taken.[16] This ratio has a clear statistical meaning: It measures how more likely it is that the distribution from which sales have been determined is $F_0(\cdot)$ rather than $F_1(\cdot)$. When $r(x)$ is high, observing sales equal to $x$ allows the manufacturer to draw a statistical inference that it is much more likely that the distribution of sales was actually $F_0(.)$; that is, the salesperson did not expend effort promoting the product. In this case,

$$\lambda + \mu\left[1 - \frac{f_0(x)}{f_1(x)}\right]$$

is small (but necessarily positive) and $S(x)$ must also be small as well. When $r(x)$ is small, the manufacturer should feel rather confident that the salesperson expended effort and it should, then, optimally reward him highly. That is, sales levels that are relatively more likely when the agent has behaved in the desired manner result in larger payments to the agent than sales levels that are relatively rare when the agent has behaved in the desired manner.

[16] Technically, if $F_a(\cdot)$ is differentiable (i.e., $f_a(\cdot)$ is a probability density function), then the likelihood ratio can be interpreted as the probability that sales lie in $(x, x + dx)$ when $a = 0$ divided by the probability of sales lying in that interval when $a = 1$.

The minimum-cost incentive contract that induces the costly action $a = 1$ in essence commits the principal (manufacturer) to behave like a Bayesian statistician who holds some diffuse prior over which action the agent has taken:[17] She should use the observation of sales to revise her beliefs about what action the agent took and she should reward the agent more for outcomes that cause her to revise *upward* her beliefs that he took the desired action and she should reward him less (punish him) for outcomes that cause a *downward* revision in her beliefs.[18] As a consequence, the payment schedule is connected to sales only through their statistical content (the *relative* differences in the densities), *not* through their accounting properties. In particular, there is now no reason to believe that higher sales (larger $x$) should be rewarded more than lower sales. As an example, suppose that there are three possible sales levels: low, medium, and high ($x_L$, $x_M$, and $x_H$, respectively). Suppose, in addition, that

$$f_a(x) = \begin{cases} \frac{1}{3}, \text{ if } x = x_L \\ \frac{2-a}{6}, \text{ if } x = x_M \\ \frac{2+a}{6}, \text{ if } x = x_H \end{cases} .$$

Then

$$\lambda + \mu\left[1 - \frac{f_0(x)}{f_1(x)}\right] = \begin{cases} \lambda, \text{ if } x = x_L \\ \lambda - \mu, \text{ if } x = x_M \\ \lambda + \frac{\mu}{3}, \text{ if } x = x_H \end{cases} .$$

Hence, low sales are rewarded more than medium sales—low sales are uninformative about the salesperson's action, whereas medium sales suggest that the salesperson has *not* invested. Admittedly, non-monotonic compensation is rarely, if ever, observed in real life. We will see below what additional properties are required, in this model, to ensure monotonic compensation.

Note, somewhat implicit in our analysis to this point, is an assumption that $f_1(x) > 0$ except, possibly, on a subset of $x$ that are impossible (have zero measure). Without this assumption, (9) would entail division by zero, which is, of course, not permitted. If, however, we let $f_1(\cdot)$ go to zero on some subset of $x$ that had positive measure under $F_0(\cdot)$, then we see that $\mu$ must also tend to zero since

$$\lambda + \mu\left[1 - \frac{f_0(x)}{f_1(x)}\right]$$

must be positive. In essence, then, the shadow price (cost) of the incentive constraint is vanishing as $f_1(\cdot)$ goes to zero. This makes perfect sense: Were $f_1(\cdot)$ zero on some subset of $x$ that could occur (had positive measure) under $F_0(\cdot)$, then the occurrence of any $x$ in this subset, $\mathcal{X}_0$, would be proof that the

[17] A diffuse prior is one that assigns positive probability to each possible action.

[18] Of course, as a rational player of the game, the principal can infer that, if the contract is incentive compatible, the agent will have taken the desired action. Thus, there is not, in some sense, a real inference problem. Rather the issue is that, to be incentive compatible, the principal must commit to act *as if* there were an inference problem.

agent had failed to take the desired action. We can use this, then, to design a contract that induces $a = 1$, but which costs the principal no more than the optimal full-information fixed-payment contract $S(x) = s_1^F$. That is, the incentive problem ceases to be costly; so, not surprisingly, its shadow cost is zero.

To see how we can construct such a contract when $f_1(x) = 0$ for all $x \in \mathcal{X}_0$, let

$$S(x) = \begin{cases} \underline{s} + \varepsilon, \text{ if } x \in \mathcal{X}_0 \\ s_1^F, \text{ if } x \notin \mathcal{X}_0 \end{cases},$$

where $\varepsilon > 0$ is arbitrarily small ($\underline{s}$, recall, is the greatest lower bound of the domain of $u(\cdot)$). Then

$$\begin{aligned} \int_0^{+\infty} u(S(x))dF_1(x) &= u\left(s_1^F\right) \text{ and} \\ \int_0^{+\infty} u(S(x))dF_0(x) &= \int_{\mathcal{X}_0} u\left(\underline{s}+\varepsilon\right) dF_0(x) + \int_{\mathbb{R}_+\setminus\mathcal{X}_0} u\left(s_1^F\right) dF_0(x) \\ &= u\left(\underline{s}+\varepsilon\right) F_0(\mathcal{X}_0) + u\left(s_1^F\right)\left(1 - F_0(\mathcal{X}_0)\right). \end{aligned}$$

From the last expression, it's clear that $\int_0^{+\infty} u(S(x))dF_0(x) \to -\infty$ as $\varepsilon \to 0$; hence, the IC constraint is met trivially. By the definition of $s_1^F$, IR is also met. That is, this contract implements $a = 1$ at full-information cost. Again, as we saw in the two-outcome model, having a *shifting support* (i.e., the property that $F_0(\mathcal{X}_0) > 0 = F_1(\mathcal{X}_0)$) allows us to implement the desired action at full-information cost. SHIFTING SUPPORT

Let us conclude this sub-section by explaining how to answer the final question about the optimal contract. The manufacturer faces the following alternative: Either it imposes the fixed-payment contract $s_0^F$, which induces action $a = 0$, or it imposes the contract $S(\cdot)$ that has been derived above, which induces action $a = 1$. The expected-profit-maximizing choice results from the simple comparison of these two contracts; that is, the manufacturer imposes the incentive contract $S(\cdot)$ if and only if:

$$\mathbb{E}_0[x] - s_0^F < \mathbb{E}_1[x] - \mathbb{E}_1[S(x)] \tag{10}$$

The right-hand side of this inequality corresponds to the value of the maximization program (3). Given that the incentive constraint is binding in this program, this value is strictly smaller than the value of the same program without the incentive constraint; hence, just as we saw in the two-outcome case, the value is smaller than full-information profits, $E_1[x] - s_1^F$. Observe, therefore, that it is possible that

$$\mathbb{E}_1[x] - s_1^F > \mathbb{E}_0[x] - s_0^F > \mathbb{E}_1[x] - \mathbb{E}_1[S(x)]:$$

Under full information the principal would induce $a = 1$, but not if there's a hidden-action problem. In other words, imperfect observability of the agent's action imposes a cost on the principal that may induce her to distort the action that she induces the agent to take.

## 2.5 Monotonicity of the optimal contract

Let us suppose that (10) is, indeed, satisfied so that the contract $S(\cdot)$ derived above is the optimal contract. Can we exhibit additional and meaningful assumptions that would imply interesting properties of the optimal contract?

We begin with monotonicity, the idea that greater sales should mean greater compensation for the salesperson. As we saw above (page 20), there is no guarantee in the multiple-outcome model that this property should hold everywhere. From (9), it does hold if and only if the likelihood ratio, $r(x)$, is not decreasing and increasing at least somewhere. As this is an important property, it has a name:

**Definition 1 (MLRP)** *The likelihood ratio* $r(x) = f_0(x)/f_1(x)$ *satisfies the* monotone likelihood ratio property *(MLRP) if* $r(\cdot)$ *is non-increasing almost everywhere and strictly increasing on at least some set of* $x$*'s that occur with positive probability given action* $a = 1$*.*

The MLRP states that the greater is the outcome (i.e., $x$), the greater the *relative* probability of $x$ given $a = 1$ than given $a = 0$.[19] In other words, under MLRP, better outcomes are more likely when the salesperson expends effort than when he doesn't. To summarize:

**Proposition 1** *In the model of this section, if the likelihood ratio,* $r(\cdot)$*, satisfies the monotone likelihood ratio property, then the salesperson's compensation,* $S(\cdot)$*, under the optimal incentive contract for inducing him to expend effort (i.e., to choose* $a = 1$*) is non-decreasing everywhere.*

In fact, because we know that $S(\cdot)$ can't be constant—a fixed-payment contract is not incentive compatible for inducing $a = 1$—we can conclude that $S(\cdot)$ must, therefore, be increasing over some set of $x$.

[19] Technically, if $f_a(x) = F_a'(x)$, then we should say "the greater the relative probability of an $\hat{x} \in (x, x + dx)$ given $a = 1$ than given $a = 0$."

**Technical Aside**

Before analyzing the MLRP further, it's worth noting that even if MLRP does *not* hold, $r(\cdot)$ must be decreasing over some measurable range. To see this, suppose it were not true; that is, suppose that $r(\cdot)$ is almost everywhere non-decreasing under distribution $F_1(\cdot)$. Note this entails that $x$ and $r(x)$ are *non-negatively* correlated under $F_1(\cdot)$. To make the exposition easier, suppose for the purpose of this aside that $f_a(\cdot) = F_a'(\cdot)$. Then

$$\mathbb{E}_1\left[\frac{f_0(x)}{f_1(x)}\right] = \int_0^\infty \left(\frac{f_0(x)}{f_1(x)}\right) f_1(x)dx = \int_0^\infty f_0(x)dx = 1$$

Because $x$ and $r(x)$ are non-negatively correlated, we thus have

$$\int_0^\infty x[r(x) - \mathbb{E}_1\{r(x)\}]f_1(x)dx \geq 0.$$

Substituting, this implies

$$0 \leq \int_0^\infty x[\frac{f_0(x)}{f_1(x)} - 1]f_1(x)dx = \int_0^\infty xf_0(x)dx - \int_0^\infty xf_1(x)dx = \mathbb{E}_0[x] - \mathbb{E}_1[x].$$

But this contradicts our assumption that investing ($a = 1$) yields *greater* expected revenues than does not investing ($a = 0$). Hence, by contradiction it must be that $r(\cdot)$ is decreasing over some measurable range. But then this means that $S(\cdot)$ is increasing over some measurable range. However, without MLRP, we can't conclude that it's not also decreasing over some other measurable range.

**Conclusion:** *If $\mathbb{E}_0[x] < \mathbb{E}_1[x]$, then* $S(\cdot)$ is *increasing over some set of $x$ that has positive probability of occurring given action $a = 1$ even if MLRP does* not *hold.*

Is MLRP a reasonable assumption? To some extent is simply a strengthening of our assumption that $\mathbb{E}_1[x] > \mathbb{E}_0[x]$, given it can readily be shown that MLRP implies $\mathbb{E}_1[x] > \mathbb{E}_0[x]$.[20] Moreover, many standard distributions satisfy MLRP. But it quite easy to exhibit meaningful distributions that do not. For instance, consider our example above (page 20). We could model these distributions as the consequence of a two-stage stochastic phenomenon: With probability 1/3, a second new product is successfully introduced that eliminates the demand for the manufacturer's product (i.e., $x_L = 0$). With probability 2/3, this second product is not successfully introduced and it is "business as usual," with sales being more likely to be $x_M$ if the salesperson doesn't expend effort and more likely to be $x_H$ if he does. These "compounded" distributions do not satisfy MLRP. In such a situation, MLRP is not acceptable and the optimal reward schedule is not monotonic as we saw.

Although there has been a lot of discussion in the literature on the monotonicity issue, it may be overemphasized. If we return to the economic reality that the mathematics seeks to capture, the discussion relies on the assumption that a payment schedule with the feature that the salesperson is penalized for

[20] This can most readily be seen from the Technical Aside: Simply assume that $r(\cdot)$ satisfies MLRP, which implies $x$ and $r(x)$ are *negatively* correlated. Then, following the remaining steps, it quickly falls out that $\mathbb{E}_1[x] > \mathbb{E}_0[x]$.

increasing sales in some range does not actual induce a new agency problem. For instance, if the good is perishable or costly to ship, it might be possible for the salesperson to pretend, when sales are $x_M$, that they are $x_L$ (if the allegedly unsold amount of the good cannot be returned). That is, a non-monotonic incentive scheme could introduce a new agency problem of inducing the salesperson to report his sales honestly. Of course, if the manufacturer can verify the salesperson's inventory, a non-monotonic scheme might be possible. But think of another situation where the problem is to provide a worker (the agent) incentives to produces units of output; isn't it natural, then, to think that the worker could very well stop his production at $x_L$ or destroy his extra production $x_M - x_L$? Think of yet another situation where the problem is to provide incentives to a manager; aren't there many ways to spend money in a hardly detectable way so as to make profits look smaller than what they actually are? In short, the point is that if the agent can freely and secretly diminish his performance, then it makes no sense for the principal to have a reward schedule that is decreasing with performance over some range. In other words, there is often an *economic* justification for monotonicity even when MLRP doesn't hold.

## 2.6 Informativeness of the performance measures

Now, we again explore the question of the informativeness of the performance measures used by the principal. To understand the issue, suppose that the manufacturer in our example can also observe the sales of another of its products sold by the salesperson. Let $y$ denote these sales of the other good. These sales are, in part, also random, affected by forces outside the parties' control; but also, possibly, determined by how much effort the salesperson expends promoting the first product. For example, consumers could consider the two goods complements (or substitutes). Sales $y$ will then co-vary positively (or negatively) with sales $x$. Alternatively, both goods could be normal goods, so that sales of $y$ could then convey information about general market conditions (the incomes of the customers). Of course, it could also be that the demand for the second product is wholly unrelated to the demand for the first; in which case sales $y$ would be insensitive to the salesperson's action.

Let $f_0(x, y)$ and $f_1(x, y)$ denote the joint probability densities of sales $x$ and $y$ for action $a = 0$ and $a = 1$. An incentive contract can now be a function of both performance variables; i.e., $s = S(x, y)$. It is immediate that the same approach as before carries through and yields the following optimality condition:[21]

$$u'\left[S(x,y)\right]\left(\lambda + \mu\left[1 - \frac{f_0(x,y)}{f_1(x,y)}\right]\right) - 1 = 0. \tag{11}$$

When is it optimal to make compensation a function of $y$ as well as of $x$? The answer is straightforward: When the likelihood ratio, $r(x, y) = \frac{f_0(x,y)}{f_1(x,y)}$, actually

[21] Although we use the same letters for the Lagrange multipliers, it should be clear that their values at the optimum are not related to their values in the previous, one-performance-measure, contracting problem.

depends upon $y$. Conversely, when the likelihood ratio is independent of $y$, then there is no gain from contracting on $y$ to induce $a = 1$; indeed, it would be sub-optimal in this case since such a compensation scheme would fail to satisfy (11). The likelihood ratio is independent of $y$ if and only if the following holds: There exist three functions $h(\cdot,\cdot)$, $g_0(\cdot)$ and $g_1(\cdot)$ such that, for all $(x, y)$,

$$f_a(x, y) = h(x, y)g_a(x). \tag{12}$$

Necessity is obvious (divide $f_0(x, y)$ by $f_1(x, y)$ and observe the ratio, $g_0(x)/g_1(x)$, is independent of $y$). Sufficiency is also straightforward: Set $h(x, y) = f_1(x, y)$, $g_1(x) = 1$, and $g_0(x) = r(x)$. This condition of multiplicative separability, (12), has a well-established meaning in statistics: If (12) holds, then $x$ is a *sufficient statistic* for the action $a$ given data $(x, y)$. In words, were we trying to infer $a$, our inference would be just as good if we observed only $x$ as it would be if we observed the pair $(x, y)$. That is, conditional on knowing $x$, $y$ is *un*informative about $a$.

This conclusion is, therefore, quite intuitive, once we recall that the value of performance measures to our contracting property rests solely on their statistical properties. The optimal contract should be based on all performance measures that convey information about the agent's decision; but it is not desirable to include performance measures that are statistically redundant with other measures. As a corollary, there is no gain from considering *ex post* random contracts (e.g., a contract that based rewards on $x + \eta$, where $\eta$ is some random variable—noise—distributed independently of $a$ that is added to $x$).[22] As a second corollary, if the principal could freely eliminate noise in the performance measure—i.e., switch from observing $x + \eta$ to observing $x$—she would do better (at least weakly).

## 2.7 Conclusions from the two-action model

It may be worth summing up all the conclusions we have reached within the two-action model in a proposition:

**Proposition 2** *If the agent is strictly risk averse, there is no shifting support, and the principal seeks to implement the costly action (i.e., $a = 1$), then the principal's expected profits are smaller than under full (perfect) information. In some instances, this reduction in expected profits may lead the principal to implement the "free" action (i.e., $a = 0$).*

- *When (10) holds, the reward schedule imposes some risk on the risk-averse agent: Performances that are more likely when the agent takes the correct action $a = 1$ are rewarded more than performances that are more likely under $a = 0$.*
- *Under MLRP (or when the agent can destroy output), the optimal reward schedule is non-decreasing in performance.*

[22] *Ex ante* random *contracts* may, however, be of some value, as explained later.

- *The optimal reward schedule depends only upon performance measures that are sufficient statistics for the agent's action.*

To conclude, let us simply stress the two major themes that we would like the reader to remember from this section. First, imperfect information implies that the contractual reward designed by the principal should perform two tasks: Share the risks involved in the relationship and provide incentives to induce the agent to undertake the desired action. Except in trivial cases (e.g., a risk-neutral agent or a shifting support), these two goals are in conflict. Consequently, the optimal contract may induce an inefficient action *and* a Pareto suboptimal sharing of risk. Second, the optimal reward schedule establishes a link between rewards and performances that depends upon the statistical properties of the performance measures with respect to the agent's action.

### 2.8 Bibliographic notes

The analysis presented so far is fairly standard. The two-step approach—first determine, separately, the optimal contracts for implementing $a = 0$ and $a = 1$, then choose which yields greater profits—is due to Grossman and Hart (1983). The analysis, in the two-outcome case, when $q$ varies is also based on their work. They also consider the monotonicity of the optimal contract, although our analysis here draws more from Holmström (1979). Holmström is also our source for the sufficient-statistic result. Finally, the expression

$$\frac{1}{u'\left[S\left(x\right)\right]} = \lambda + \mu\left[1 - \frac{f_0(x)}{f_1(x)}\right],$$

which played such an important part in our analysis, is frequently referred to as the *modified* Borch sharing rule, in honor of Borch (1968), who worked out the rules for optimal risk sharing *absent* a moral-hazard problem (hence, the adjective "modified").

## 3 General formal setting

As we've just seen, the two-action model yields strong results. But the model incorporates a lot of structure and it relies on strong assumptions. Consequently, it's hard to understand which findings are robust and which are merely artifacts of an overly simple formalization. The basic ideas behind the incentive model are, we think, quite deep and we ought to show whether and how they generalize in less constrained situations.

Our approach is to propose a very general framework that captures the situation described in the opening section. Such generality comes at the cost of tractability, so we will again find ourselves making specific assumptions. But doing so, we will try to motivate the assumptions we have to make and discuss their relevance or underline how strong they are.

The situation of incentive contracting under hidden action or imperfect monitoring involves:

- a principal;
- an agent;
- a set of possible actions, $\mathcal{A}$, from which the agent chooses (we take $\mathcal{A}$ to be exogenously determined here);
- a set of verifiable signals or performance measures, $\mathcal{X}$;
- a set of benefits, $\mathcal{B}$, for the principal that are affected by the agent's action (possibly stochastically);
- rules (functions, distributions, or some combination) that relate elements of $\mathcal{A}$, $\mathcal{X}$, and $\mathcal{B}$;
- preferences for the principal and agent; and
- a bargaining game that establishes the contract between principal and agent (here, recall, we've fixed the bargaining game as the principal makes a take-it-or-leave-it offer, so that the only element of the bargaining game of interest here is the agent's reservation utility, $U_R$).[23]

In many settings, including the one explored above, the principal's benefit is the same as the verifiable performance measure (i.e., $b = x$). But this need not be the case. We could, for instance, imagine that there is a *function* mapping the elements of $\mathcal{A}$ onto $\mathcal{B}$. For example, the agent's action could be fixing the "true" quality of a product produced for the principal. This quality is also, then, the principal's benefit (i.e., $b = a$). The only *verifiable* measure of quality, however, is some noisy (i.e., stochastic) measure of true quality (e.g., $x = a + \eta$, where $\eta$ is some randomly determined distortion). As yet another possibility, the benchmark case of full information entails $\mathcal{X} = \mathcal{X}' \times \mathcal{A}$, where $\mathcal{X}'$ is some set of performance measures other than the action.

We need to impose some structure on $\mathcal{X}$ and $\mathcal{B}$ and their relationship to $\mathcal{A}$: We take $\mathcal{X}$ to be a Euclidean vector space and we let $dF(\cdot|a)$ denote the probability measure over $\mathcal{X}$ conditional on $a$. Similarly, we take $\mathcal{B}$ to be a Euclidean vector space and we let $dG(\cdot,\cdot|a)$ denote the joint probability measure over $\mathcal{B}$ and $\mathcal{X}$ conditional on $a$ (when $b \equiv x$, we will write $dF(\cdot|a)$ instead of $dG(\cdot,\cdot|a)$). This structure is rich enough to encompass the possibilities enumerated in the previous paragraph (and more).

Although we could capture the preferences of the principal and agent with*out* assuming the validity of the expected-utility approach to decision-making under uncertainty (we could, for instance, take as primitives the indifference curves

[23] We could also worry about whether the principal wants to participate—even make a take-it-or-leave-it offer—but because our focus is on the contract design and its execution, stages of the game not reached if she doesn't wish to participate, we will not explicitly consider this issue here.

shown in Figures 1 and 2), this approach has not been taken in the literature.[24] Instead, the expected-utility approach is assumed to be valid and we let $W(s, x, b)$ and $U(s, x, a)$ denote the respective von Neumann-Morgenstern utility of the principal and of the agent, where $s$ denotes the transfer from the principal to the agent (to principal from agent if $s < 0$).

In this situation, the obvious contract is a function that maps $\mathcal{X}$ into $\mathbb{R}$. We define such a contract as

**Definition 2** *A* simple *incentive contract is a reward schedule $S : \mathcal{X} \to \mathbb{R}$ that determines the level of reward $s = S(x)$ to be decided as a function of the realized performance level $x$.*

There is admittedly no other verifiable variable that can be used to write more elaborate contracts. There is, however, the possibility of creating verifiable variables, by having one or the other or both players take *verifiable* actions from some specified action spaces. Consistent with the mechanism-design approach, the most natural interpretation of these new variables are that they are public announcements made by the players; but nothing that follows requires this interpretation. For example, suppose both parties have to report to the third party charged with enforcing the contract their observation of $x$, or the agent must report which action he has chosen. We could even let the principal make a "good faith" report of what action she believes the agent took, although this creates its own moral-hazard problem because, in most circumstances, the principal could gain *ex post* by claiming she believes the agent's action was unacceptable. It turns out, as we will show momentarily, that there is nothing to be gained by considering such elaborate contracts; that is, there is no such contract that can improve over the optimal simple contract.

To see this, let us suppose that a contract determines a normal-form game to be played by both players after the agent has taken his action.[25,26] In particular, suppose the agent takes an action $h \in \mathcal{H}$ after choosing his action, but prior to the realization of $x$; that he takes an action $m \in \mathcal{M}$ after the realization of $x$; and that the principal also takes an action $n \in \mathcal{N}$ after $x$ has been realized. One or more of these sets could, but need not, contain a single element, a "null" action. We assume that the actions in these sets are costless—if we show that costless

[24]Given some well-documented deficiencies in expected-utility theory (see, e.g., Epstein, 1992; Rabin, 1997), this might, at first, seem somewhat surprising. However, as Epstein, §2.5, notes many of the predictions of expected-utility theory are robust to relaxing some of the more stringent assumptions that support it (e.g., such as the independence axiom). Given the tractability of the expected-utility theory combined with the general empirical support for the predictions of agency theory, we see the gain from sticking with expected-utility theory as outweighing the losses, if any, associated with that theory.

[25]Note this may require that there be some way that the parties can verify that the agent has taken an action. This may simply be the passage of time: The agent must take his action before a certain date. Alternatively, there could be a verifiable signal that the agent has acted (but which does not reveal *how* he's acted).

[26]Considering a extensive-form game with the various steps just considered would not alter the reasoning that follows; so, we avoid these unnecessary details by restricting attention to a normal-form game.

elaboration does no better than simple contracts, then costly elaboration also cannot do better than simple contracts. Finally, let the agent's compensation under this elaborate contract be: $s = \tilde{S}(x, h, m, n)$. We can now establish the following:

**Proposition 3 (Simple contracts are sufficient)** *For any general contract $\left\langle \mathcal{H}, \mathcal{M}, \mathcal{N}, \tilde{S}(\cdot) \right\rangle$ and associated (perfect Bayesian) equilibrium, there exists a simple contract $S(\cdot)$ that yields the same equilibrium outcome.*

---

**Proof:** Consider a (perfect Bayesian) equilibrium of the original contract, involving strategies $(a^*, h^*(\cdot), m^*(\cdot, \cdot))$ for the agent, where $h^*(a)$ and $m^*(a, x)$ describe the agent's choice within $\mathcal{H}$ and $\mathcal{M}$ after he's taken action $a$ and performance $x$ has been observed. Similarly, $n^*(x)$ gives the principal's choice of action as a function of the observed performance. Let us now consider the simple contract defined as follows: For all $x \in \mathcal{X}$,

$$S(x) \equiv \tilde{S}(x, h^*(a^*), m^*(x, a^*), n^*(x)).$$

Suppose that, facing this contract, the agent chooses an action $a$ different from $a^*$. Then, this implies that:

$$\int_{\mathcal{X}} U(S(x), x, a) dF(x|a) > \int_{\mathcal{X}} U(S(x), x, a^*) dF(x|a^*),$$

or, using the definition of $S(\cdot)$,

$$\begin{aligned} & \int_{\mathcal{X}} U\left[\tilde{S}(x, h^*(a^*), m^*(x, a^*), n^*(x)), x, a\right] dF(x|a) \\ > & \int_{\mathcal{X}} U\left[\tilde{S}(x, h^*(a^*), m^*(x, a^*), n^*(x)), x, a^*\right] dF(x|a^*). \end{aligned}$$

Since, *in the equilibrium of the normal-form game that commences* after *the agent chooses his action*, $h^*(\cdot)$ and $m^*(\cdot, \cdot)$ must satisfy the following inequality:

$$\begin{aligned} & \int_{\mathcal{X}} U\left[\tilde{S}(x, h(a), m^*(x, a), n^*(x)), x, a\right] dF(x|a) \\ \geq & \int_{\mathcal{X}} U\left[\tilde{S}(x, h^*(a^*), m^*(x, a^*), n^*(x)), x, a\right] dF(x|a), \end{aligned}$$

it follows that

$$\begin{aligned} & \int_{\mathcal{X}} U\left[\tilde{S}(x, h^*(a), m^*(x, a), n^*(x)), x, a\right] dF(x|a) \\ > & \int_{\mathcal{X}} U\left[\tilde{S}(x, h^*(a^*), m^*(x, a^*), n^*(x)), x, a^*\right] dF(x|a^*). \end{aligned}$$

This contradicts the fact the $a^*$ is an equilibrium action in the game defined by the original contract. Hence, the simple contract $S(\cdot)$ gives rise to the same action choice, and therefore the same distribution of outcomes than the more complicated contract. ■

---

As a consequence, there is no need to consider sophisticated announcement mechanisms in this setting, at least in the simple situation we have described.[27] It is worth noting that this style of proof is used many times in the study of contract theory. In particular, it is used to prove the *Revelation Principle*.

The contracting problem under imperfect information can now easily be stated. The principal, having the bargaining power in the negotiation process, simply has to choose a (simple) contract, $S(\cdot)$, so as to maximize her expected utility from the relationship given two constraints. First, the contract $S(\cdot)$ induces the agent to choose an action that maximizes his expected utility (i.e., the IC constraint must be met). Second, given the contract and the action it will induce, the agent must receive an expected utility at least as great as his reservation utility (i.e., the IR constraint must be met). In this general setting, the IC constraint can be stated as the action induced, $a$, must satisfy

$$a \in \arg\max_{a'} \int_{\mathcal{X}} U(S(x), x, a') dF(x|a'). \tag{13}$$

Observe that choosing $S(\cdot)$ amounts to choosing $a$ as well, at least when there exists a unique optimal choice for the agent. To take care of the possibility of multiple optima, one can simply imagine that the principal chooses a pair $(S(\cdot), a)$ subject to the incentive constraint (13). The IR constraint takes the simple form:

$$\max_{a'} \int_{\mathcal{X}} U(S(x), x, a') dF(x|a') \geq U_R. \tag{14}$$

The principal's problem is, thus,

$$\begin{aligned} &\max_{(S(\cdot),a)} \int_{\mathcal{X}} W(S(x), x, b) dG(b, x|a) \\ &\text{s.t. (13) and (14).} \end{aligned} \tag{15}$$

Observe, as we did in Section 2, it is perfectly permissible to solve this maximization program in two steps. First, for each action $a$, find the expected-profit-maximizing contract that implements action $a$ subject to the IC and IR constraints; this amounts to solving a similar program, taking action $a$ as fixed:

$$\begin{aligned} &\max_{S(.)} \int_{\mathcal{X}} W(S(x), x, b) dG(b, x|a) \\ &\text{s.t. (13) and (14).} \end{aligned} \tag{16}$$

Second, optimize the principal's objectives with respect to the action to implement; if we let $S_a(\cdot)$ denote the expected-profit-maximizing contract for implementing $a$, this second step consists of:

$$\max_{a \in \mathcal{A}} \int_{\mathcal{X}} W(S_a(x), x, b) dG(b, x|a).$$

[27] We will see that more sophisticated contracts may be strictly valuable when, for example, the agent gets an early private signal about performances. Intuitively though, sophisticated contracts are only useful when one player gets some private information about the realization of the state of nature along the relationship.

In this more general framework, it's worth revisiting the full-information benchmark. Before doing that, however, it is worth assuming that the domain of $U(\cdot, x, a)$ is sufficiently broad:

- Existence of a punishment: There exists some $s_P$ in the domain of $U(\cdot, x, a)$ for all $x \in \mathcal{X}$ and $a \in \mathcal{A}$ such that
$$\int_{\mathcal{X}} U(s_P, x, a')dF(x|a') < U_R$$
for all $a' \in \mathcal{A}$.

- Existence of a sufficient reward: There exists some $s_R$ in the domain of $U(\cdot, x, a)$ for all $x \in \mathcal{X}$ and $a \in \mathcal{A}$ such that
$$\int_{\mathcal{X}} U(s_R, x, a')dF(x|a') \geq U_R$$
for all $a' \in \mathcal{A}$.

In light of the second assumption, we can always satisfy (14) for any action $a$ (there is no guarantee, however, that we can also satisfy (13)).

With these two assumptions in hand, suppose that we're in the full-information case; that is, $\mathcal{X} = \mathcal{X}' \times \mathcal{A}$ (note $\mathcal{X}'$ could be a single-element space, so that we're also allowing for the possibility that, effectively, the only performance measure is the action itself). In the full-information case, the principal can rely on *forcing contracts*; that is, contracts that effectively leave the agent with no choice over the action he chooses. Hence, writing $(x', a)$ for an element of $\mathcal{X}$, a forcing contract for implementing $\hat{a}$ is

$$\begin{aligned} S(x', a) &= s_P \text{ if } a \neq \hat{a} \\ &= S^F(x') \text{ if } a = \hat{a}, \end{aligned}$$

where $S^F(\cdot)$ satisfies (14). Since $S^F(x') = s_R$ satisfies (14) by assumption, we know that we can find *a* $S^F(\cdot)$ that satisfies (14). In equilibrium, the agent will choose to sign the contract—the IR constraint is met—and he will take action $\hat{a}$ since this is his only possibility for getting at least his reservation utility. Forcing contracts are very powerful because they transform the contracting problem into a simple *ex ante* Pareto computation program:

$$\max_{(S(\cdot),a)} \int_{\mathcal{X}} W(S(x), x, b)dG(b, x|a) \qquad (18)$$
$$\text{s.t. } (14),$$

where only the agent's participation constraint matters. This *ex ante* Pareto program determines the efficient risk-sharing arrangement for the full-information optimal action, as well as the full-information optimal action itself. Its solution characterizes the optimal contract under perfect information.

At this point, we've gone about as far as we can go without imposing more structure on the problem. The next few sections consider more structured variations of the problem.

# 4 The Finite Model

In this section, we suppose that $\mathcal{A}$, the set of possible actions, is finite with $J$ elements. Likewise, the set of possible verifiable performance measures, $\mathcal{X}$, is also taken to be finite with $N$ elements, indexed by $n$ (although, at the end of this section, we'll discuss the case where $\mathcal{X} = \mathbb{R}$). Given the world is, in reality, finite, this is the most general version of the principal-agent model (although not necessarily the most analytically tractable).[28]

We will assume that the agent's utility is additively separable between payments and action. Moreover, it is not, directly, dependent on performance. Hence,

$$U(s, x, a) = u(s) - c(a);$$

where $u : \mathcal{S} \to \mathbb{R}$ maps some subset $\mathcal{S}$ of $\mathbb{R}$ into $\mathbb{R}$ and $c : \mathcal{A} \to \mathbb{R}$ maps the action space into $\mathbb{R}$. As before, we assume that $\mathcal{S} = (\underline{s}, \infty)$, where $u(s) \to -\infty$ as $s \downarrow \underline{s}$. Observe that this assumption entails the existence of a punishment, $s_P$, as described in the previous section. We further assume that $u(\cdot)$ is strictly monotonic and concave (at least weakly). Typically, we will assume that $u(\cdot)$ is, in fact, strictly concave, implying the agent is risk averse. Note, that the monotonicity of $u(\cdot)$ implies that the inverse function $u^{-1}(\cdot)$ exists and, since $u(\mathcal{S}) = \mathbb{R}$, is defined for all $u \in \mathbb{R}$.

We assume, now, that $\mathcal{B} \subset \mathbb{R}$ and that the principal's utility is a function only of the difference between her benefit, $b$, and her payment to the agent; that is,

$$W(s, x, b) = w(b - s),$$

where $w(\cdot)$ is assumed to be strictly increasing and concave. In fact, in most applications—particularly most applications of interest in the study of strategy and organization—it is reasonable to assume that the principal is risk neutral. We will maintain that assumption here (the reader interested in the case of a risk-averse principal should consult Holmström, 1979, among other work). In what follows, let $B(a) = \mathbb{E}\{b|a\}$.

In addition to being discrete, we assume that there exists some partial order on $\mathcal{X}$ (i.e., to give meaning to the idea of "better" or "worse" performance) and that, with respect to this partial order, $\mathcal{X}$ is a chain (i.e., if $\preceq$ is the partial order on $\mathcal{X}$, then $x \preceq x'$ or $x' \preceq x$ for any two elements, $x$ and $x'$ in $\mathcal{X}$). Because identical signals are irrelevant, we may also suppose that no two elements of $\mathcal{X}$ are the same (i.e., $x \prec x'$ or $x' \prec x$ for any two elements in $\mathcal{X}$). The most natural interpretation is that $\mathcal{X}$ is a subset of distinct real numbers—different "performance scores"—with $\leq$ as the partial order. Given these assumptions, we can write $\mathcal{X} = \{x_1, \ldots, x_N\}$, where $x_m \prec x_n$ if $m < n$. Likewise the

[28]Limitations on measurement—of both inputs and outputs—make the world discrete. For instance, effort, measured as time on the job, can take only discrete values because only discrete intervals of time can be measured. Similarly, output, measured as volume, can take only discrete values because only discrete intervals of volume can be measured. Because the world is also bounded—each of us is allotted only so much time and there are only so many atoms in the universe (under current cosmological theories)—it follows that the world is finite.

distribution function $F(x|a)$ gives, for each $x$, the probability that an $x' \preceq x$ is realized conditional on action $a$. The corresponding density function is then defined by

$$f(x_n|a) = \begin{cases} F(x_1|a) & \text{if } n = 1 \\ F(x_n|a) - F(x_{n-1}|a) & \text{if } n > 1 \end{cases}.$$

In much of what follows, it will be convenient to write $f_n(a)$ for $f(x_n|a)$. It will also be convenient to write the density as a vector:

$$\mathbf{f}(a) = (f_1(a), \ldots, f_N(a))^\top$$

(observe, unless indicated otherwise, vectors are *column* vectors; hence the final $\top$, which denotes matrix Transpose, is needed to transform the row vector into a column vector).

## 4.1 The "two-step" approach

As we have done already in Section 2, we will pursue a two-step approach to solving the principal-agent problem:

**Step 1:** For each $a \in \mathcal{A}$, the principal determines whether it can be implemented. Let $\mathcal{A}^I$ denote the set of implementable actions. For each $a \in \mathcal{A}^I$, the principal determines the least-cost contract for implementing $a$ subject to the IC and IR constraints. Let $C(a)$ denote the principal's expected cost (expected payment) of implementing $a$ under this least-cost contract.

**Step 2:** The principal then determines the solution to the maximization problem

$$\max_{a \in \mathcal{A}^I} B(a) - C(a).$$

If $a^*$ is the solution to this maximization problem, the principal offers the least-cost contract for implementing $a^*$.

Note that this two-step process is analogous to a standard production problem, in which a firm, first, solves its cost-minimization problems to determine the least-cost way of producing any given amount of output (i.e., derives its cost function); and, then, it produces the amount of output that maximizes the difference between revenues (benefits) and cost. As with production problems, the first step is generally the harder step.

## 4.2 The full-information benchmark

As before, we consider as a benchmark the case where the principal can observe and verify the agent's action. Consequently, as we discussed at the end of Section 3, the principal can implement any action $\hat{a}$ that she wants using a forcing contract: The contract punishes the agent sufficiently for choosing actions $a \neq \hat{a}$ that he would never choose any action other than $\hat{a}$; and the contract rewards the agent sufficiently for choosing $\hat{a}$ that he is just willing to sign the principal's contract. This last condition can be stated formally as

$$u(\hat{s}) - c(\hat{a}) = U_R, \tag{IR$^F$}$$

where $\hat{s}$ is what the agent is paid if he chooses action $\hat{a}$. Solving this last expression for $\hat{s}$ yields

$$\hat{s} = u^{-1}[U_R + c(\hat{a})] \equiv C^F(\hat{a}).$$

The function $C^F(\cdot)$ gives the cost, under full information, of implementing actions.

## 4.3 The hidden-action problem

Now, and henceforth, we assume that a hidden-action problem exists. Consequently, the only feasible contracts are those that make the agent's compensation contingent on the verifiable performance measure. Let $s(x)$ denote the payment made to the agent under such a contract if $x$ is realized. It will prove convenient to write $s_n$ for $s(x_n)$ and to consider the compensation *vector* $\mathbf{s} = (s_1, \ldots, s_N)^\top$. The optimal—expected-cost-minimizing—contract for implementing $\hat{a}$ (assuming it can be implemented) is the contract that solves the following program:[29]

$$\min_{\mathbf{s}} \mathbf{f}(\hat{a})^\top \mathbf{s}$$

subject to

$$\sum_{n=1}^{N} f_n(\hat{a}) u(s_n) - c(\hat{a}) \geq U_R$$

(the IR constraint) and

$$\hat{a} \in \max_a \sum_{n=1}^{N} f_n(a) u(s_n) - c(a)$$

(the IC constraint—see (13)). Observe that an equivalent statement of the IC constraint is

$$\sum_{n=1}^{N} f_n(\hat{a}) u(s_n) - c(\hat{a}) \geq \sum_{n=1}^{N} f_n(a) u(s_n) - c(a) \quad \forall a \in \mathcal{A}.$$

As we've seen above, it is often easier to work in terms of utility payments than in terms of monetary payments. Specifically, because $u(\cdot)$ is invertible, we can express a contract as an $N$-dimensional vector of contingent utilities,

[29] Observe, given the separability between the principal's benefit and cost, minimizing her expected wage payment is equivalent to maximizing her expected profit.

$\mathbf{u} = (u_1, \ldots, u_N)^\top$, where $u_n = u(s_n)$. Using this "trick," the principal's program becomes

$$\min_{\mathbf{u}} \sum_{n=1}^{N} f_n(\hat{a})\, u^{-1}(u_n) \tag{19}$$

subject to

$$\mathbf{f}(\hat{a})^\top \mathbf{u} - c(\hat{a}) \geq U_R \tag{IR}$$

and

$$\mathbf{f}(\hat{a})^\top \mathbf{u} - c(\hat{a}) \geq \mathbf{f}(a)^\top \mathbf{u} - c(a) \quad \forall a \in \mathcal{A}. \tag{IC}$$

**Definition 3** *An action $\hat{a}$ is* implementable *if there exists at least one contract solving (IR) and (IC).*

A key result is the following:

**Proposition 4** *If $\hat{a}$ is implementable, then there exists a contract that implements $\hat{a}$ and satisfies (IR) as an equality. Moreover, (IR) is met as an equality (i.e., is binding) under the optimal contract for implementing $\hat{a}$.*

---

**Proof:** Suppose not: Let $\mathbf{u}$ be a contract that implements $\hat{a}$ and suppose that

$$\mathbf{f}(\hat{a})^\top \mathbf{u} - c(\hat{a}) > U_R.$$

Define

$$\varepsilon = \mathbf{f}(\hat{a})^\top \mathbf{u} - c(\hat{a}) - U_R.$$

By assumption, $\varepsilon > 0$. Consider a new contract, $\tilde{\mathbf{u}}$, where $\tilde{u}_n = u_n - \varepsilon$. By construction, this new contract satisfies (IR). Moreover, because

$$\mathbf{f}(a)^\top \tilde{\mathbf{u}} = \mathbf{f}(a)^\top \mathbf{u} - \varepsilon,$$

for all $a \in \mathcal{A}$, this new contract also satisfies (IC). Observe, too, that this new contract is superior to $\mathbf{u}$: It satisfies the contracts, while costing the principal less. Hence, a contract cannot be optimal unless (IR) is an equality under it. ∎

---

In light of this proposition, it follows that an action $\hat{a}$ can be implemented if there is a contract $\mathbf{u}$ that solves the following system:

$$\mathbf{f}(\hat{a})^\top \mathbf{u} - c(\hat{a}) = U_R \tag{20}$$

and

$$\mathbf{f}(a)^\top \mathbf{u} - c(a) \leq U_R \quad \forall a \in \mathcal{A} \backslash \{\hat{a}\} \tag{21}$$

(where (21) follows from (IC) and (20)). We are now in position to establish the following proposition:

**Proposition 5** *Action $\hat{a}$ is implementable if and only if there is no strategy for the agent that induces the same density over signals as $\hat{a}$ and which costs the agent less, in terms of expected disutility, than $\hat{a}$ (where "strategy" refers to mixed, as well as, pure strategies).*

---

**Proof:** Let $j = 1, \ldots, J-1$ index the elements in $\mathcal{A}$ *other* than $\hat{a}$. Then the system (20) and (21) can be written as $J+1$ inequalities:

$$\begin{gathered} \mathbf{f}(\hat{a})^{\top} \mathbf{u} \leq U_R + c(\hat{a}) \\ [-\mathbf{f}(\hat{a})]^{\top} \mathbf{u} \leq -U_R - c(\hat{a}) \\ \mathbf{f}(a_1)^{\top} \mathbf{u} \leq U_R + c(a_1) \\ \vdots \\ \mathbf{f}(a_{J-1})^{\top} \mathbf{u} \leq U_R + c(a_{J-1}) \end{gathered}$$

By a well-known result in convex analysis (see, e.g., Rockafellar, 1970, page 198), there is a $\mathbf{u}$ that solves this system if and only if there is *no* vector

$$\mu = (\hat{\mu}_+, \hat{\mu}_-, \mu_1, \ldots, \mu_{J-1})^{\top} \geq \mathbf{0}_{J+1}$$

(where $\mathbf{0}_K$ is a $K$-dimensional vector of zeros) such that

$$\hat{\mu}_+ \mathbf{f}(\hat{a}) + \hat{\mu}_- [-\mathbf{f}(\hat{a})] + \sum_{j=1}^{J-1} \mu_j \mathbf{f}(a_j) = \mathbf{0}_N \tag{22}$$

and

$$\hat{\mu}_+ [U_R + c(\hat{a})] + \hat{\mu}_- [-U_R - c(\hat{a})] + \sum_{j=1}^{J-1} \mu_j [U_R + c(a_j)] < 0. \tag{23}$$

Observe that if such a $\mu$ exists, then (23) entails that not all elements can be zero. Define $\mu_* = \hat{\mu}_+ - \hat{\mu}_-$. By post-multiplying (22) by $\mathbf{1}_N$ (an $N$-dimensional vector of ones), we see that

$$\mu_* + \sum_{j=1}^{J-1} \mu_j = 0. \tag{24}$$

Equation (24) implies that $\mu_* < 0$. Define $\sigma_j = \mu_j / (-\mu_*)$. By construction each $\sigma_j \geq 0$ (with at least some being strictly greater than 0) and, from (24), $\sum_{j=1}^{J-1} \sigma_j = 1$. Hence, we can interpret these $\sigma_j$ as probabilities and, thus, as a mixed strategy over the elements of $\mathcal{A} \backslash \{\hat{a}\}$. Finally, observe that if we divide both sides of (22) and (23) by $-\mu_*$ and rearrange, we can see that (22) and (23) are equivalent to

$$\mathbf{f}(\hat{a}) = \sum_{j=1}^{J-1} \sigma_j \mathbf{f}(a_j) \tag{25}$$

and

$$c(\hat{a}) > \sum_{j=1}^{J-1} \sigma_j c(a_j); \tag{26}$$

that is, there is a contract $\mathbf{u}$ that solves the above system of inequalities if and only if there is no (mixed) strategy that induces the same density over the performance measures as $\hat{a}$ (i.e., (25)) and that has lower expected cost (i.e., (26)). ■

The truth of the necessity condition (only if part) of Proposition 5 is straightforward: Were there such a strategy—one that always produced the same expected utility over money as $\hat{a}$, but which cost the agent less than $\hat{a}$—then it would clearly be impossible to implement $\hat{a}$ as a pure strategy. What is less obvious is the sufficiency (if part) of the proposition. Intuitively, if the density over the performance measure induced by $\hat{a}$ is distinct from the density induced by any other strategy, then the performance measure is informative with respect to determining whether $\hat{a}$ was the agent's strategy or whether he played a different strategy. Because the range of $u(\cdot)$ is unbounded, even a small amount of information can be exploited to implement $\hat{a}$ by rewarding the agent for performance that is relatively more likely to occur when he plays the strategy $\hat{a}$, or by punishing him for performance that is relatively unlikely to occur when he plays the strategy $\hat{a}$, or both. Of course, even if there are other strategies that induce the same density as $\hat{a}$, $\hat{a}$ is still implementable if the agent finds these other strategies more costly than $\hat{a}$.

**Technical Aside**

We can formalize this notion of informationally distinct as follows: The condition that no strategy duplicate the density over performance measures induced by $\hat{a}$ is equivalent to saying that there is *no* density (strategy) $(\sigma_1, \ldots, \sigma_{J-1})$ over the *other* $J-1$ elements of $\mathcal{A}$ such that

$$\mathbf{f}(\hat{a}) = \sum_{j=1}^{J-1} \sigma_j \mathbf{f}(a_j).$$

Mathematically, that's equivalent to saying that $\mathbf{f}(\hat{a})$ is *not* a convex combination of $\{\mathbf{f}(a)\}_{a\in\mathcal{A}\setminus\{\hat{a}\}}$; or, equivalently that $\mathbf{f}(\hat{a})$ is *not* in the *convex hull* of $\{\mathbf{f}(a)|a \neq \hat{a}\}$. See Hermalin and Katz (1991) for more on this "convex-hull" condition and its interpretation. Finally, from Proposition 5, the condition that $\mathbf{f}(\hat{a})$ *not* be in the convex hull of $\{\mathbf{f}(a)|a \neq \hat{a}\}$ is *sufficient* for $\hat{a}$ to be implementable.

Before solving the principal's problem (Step 1, page 33), it's worth considering, and then dismissing, two "pathological" cases. The first is the ability to implement *least-cost actions* at their full-information cost. The second is the ability to implement any action at its full-information cost when there is a *shifting support* (of the right kind).

**Definition 4** *An action $\tilde{a}$ is a* least-cost action *if $\tilde{a} \in \arg\min_{a\in\mathcal{A}} c(a)$. That is, $\tilde{a}$ is a least-cost action if the agent's disutility from choosing any other action is at least as great as his disutility from choosing $\tilde{a}$.*

**Proposition 6** *If $\tilde{a}$ is a least-cost action, then it is implementable at its full-information cost.*

**Proof:** Consider the fixed-payment contract that pays the agent $u_n = U_R + c(\tilde{a})$ for all $n$. This contract clearly satisfies (IR) and, because $c(\tilde{a}) \leq c(a)$ for all $a \in \mathcal{A}$, it also satisfies (IC). The cost of this contract to the principal is $u^{-1}[U_R + c(\tilde{a})] = C^F(\tilde{a})$, the full-information cost. ■

---

Of course, there is nothing surprising to Proposition 6: When the principal wishes to implement a least-cost action, her interests and the agent's are perfectly aligned; that is, there is no agency problem. Consequently, it is not surprising that the full-information outcome obtains.

**Definition 5** *There is a* meaningful shifting support *associated with action $\tilde{a}$ if there exists a subset of $\mathcal{X}$, $\mathcal{X}_0$, such that $F(\mathcal{X}_0|a) > 0 = F(\mathcal{X}_0|\tilde{a})$ for* all *actions $a$ such that $c(a) < c(\tilde{a})$.*

**Proposition 7** *Let there be a meaningful shifting support associated with action $\tilde{a}$. Then action $\tilde{a}$ is implementable at its full-information cost.*

---

**Proof:** Fix some arbitrarily small $\varepsilon > 0$ and define $u_P = u(\underline{s} + \varepsilon)$. Consider the contract $\mathbf{u}$ that sets $u_m = u_P$ if $x_m \in \mathcal{X}_0$ (where $\mathcal{X}_0$ is defined above) and that sets $u_n = U_R + c(\tilde{a})$ if $x_n \notin \mathcal{X}_0$. It follows that $\mathbf{f}(\tilde{a})^\top \mathbf{u} = U_R + c(\tilde{a})$; that $\mathbf{f}(a)^\top \mathbf{u} \to -\infty$ as $\varepsilon \downarrow 0$ for $a$ such that $c(a) < c(\tilde{a})$; and that

$$\mathbf{f}(a)^\top \mathbf{u} - c(a) \leq U_R + c(\tilde{a}) - c(a) \leq \mathbf{f}(\tilde{a})^\top \mathbf{u} - c(\tilde{a})$$

for $a$ such that $c(a) \geq c(\tilde{a})$. Consequently, this contract satisfies (IR) and (IC). Moreover, the *equilibrium* cost of this contract to the principal is $u^{-1}[U_R + c(\tilde{a})]$, the full-information cost. ■

---

Intuitively, when there is a meaningful shifting support, observing an $x \in \mathcal{X}_0$ is proof that the agent took an action other than $\tilde{a}$. Because the principal has this proof, she can punish the agent as severely as she wishes when such an $x$ appears (in particular, she doesn't have to worry about how this punishment changes the risk faced by the agent, given the agent is never in jeopardy of suffering this punishment if he takes the desired action, $\tilde{a}$).[30] Moreover, such a draconian punishment will deter the agent from taking an action that induces a positive probability of suffering the punishment. In effect, such actions have been dropped from the original game, leaving $\tilde{a}$ as a least-cost action of the new game. It follows, then, from Proposition 6, that $\tilde{a}$ can be implemented at its full-information cost.

[30] It is worth noting that this argument relies on being able to punish the agent sufficiently in the case of an $x \in \mathcal{X}_0$. Whether the use of such punishments is really feasible could, in some contexts, rely on assumptions that are overly strong. First, that the agent hasn't (or can contractually waive) protections against severe punishments. For example, in the English common-law tradition, this is generally not true; moreover, courts in these countries are generally loath to enforce contractual clauses that are deemed to be penalties. Second, that the agent has faith in his understanding of the distributions (i.e., he is sure that taking action $\tilde{a}$ guarantees that an $x \in \mathcal{X}_0$ won't occur). Third, that the agent has faith in his own rationality; that is, in particular, he is sufficiently confident that won't make a mistake (i.e., choose an $a$ such that $F(\mathcal{X}_0|a) > 0$).

It is worth noting that the full-information benchmark is just a special case of Proposition 7, in which the support of $\mathbf{f}(a)$ lies on a separate plane, $\mathcal{X}' \times \{a\}$, for each action $a$.

Typically, it is assumed that the action the principal wishes to implement is neither a least-cost action, nor has a meaningful shifting support associated with it. Henceforth, we will assume that the action that principal wishes to implement, $\hat{a}$, is not a least-cost action (i.e., $\exists a \in \mathcal{A}$ such that $c(a) < c(\hat{a})$). Moreover, we will rule out all shifting supports by assuming that $f_n(a) > 0$ for all $n$ and all $a$.

We now consider the whether there is a solution to Step 1 when there is no shifting support and the action to be implemented is not a least-cost action. That is, we ask the question: If $\hat{a}$ is implementable is there an optimal contract for implementing it? We divide the analysis into two cases: $u(\cdot)$ affine (risk neutral) and $u(\cdot)$ strictly concave (risk averse).

**Proposition 8** *Assume $u(\cdot)$ is affine and that $\hat{a}$ is implementable. Then $\hat{a}$ is implementable at its full-information cost.*

---

**Proof:** Let $\mathbf{u}$ solve (IR) and (IC). From Proposition 4, we may assume that (IR) is binding. Then, because $u(\cdot)$ and, thus, $u^{-1}(\cdot)$ are affine:

$$\sum_{n=1}^{N} f_n(\hat{a})\, u^{-1}(u_n) = u^{-1}\left(\sum_{n=1}^{N} f_n(\hat{a})\, u_n\right) = u^{-1}\left[U_R + c(\hat{a})\right],$$

where the last inequality follows from the fact that (IR) is binding. ■

---

Note, given that we can't do better than implement an action at full-information cost, this proposition also tells us that, with a risk-neutral agent, an optimal contract exists for inducing any implementable action. The hidden-action problem (the lack of full information) is potentially costly to the principal for two reasons. First, it may mean a desired action is not implementable. Second, even if it is implementable, it may be implementable at a higher cost. Proposition 8 tells us that this second source of cost must be due solely to the agent's risk aversion; an insight consistent with those derived earlier.

In fact, if we're willing to assume that the principal's benefit is alienable—that is, she can sell the rights to receive it to the agent—and that the agent is risk neutral, then we can implement the optimal full-information action, $a^*$ (i.e., the solution to Step *2* under full information) at full-information cost. In other words, we can achieve the complete full-information solution in this case:

**Proposition 9 (Selling the store)** *Assume that $u(\cdot)$ is affine and that the principal's benefit is alienable. Then the principal can achieve the same expected utility with a hidden-action problem as she could under full information.*

---

**Proof:** Under full information, the principal would induce $a^*$ where

$$a^* \in \arg\max_{a \in \mathcal{A}} B(a) - C^F(a).$$

Define
$$t^* = B(a^*) - C^F(a^*).$$
Suppose the principal offers to sell the right to her benefit to the agent for $t^*$. If the agent accepts, then the principal will enjoy the same utility she would have enjoyed under full information. Will the agent accept? Note that because $u(\cdot)$ is affine, there is no loss of generality in assuming it is the identity function. If he accepts, he faces the problem
$$\max_{a \in \mathcal{A}} \int_{\mathcal{B}} (b - t^*)\, dG(b|a) - c(a).$$
This is equivalent to
$$\max_{a \in \mathcal{A}} B(a) - c(a) - B(a^*) + c(a^*) + U_R; \text{ or to}$$
$$\max_{a \in \mathcal{A}} B(a) - [c(a) + U_R] - B(a^*) + c(a^*) + 2U_R.$$
Because $B(a) - [c(a) + U_R] = B(a) - C^F(a)$, rational play by the agent conditional on accepting means his utility will be $U_R$; which also means he'll accept. ■

---

People often dismiss the case where the agent is risk neutral by claiming that there is no agency problem because the principal could "sell the store (productive asset)" to the agent. As this last proposition makes clear, such a conclusion relies critically on the ability to literally sell the asset; that is, if the principal's benefit is not alienable, then this conclusion might not hold.[31] In other words, it is not solely the agent's risk aversion that causes problems with a hidden action.

**Corollary 1** *Assume that $u(\cdot)$ is affine and that the principal's benefit equals the performance measure (i.e., $\mathcal{B} = \mathcal{X}$ and $G(\cdot|a) = F(\cdot|a)$). Then the principal can achieve the same expected utility with a hidden-action problem as she could under full information.*

---

**Proof:** Left to the reader. [Hint: Let $s(x) = x - t$, where $t$ is a constant.] ■

---

Now we turn our attention to the case where $u(\cdot)$ is strictly concave (the agent is risk averse). Observe (i) this entails that $u^{-1}(\cdot)$ is strictly convex; (ii), because $\mathcal{S}$ is an open interval, that $u(\cdot)$ is continuous; and (iii) that $u^{-1}(\cdot)$ is continuous.

**Proposition 10** *Assume that $u(\cdot)$ is strictly concave. If $\hat{a}$ is implementable, then there exists a unique contract that implements $\hat{a}$ at minimum expected cost.*

[31] To see this, suppose the benefit is *un*alienable. Assume, too, that $\mathcal{A} = \{1/4, 1/2, 3/4\}$, $\mathcal{X} = \{1, 2\}$, $c(a) = \sqrt{a}$, $f_2(a) = a$, $U_R = 0$, and $B(a) = 4 - 4\left(a - \frac{1}{2}\right)^2$. Then it is readily seen that $a^* = 1/2$. However, from Proposition 5, $a^*$ is *not* implementable, so the full-information outcome is unobtainable when the action is hidden (even though the agent is risk neutral).

**Proof:** Existence.[32] Define

$$\Omega\left(\mathbf{u}\right)=\sum_{n=1}^{N}f_n\left(\hat{a}\right)u^{-1}\left(u_n\right). \tag{27}$$

The strict convexity and continuity of $u^{-1}(\cdot)$ implies that $\Omega(\cdot)$ is also a strictly convex and continuous function. Observe that the principal's problem is to choose $\mathbf{u}$ to minimize $\Omega(\mathbf{u})$ subject to (IR) and (IC). Let $\mathcal{U}$ be the set of contracts that satisfy (IR) and (IC) (by assumption, $\mathcal{U}$ is not empty). Were $\mathcal{U}$ closed and bounded, then a solution to the principal's problem would certainly exist because $\Omega(\cdot)$ is a continuous real-valued function.[33] Unfortunately, $\mathcal{U}$ is not bounded (although it is closed given that all the inequalities in (IR) and (IC) are weak inequalities). Fortunately, we can artificially bound $\mathcal{U}$ by showing that any solution outside some bound is inferior to a solution inside the bound. Consider any contract $\mathbf{u}^0\in\mathcal{U}$ and consider the contract $\mathbf{u}^*$, where $u_n^*=U_R+c(\hat{a})$. Let $\mathcal{U}^{IR}$ be the set of contracts that satisfy (IR). Note that $\mathcal{U}\subset\mathcal{U}^{IR}$. Note, too, that both $\mathcal{U}$ and $\mathcal{U}^{IR}$ are convex sets. Because $\Omega(\cdot)$ has a minimum on $\mathcal{U}^{IR}$, namely $\mathbf{u}^*$, the set

$$\mathcal{V}\equiv\left\{\mathbf{u}\in\mathcal{U}^{IR}|\Omega\left(\mathbf{u}\right)\leq\Omega\left(\mathbf{u}^0\right)\right\}$$

is closed, bounded, and convex.[34] By construction, $\mathcal{U}\cap\mathcal{V}$ is non-empty; moreover, for any $\mathbf{u}^1\in\mathcal{U}\cap\mathcal{V}$ and any $\mathbf{u}^2\in\mathcal{U}\backslash\mathcal{V}$, $\Omega\left(\mathbf{u}^2\right)>\Omega\left(\mathbf{u}^1\right)$. Consequently, nothing is lost be limiting the search for an optimal contract to $\mathcal{U}\cap\mathcal{V}$. The set $\mathcal{U}\cap\mathcal{V}$ is closed and bounded and $\Omega(\cdot)$ is continuous, hence it follows that an optimal contract must exist.

Uniqueness. Suppose the optimal contract, $\mathbf{u}$, were not unique. That is, there exists another contract $\tilde{\mathbf{u}}$ such that $\Omega(\mathbf{u})=\Omega(\tilde{\mathbf{u}})$ (where $\Omega(\cdot)$ is defined by (27)). It is readily seen that if these two contracts each satisfy both the (IR) and (IC) constraints, then any convex combination of them must as well (i.e., both are elements of $\mathcal{U}$, which is convex). That is, the contract

$$\mathbf{u}_\lambda\equiv\lambda\mathbf{u}+(1-\lambda)\tilde{\mathbf{u}},$$

$\lambda\in(0,1)$ must be feasible (i.e., satisfy (IR) and (IC)). Since $\Omega(\cdot)$ is strictly convex, Jensen's inequality implies

$$\Omega\left(\mathbf{u}_\lambda\right)<\lambda\Omega\left(\mathbf{u}\right)+(1-\lambda)\Omega\left(\tilde{\mathbf{u}}\right)=\Omega\left(\mathbf{u}\right).$$

But this contradicts the optimality of $\mathbf{u}$. By contradiction, uniqueness is established. ■

Having concluded that a solution to Step 1 exists, we can—at last—calculate what it is. From Proposition 8, the problem is trivial if $u(\cdot)$ is affine, so we will consider only the case in which $u(\cdot)$ is strictly concave. The principal's problem

[32] The existence portion of this proof is somewhat involved mathematically and can be omitted without affecting later comprehension of the material.

[33] This is a well-known result from analysis (see, e.g., Fleming, 1977, page 49).

[34] The convexity of $\mathcal{V}$ follows because $\Omega(\cdot)$ is a convex function and $\mathcal{U}^{IR}$ is a convex set. That $\mathcal{V}$ is closed follows given $\mathcal{U}^{IR}$ is also closed. To see that $\mathcal{V}$ is bounded, recognize that, as one "moves away" from $\mathbf{u}^*$—while staying in $\mathcal{U}^{IR}$—$\Omega(\mathbf{u})$ increases. Because $\Omega(\cdot)$ is convex, any such movement away from $\mathbf{u}^*$ must eventually (i.e., for finite $\mathbf{u}$) lead to a $\Omega(\mathbf{u})>\Omega\left(\mathbf{u}^0\right)$ (convex functions are unbounded above). Hence $\mathcal{V}$ is bounded.

is a standard nonlinear programming problem: Minimize a convex function (i.e., $\sum_{n=1}^{N} f_n(\hat{a}) u^{-1}(u_n)$) subject to $J$ constraints (one individual rationality constraint and $J-1$ incentive compatibility constraints, one for each action other than $\hat{a}$). If we further assume, as we do henceforth, that $u(\cdot)$ is differentiable, then the standard Lagrange-multiplier techniques can be employed. Specifically, let $\lambda$ be the Lagrange multiplier on the IR constraint and let $\mu_j$ be the Lagrange multiplier on the IC constraint between $\hat{a}$ and $a_j$, where $j = 1, \ldots, J-1$ indexes the elements of $\mathcal{A}$ other than $\hat{a}$. It is readily seen that the first-order condition with respect to the contract are

$$\frac{f_n(\hat{a})}{u'[u^{-1}(u_n)]} - \lambda f_n(\hat{a}) - \sum_{j=1}^{J-1} \mu_j [f_n(\hat{a}) - f_n(a_j)] = 0;\ n = 1, \ldots, N.$$

We've already seen (Proposition 4) that the IR constraint binds, hence $\lambda > 0$. Because $\hat{a}$ is not a least-cost action and there is no shifting support, it is readily shown that at least one IC constraint binds (i.e., $\exists j$ such that $\mu_j > 0$). It's convenient to rewrite the first-order condition as

$$\frac{1}{u'[u^{-1}(u_n)]} = \lambda + \sum_{j=1}^{J-1} \mu_j \left(1 - \frac{f_n(a_j)}{f_n(\hat{a})}\right);\ n = 1, \ldots, N. \tag{28}$$

Note the resemblance between (28) and (9) in Section 2.4. The difference is that, now, we have more than one Lagrange multiplier on the actions (since we now have more than two actions). In particular, we can give a similar interpretation to the likelihood ratios, $f_n(a_j)/f_n(\hat{a})$, that we had in that earlier section; with the caveat that we now must consider more than one action.

## 4.4 Properties of the optimal contract

Having solved for the optimal contract, we can now examine its properties. In particular, we will consider three questions:

1. Under what conditions does the expected cost of implementing an action under the optimal contract for the hidden-action problem exceed the full-information cost of implementing that action?

2. Recall the performance measures, $x$, constitute distinct elements of a chain. Under what conditions is the agent's compensation increasing with the value of the signal (i.e., when does $x \prec x'$ imply $s(x) \leq s(x')$)?

3. Consider two principal-agent models that are identical except that the information structure (i.e., $\{\mathbf{f}(a) | a \in \mathcal{A}\}$) in one is more informative than the information structure in the other. How do the costs of implementing actions vary between these two models.

The answer to the first question is given by

**Proposition 11** *Consider a hidden-information problem. Assume there is no shifting support (i.e., $f_n(a) > 0$ for all $n$ and all $a$). Assume, too, that $u(\cdot)$ is strictly concave. If $\hat{a}$ is not a least-cost action, then it cannot be implemented at its full-information cost.*

---

**Proof:** If $\hat{a}$ is not implementable, then the result is obvious; hence, we'll assume $\hat{a}$ is implementable. Define $\mathcal{U}^{IR}$ to be the set of all contracts that satisfy the IR constraint for $\hat{a}$. Let $\mathbf{u}^*$ be the contract in which $u_n^* = U_R + c(\hat{a})$ for all $n$. Note $\mathbf{u}^* \in \mathcal{U}^{IR}$. Finally define,

$$\Omega(\mathbf{u}) = \sum_{n=1}^{N} f_n(\hat{a})\, u^{-1}(u_n)\,.$$

Because $u(\cdot)$ is strictly concave, the principal's expected cost if the agent chooses $\hat{a}$ under contract $\mathbf{u}$, $\Omega(\mathbf{u})$, is a strictly convex function of $\mathbf{u}$. By Jensen's inequality and the fact that there is no shifting support, $\Omega(\cdot)$, therefore, has a unique minimum in $\mathcal{U}^{IR}$, namely $\mathbf{u}^*$. Clearly, $\Omega(\mathbf{u}^*) = C^F(\hat{a})$. The result, then, follows if we can show that $\mathbf{u}^*$ is not incentive compatible. Given that $\hat{a}$ is not a least-cost action, there exists an $a$ such that $c(\hat{a}) > c(a)$. But

$$\mathbf{f}(a)^\top \mathbf{u}^* - c(a) = U_R + c(\hat{a}) - c(a) > U_R = \mathbf{f}(\hat{a})^\top \mathbf{u}^* - c(\hat{a})\,;$$

that is, $\mathbf{u}^*$ is not incentive compatible. ■

---

Note the elements that go into this proposition *if $\hat{a}$ is implementable:* There must be an agency problem—mis-alignment of interests (i.e., $\hat{a}$ is not least cost); there must, in fact, be a *significant* hidden-action problem (i.e., no shifting support); and the agent must be risk averse. We saw earlier that without any one of these elements, an implementable action is implementable at full-information cost (Propositions 6–8); that is, each element is individually necessary for cost to increase when we go from full information to a hidden action. This last proposition shows, *inter alia*, that they are collectively sufficient for the cost to increase.

Next we turn to the second question. We already know from our analysis of the two-action model that the assumptions we have so far made are *in*sufficient for us to conclude that compensation will be monotonic. From our analysis of that model, we might expect that we need some monotone likelihood ratio property. In particular, we assume

**MLRP:** Assume there is no shifting support. Then the monotone likelihood ratio property is said to hold if, for any $a$ and $\hat{a}$ in $\mathcal{A}$, $c(a) \leq c(\hat{a})$ implies that $f_n(a)/f_n(\hat{a})$ is nonincreasing in $n$.

Intuitively, MLRP is the condition that actions that the agent finds more costly be more likely to produce better performance.

Unlike the two-action case, however, MLRP is not sufficient for us to obtain monotone compensation (see Grossman and Hart, 1983, for an example in which MLRP is satisfied but compensation is non-monotone). We need an additional assumption:

**CDFP:** The agency problem satisfies the *concavity of distribution function property* if, for any $a$, $a'$, and $\hat{a}$ in $\mathcal{A}$,

$$c(\hat{a}) = \lambda c(a) + (1-\lambda) c(a') \quad \exists \lambda \in (0,1)$$

implies that $F(\cdot|\hat{a})$ first-order stochastically dominates $\lambda F(\cdot|a) + (1-\lambda) F(\cdot|a')$.[35] Another way to state the CDFP is that the distribution over performance is better—more likely to produce high signals—if the agent plays a pure strategy than it is if he plays any mixed strategy over two actions when that mixed strategy has the same expected disutility as the pure strategy.

We can now answer the second question:

**Proposition 12** *Assume there is no shifting support, that $u(\cdot)$ is strictly concave and differentiable, and that MLRP and CDFP are met. Then the optimal contract given the hidden-action problem satisfies $s_1 \leq \cdots \leq s_N$.*

---

**Proof:** Let $\hat{a}$ be the action the principal wishes to implement. If $\hat{a}$ is a least-cost action, then the result follows from Proposition 6; hence assume that $\hat{a}$ is not a least-cost action. Let $\mathcal{A}' = \{a | c(a) \leq c(\hat{a})\}$; that is, $\mathcal{A}'$ is the set of actions that cause the agent less disutility than $\hat{a}$. Consider the principal's problem of implementing $\hat{a}$ under the assumption that the space of contracts is $\mathcal{A}'$. By MLRP, $f_n(a)/f_n(\hat{a})$ is nonincreasing in $n$ for all $a \in \mathcal{A}'$, so it follows from (28) that $s_1 \leq \cdots \leq s_N$ under the optimal contract for this *restricted* problem. The result then follows if we can show that this contract remains optimal when we expand $\mathcal{A}'$ to $\mathcal{A}$—adding actions cannot reduce the cost of implementing $\hat{a}$, hence we are done if we can show that the optimal contract for the restricted problem is incentive compatible in the *un*restricted problem. That is, if there is *no* $a$, $c(a) > c(\hat{a})$, such that

$$\mathbf{f}(a)^\top \mathbf{u} - c(a) > \mathbf{f}(\hat{a})^\top \mathbf{u} - c(\hat{a}), \tag{29}$$

where $\mathbf{u} = (u(s_1), \ldots, u(s_N))^\top$. As demonstrated in the proof of Proposition 11, the incentive compatibility constraint between $\hat{a}$ and at least one $a' \in \mathcal{A}'$, $c(a') < c(\hat{a})$, is binding; i.e.,

$$\mathbf{f}(a')^\top \mathbf{u} - c(a') = \mathbf{f}(\hat{a})^\top \mathbf{u} - c(\hat{a}). \tag{30}$$

Because $c(\hat{a}) \in (c(a'), c(a))$, there exists a $\lambda \in (0,1)$ such that $c(\hat{a}) = (1-\lambda) c(a') + \lambda c(a)$. Using CDFP and the fact that $u(s_1) \leq \cdots \leq u(s_N)$, we have

$$\begin{aligned} \mathbf{f}(\hat{a})^\top \mathbf{u} - c(\hat{a}) &\geq (1-\lambda)\mathbf{f}(a')^\top \mathbf{u} + \lambda \mathbf{f}(a)^\top \mathbf{u} - c(\hat{a}) \\ &= (1-\lambda)\left[\mathbf{f}(a')^\top \mathbf{u} - c(a')\right] + \lambda\left[\mathbf{f}(a)^\top \mathbf{u} - c(a)\right]. \end{aligned}$$

But this and (30) are inconsistent with (29); that is, (29) cannot hold, as was required. ■

---

Lastly, we come to question 3. An *information structure* for a principal-agent problem is $\mathbf{F} \equiv \{\mathbf{f}(a) | a \in \mathcal{A}\}$. A principal-agent problem can, then, be summarized as $\mathfrak{P} = \langle \mathcal{A}, \mathcal{X}, \mathbf{F}, B(\cdot), c(\cdot), u(\cdot), U_R \rangle$.

[35] Recall that distribution $G(\cdot)$ *first-order stochastically dominates* distribution $H(\cdot)$ if $G(z) \leq H(z)$ for all $z$ and strictly less than for some $z$.

**Proposition 13** *Consider two principal-agent problems that are identical except for their information structures (i.e., consider*

$$\mathfrak{P}^1 = \left\langle \mathcal{A}, \mathcal{X}, \mathbf{F}^1, B(\cdot), c(\cdot), u(\cdot), U_R \right\rangle$$

*and*

$$\mathfrak{P}^2 = \left\langle \mathcal{A}, \mathcal{X}, \mathbf{F}^2, B(\cdot), c(\cdot), u(\cdot), U_R \right\rangle).$$

*Suppose there exists a stochastic transformation matrix* $\mathbf{Q}$ *(i.e., a* garbling*)*[36] *such that* $\mathbf{f}^2(a) = \mathbf{Q}\mathbf{f}^1(a)$ *for all* $a \in \mathcal{A}$*, where* $\mathbf{f}^i(a)$ *denotes an element of* $\mathbf{F}^i$*. Then, for all* $a \in \mathcal{A}$*, the principal's expected cost of optimally implementing action* $a$ *in the first principal-agent problem,* $\mathfrak{P}^1$*, is* not *greater than her expected cost of optimally implementing* $a$ *in the second principal-agent problem,* $\mathfrak{P}^2$*.*

---

**Proof:** Fix $a$. If $a$ is not implementable in $\mathfrak{P}^2$, then the result follows immediately. Suppose, then, that $a$ is implementable in $\mathfrak{P}^2$ and let $\mathbf{u}^2$ be the optimal contract for implementing $a$ in that problem. Consider the contract $\mathbf{u}^1 = \mathbf{Q}^\top \mathbf{u}^2$. We will show that $\mathbf{u}^1$ implements $a$ in $\mathfrak{P}^1$. Because

$$\mathbf{f}^1(a)^\top \mathbf{u}^1 = \mathbf{f}^1(a)^\top \mathbf{Q}^\top \mathbf{u}^2 = \mathbf{f}^2(a)^\top \mathbf{u}^2,$$

the fact that $\mathbf{u}^2$ satisfies IR and IC in $\mathfrak{P}^2$ can readily be shown to imply that $\mathbf{u}^1$ satisfies IR and IC in $\mathfrak{P}^1$. The principal's cost of *optimally* implementing $a$ in $\mathfrak{P}^1$ is no greater than her cost of implementing $a$ in $\mathfrak{P}^1$ using $\mathbf{u}^1$. By construction, $u_n^1 = \mathbf{q}_{\cdot n}^\top \mathbf{u}^2$, where $\mathbf{q}_{\cdot n}$ is the $n$th column of $\mathbf{Q}$. Because $s_n^i = u^{-1}\left(u_n^i\right)$ and $u^{-1}(\cdot)$ is convex, it follows from Jensen's Inequality that

$$s_n^1 \leq \sum_{m=1}^{N} q_{mn} s_m^2$$

(recall $\mathbf{q}_{\cdot n}$ is a probability vector). Consequently,

$$\sum_{n=1}^{N} f_n^1(a)\, s_n^1 \leq \sum_{n=1}^{N} f_n^1(a) \sum_{m=1}^{N} q_{mn} s_m^2 = \sum_{m=1}^{N} f_n^2(a)\, s_n^2.$$

The result follows. ∎

---

Proposition 13 states that if two principal-agent problems are the same, except that they have different information structures, where the information structure of the first problem is more informative than the information structure of the second problem (in the sense of Blackwell's theorem), then the principal's expected cost of optimally implementing any action is no greater in the first problem than in the second problem. By strengthening the assumptions slightly, we can, in fact, conclude that the principal's expected cost is strictly less in the first problem. In other words, making the signal more informative about the agent's action makes the principal better off. This is consistent with our earlier

[36] A stochastic transformation matrix, sometimes referred to as a garbling, is a matrix in which each column is a probability density (i.e., has non-negative elements that sum to one).

findings that (i) the value of the performance measures is solely their statistical properties as correlates of the agent's action; and (ii) the better correlates—technically, the more informative—they are, the lower the cost of the hidden-action problem.

It is worth observing that Proposition 13 implies that the optimal incentive scheme never entails paying the agent with lotteries over money (i.e., randomly mapping the realized performance levels via weights $\mathbf{Q}$ into payments).

## 4.5 A continuous performance measure

Suppose that $\mathcal{X}$ were a real interval—which, without loss of generality, we can take to be $\mathbb{R}$—rather than a discrete space and suppose, too, that $F(x|a)$ were a continuous and differentiable function with corresponding probability density function $f(x|a)$. How would this change our analysis? By one measure, the answer is not much. Only three of our proofs rely on the assumption that $\mathcal{X}$ is finite; namely the proofs of Propositions 5, 10 (and, there, only the existence part),[37] and 13. Moreover, the last of the three can fairly readily be extended to the continuous case. Admittedly, it is troubling not to have *general* conditions for implementability and existence of an optimal contract, but in many specific situations we can, nevertheless, determine the optimal contract.[38]

With $\mathcal{X} = \mathbb{R}$, the principal's problem—the equivalent of (19)—becomes

$$\min_{u(x)} \int_{-\infty}^{\infty} u^{-1}\left[u(x)\right] f(x|\hat{a})\, dx$$

subject to

$$\begin{aligned} \int_{-\infty}^{\infty} u(x) f(x|\hat{a})\, dx - c(\hat{a}) &\geq U_R; \text{ and} \\ \int_{-\infty}^{\infty} u(x) f(x|\hat{a})\, dx - c(\hat{a}) &\geq \int_{-\infty}^{\infty} u(x) f(x|a)\, dx - c(a) \quad \forall a \in \mathcal{A}. \end{aligned}$$

We know the problem is trivial if there is a shifting support, so assume the support of $x$, $\operatorname{supp}\{x\}$, is invariant with respect to $a$.[39] Assuming an optimal

[37] Where our existence proof "falls down" when $\mathcal{X}$ is continuous is that our proof relies on the fact that a continuous function from $\mathbb{R}^N \to \mathbb{R}$ has a minimum on a closed and bounded set. But, here, the contract space is no longer a subset of $\mathbb{R}^N$, but rather the space of all *functions* from $\mathcal{X} \to \mathbb{R}$; and there is no general result guaranteeing the existence of a minimum in this case.

[38] Page (1987) considers conditions for existence in this case (actually he also allows for $\mathcal{A}$ to be a continuous space). Most of the assumptions are technical, but not likely to be considered controversial. Arguably a problematic assumption in Page is that the space of possible contracts is constrained; that is, assumptions are imposed on an *endogenous* feature of the model, the contracts. In particular, if $\mathfrak{S}$ is the space of permitted contracts, then there exist $L$ and $M \in \mathbb{R}$ such that $L \leq s(x) \leq M$ for all $s(\cdot) \in \mathfrak{S}$ and all $x \in \mathcal{X}$. Moreover, $\mathfrak{S}$ is closed under the topology of pointwise convergence. On the other hand, it could be argued that range of real-life contracts must be bounded: Legal and other constraints on what payments the parties can make effectively limit the space of contracts to some set of bounded functions.

[39] That is,

$$\{x|f(x|a) > 0\} = \{x|f(x|a') > 0\}$$

contract exists to implement $\hat{a}$, that contract must satisfy the modified Borch sharing rule:

$$\frac{1}{u'\left[u^{-1}\left(u\left(x\right)\right)\right]} = \lambda + \sum_{j=1}^{J-1} \mu_j \left[1 - \frac{f\left(x|a_j\right)}{f\left(x|\hat{a}\right)}\right] \text{ for almost every } x \in \operatorname{supp}\left\{x\right\}.$$

Observe that this is just a variation on (9) or (28).

## 4.6 Bibliographic notes

Much of the analysis in this section has been drawn from Grossman and Hart (1983). In particular, they deserve credit for Propositions 4, 8, and 10–13 (although, here and there, we've made slight modifications to the statements or proofs). Proposition 5 is based on Hermalin and Katz (1991). The rest of the analysis represent well-known results.

# 5 Continuous Action Space

So far, we've limited attention to finite action spaces. Realistic though this may be, it can serve to limit the tractability of many models, particularly when we need to assume the action space is large. A large action space can be problematic for two, related, reasons. First, under the two-step approach, we are obligated to solve for the optimal contract for *each* $a \in \mathcal{A}$ (or at least each $a \in \mathcal{A}^I$) then, letting $C(a)$ be the expected cost of inducing action $a$ under its corresponding optimal contact, we next maximize $B(a) - C(a)$—expected benefit net expected cost. If $\mathcal{A}$ is large, then this is clearly a time-consuming and potentially impractical method for solving the principal-agent problem. The second reason a large action space can be impractical is because it can mean many constraints in the optimization program involved with finding the optimal contract for a given action (recall, e.g., that we had $J-1$ constraints—one for each action other than the given action). Again, this raises issues about the practicality of solving the problem.

These problems suggest that we would like a technique that allows us to solve program (15) on page 30,

$$\max_{(S(\cdot),a)} \int_{\mathcal{X}} W(S(x), x, b) dG(b, x|a) \tag{31}$$

subject to

$$a \in \arg\max_{a'} \int_{\mathcal{X}} U(S(x), x, a') dF(x|a') \tag{32}$$

and

$$\max_{a'} \int_{\mathcal{X}} U(S(x), x, a') dF(x|a') \geq U_R,$$

for all $a$ and $a'$ in $\mathcal{A}$.

directly, in a one-step procedure. Generally, to make such a maximization program tractable, we would take $\mathcal{A}$ to be a compact and continuous space (e.g., a closed interval on $\mathbb{R}$), and employ standard programming techniques. A number of complications arise, however, if we take such an approach.

Most of these complications have to do with how we treat the IC constraint, expression (32). To make life simpler, suppose that $\mathcal{A} = [\underline{a}, \bar{a}] \subset \mathbb{R}$, $\mathcal{X} = \mathbb{R}$, that $F(\cdot|a)$ is differentiable and, moreover, that the expression in (32) is itself differentiable for all $a \in \mathcal{A}$. Then, a natural approach would be to observe that *if* $a \in (\underline{a}, \bar{a})$ maximizes that expression, it must necessarily be the solution to the first-order condition to (32):

$$\int_{\mathcal{X}} \left(U_a\left[S(x), x, a\right] f(x|a) + U\left[S(x), x, a\right] f_a(x|a)\right) dx = 0 \tag{33}$$

(where subscripts denote partial derivatives). Conversely, if we knew that the *second*-order condition was also met, (33) would be equivalent to (32) and we could use it instead of (32)—*at least locally.* Unhappily, we don't, in general, know (i) that the second-order conditions is met and (ii) that, even if it is, the $a$ solving (33) is a global rather than merely local maximum. For many modeling problems in economics, we would avoid these headaches by simply assuming that (32) is globally strictly concave in $a$, which would ensure both the second-order condition and the fact that an $a$ solving (33) is a global maximum. We can't, however, do that here: The concavity of (32) will, in general, depend on $S(x)$; but since $S(\cdot)$ is *endogenous*, we can't make assumptions about it. If, then, we want to substitute (33) for (32), we need to look for other ways to ensure that (33) describes a global maximum. Not surprisingly, these ways are, in general, complicated and we direct the reader interested in a "general" approach to consider Rogerson (1985) and Jewitt (1988).

An additional complication arises with whether (31) also satisfies the properties that would allow us to conclude from first-order conditions that a global maximum has, indeed, been reached. Fortunately, in many problems, this issue is less severe because we typically impose the functional form

$$W(S(x), x, b) = x - S(x),$$

which gives the problem sufficient structure to allow us to validate a "first-order approach."

In the rest of this section, we develop a simple model in which a first-order approach is valid.

## 5.1 The first-order approach with a spanning condition

Assume, henceforth, that $\mathcal{A} = [\underline{a}, \bar{a}] \subset \mathbb{R}$, $\mathcal{X} = [\underline{x}, \bar{x}] \subset \mathbb{R}$, and that $F(\cdot|a)$ is differentiable. Let $f(\cdot|a)$ be the associated probability density function for each $a \in \mathcal{A}$. We further assume that

1. $U(S(x), x, a) = u[S(x)] - a$;

2. $u(\cdot)$ is strictly increasing and strictly concave;

3. the domain of $u(\cdot)$ is $(\underline{s}, \infty)$, $\lim_{s \downarrow \underline{s}} u(s) = -\infty$, and $\lim_{s \uparrow \infty} u(s) = \infty$;

4. $f(x|a) > 0$ for all $x \in \mathcal{X}$ and for all $a \in \mathcal{A}$ (i.e., there is no shifting support);

5. $F(x|a) = \gamma(a) F_H(x) + (1 - \gamma(a)) F_L(x)$ and $f(x|a) = \gamma(a) f_H(x) + (1 - \gamma(a)) f_L(x)$ for all $x$ and $a$, where $\gamma : \mathcal{A} \to [0,1]$ and $F_H(\cdot)$ and $F_L(\cdot)$ are distribution functions on $\mathcal{X}$;

6. $\gamma(\cdot)$ is strictly increasing, strictly concave, and twice differentiable; and

7. $f_L(x) / f_H(x)$ satisfies the MLRP (i.e., $f_L(x) / f_H(x)$ is non-increasing in $x$ and there exist $x'$ and $x''$ in $\mathcal{X}$, $x' < x''$, such that $f_L(x') / f_H(x') > f_L(x'') / f_H(x'')$).

Observe Assumptions 5–7 allow us, *inter alia*, to assume that $c(a) = a$ without loss of generality. Assumption 5 is known as a *spanning condition*. SPANNING CONDITION

In what follows, the following result will be critical:

**Lemma 2** *$F_H(\cdot)$ dominates $F_L(\cdot)$ in the sense of first-order stochastic dominance.*

---

**Proof:** We need to show $F_H(x) \leq F_L(x)$ for all $x \in \mathcal{X}$ and strictly less for some $x$. To this end, define

$$\Delta(z) = \int_{\underline{x}}^{z} [f_H(x) - f_L(x)] \, dx.$$

We wish to show that $\Delta(z) \leq 0$ for all $z \in \mathcal{X}$ and strictly less for some $z$. Observe that

$$\Delta(z) = \int_{\underline{x}}^{z} \left[1 - \frac{f_L(x)}{f_H(x)}\right] f_H(x) \, dx.$$

Let

$$\delta(x) = 1 - \frac{f_L(x)}{f_H(x)}.$$

By MLRP, $\delta(\cdot)$ is non-decreasing everywhere and increases at one $x$ at least. Because $\Delta(\bar{x}) = 0$, $f_H(\cdot) > 0$, $\bar{x} - \underline{x} > 0$, and $\delta(\cdot)$ is not constant, it follows that $\delta(\cdot)$ must be negative on some sub-interval of $\mathcal{X}$ and positive on some other. Because $\delta(\cdot)$ is non-decreasing, there must, therefore, exist an $\hat{x} \in (\underline{x}, \bar{x})$ such that

$$\begin{aligned} \delta(x) &\leq 0 \text{ for all } x < \hat{x} \text{ (and strictly less for } \underline{x} < x < x' \leq \hat{x}\text{, for some } x'\text{); and} \\ \delta(x) &\geq 0 \text{ for all } x > \hat{x} \text{ (and strictly greater for } \bar{x} > x > x'' \geq \hat{x}\text{, for some } x''\text{).} \end{aligned}$$

For $z \leq \hat{x}$, this implies that $\Delta(z) < 0$—it is the integral of a quantity that is negative over some range of the integral and never positive anywhere on that range. Finally, consider, $z \in (\hat{x}, \bar{x})$. $\Delta'(z) = \delta(z) f_H(z) \geq 0$ for all $z$ in that interval. Hence, because $\Delta(\hat{x}) < 0$ and $\Delta(\bar{x}) = 0$, it must be that $\Delta(z) \leq 0$ for all $z \in (\hat{x}, \bar{x})$. We've just shown that $\Delta(z) \leq 0$ for all $z \in \mathcal{X}$ and strictly less for some $z$, which yields the result. ■

---

A consequence of this Lemma is that if $\phi(\cdot)$ is an increasing function, then

$$\int_{\underline{x}}^{\bar{x}} \phi(x)\left[f_H(x) - f_L(x)\right] dx > 0. \tag{34}$$

It follows, then, that if $S(\cdot)$ (and, thus, $u[S(x)]$) is increasing, then (32) is globally concave in $a$. To see this, observe

$$\int_{\mathcal{X}} U(S(x), x, a) dF(x|a) = \int_{\underline{x}}^{\bar{x}} u[S(x)]\left[\gamma(a) f_H(x) + (1-\gamma(a)) f_L(x)\right] dx - a;$$

so we have

$$\frac{d}{da}\int_{\mathcal{X}} U(S(x), x, a) dF(x|a) = \int_{\underline{x}}^{\bar{x}} u[S(x)]\left[f_H(x) - f_L(x)\right]\gamma'(a)\, dx > 0$$

by (34) and the assumption that $\gamma'(\cdot) > 0$. Moreover,

$$\frac{d^2}{da^2}\int_{\mathcal{X}} U(S(x), x, a) dF(x|a) = \int_{\underline{x}}^{\bar{x}} u[S(x)]\left[f_H(x) - f_L(x)\right]\gamma''(a)\, dx < 0$$

by (34) and the assumption that $\gamma''(\cdot) < 0$. To summarize:

**Corollary 2** *If $S(\cdot)$ is increasing, then the agent's choice-of-action problem is globally concave. That is, we're free to substitute*

$$\int_{\underline{x}}^{\bar{x}} u[S(x)]\left[f_H(x) - f_L(x)\right]\gamma'(a)\, dx = 0 \tag{35}$$

*for (32).*

We'll now proceed as follows. We'll suppose that $S(\cdot)$ is increasing and we'll solve the principal's problem. Of course, when we're done, we'll have to double check that our solution indeed yields an increasing $S(\cdot)$. It will, but if it didn't, then our approach would be invalid. The principal's problem is

$$\max_{S(\cdot),a}\int_{\underline{x}}^{\bar{x}}\left[x - S(x)\right] f(x|a)\, dx$$

subject to (35) and the IR constraint,

$$\int_{\underline{x}}^{\bar{x}} u[S(x)]\, f(x|a)\, dx - a \geq U_R.$$

As we've shown many times now, this last constraint must be binding; so we have a classic constrained optimization program. Letting $\lambda$ be the Lagrange multiplier on the IR constraint and letting $\mu$ be the Lagrange multiplier on (35), we obtain the first-order conditions:

$$-f(x|a) + \mu u'[S(x)]\left[f_H(x) - f_L(x)\right]\gamma'(a) + \lambda u'[S(x)]\, f(x|a) = 0$$

differentiating by $S(x)$; and

$$\begin{aligned}&\int_{\underline{x}}^{\bar{x}}[x-S(x)][f_H(x)-f_L(x)]\gamma'(a)\,dx\\&+\mu\int_{\underline{x}}^{\bar{x}}u[S(x)][f_H(x)-f_L(x)]\gamma''(a)\,dx\\=\;&0\end{aligned}\tag{36}$$

differentiating by $a$ (there's no $\lambda$ expression in the second condition because, by (35), the derivative of the IR constraint with respect to $a$ is zero). We can rearrange the first condition into our familiar modified Borch sharing rule:

$$\begin{aligned}\frac{1}{u'[S(x)]}&=\lambda+\mu\frac{[f_H(x)-f_L(x)]\gamma'(a)}{\gamma(a)f_H(x)+(1-\gamma(a))f_L(x)}\\&=\lambda+\mu\frac{[1-r(x)]\gamma'(a)}{\gamma(a)[1-r(x)]+r(x)},\end{aligned}$$

where $r(x)=f_L(x)/f_H(x)$. Recall that $1/u'[\cdot]$ is an increasing function; hence, to test whether $S(\cdot)$ is indeed increasing, we need to see whether the right-hand side is *decreasing* in $r(x)$ since $r(\cdot)$ is decreasing. Straightforward calculations reveal that the derivative of the right-hand side is

$$\frac{-\gamma'(a)}{(r(x)+(1-r(x))\gamma(a))^2}<0.$$

We've therefore shown that $S(\cdot)$ is indeed increasing as required; that is, our use of (35) for (32) was valid.

Observe, from (36), that, because the *agent's* second-order condition is met, the first line in (36) must be positive; that is,

$$\int_{\underline{x}}^{\bar{x}}[x-S(x)][f_H(x)-f_L(x)]\,dx>0.$$

But this implies that, for this $S(\cdot)$, the *principal's* problem is globally concave in $a$:

$$\frac{d^2}{da^2}\int_{\underline{x}}^{\bar{x}}[x-S(x)]f(x|a)\,dx=\int_{\underline{x}}^{\bar{x}}[x-S(x)][f_H(x)-f_L(x)]\gamma''(a)\,dx<0.$$

Moreover, for any $S(\cdot)$, the principal's problem is (trivially) concave in $S(\cdot)$. Hence, we may conclude that the first-order approach is, indeed, valid for this problem.

Admittedly, the spanning condition is a fairly stringent condition; although it does have an economic interpretation. Suppose there are two distributions from which the performance measure could be drawn, "favorable" (i.e., $F_H(\cdot)$) and "unfavorable" (i.e., $F_L(\cdot)$). The harder—higher $a$—the agent chooses, the greater the probability, $\gamma(a)$, that the performance measure will be drawn from the favorable distribution. For instance, suppose there are two types of potential customers, those who tend to buy a lot—the $H$ type—and those who tend not to buy much—the $L$ type. By investing more effort, $a$, in learning his territory, the salesperson (agent) increases the probability that he will sell to $H$ types rather than $L$ types.

## 5.2 Bibliographic notes

The first papers to use the first-order approach were Holmström (1979) and Shavell (1979). Grossman and Hart (1983) was, in large part, a response to the potential invalidity of the first-order approach. Our analysis under the spanning condition draws, in part, from Hart and Holmström (1987).

# References

**Borch, Karl H.**, *The Economics of Uncertainty*, Princeton, NJ: Princeton University Press, 1968.

**Epstein, Larry G.**, "Behavior Under Risk: Recent Developments in Theory and Applications," in Jean-Jacques Laffont, ed., *Advances in Economic Theory: Sixth World Congress*, Vol. 2, Cambridge, England: Cambridge University Press, 1992.

**Fleming, Wendell**, *Functions of Several Variables*, Berlin: Springer-Verlag, 1977.

**Grossman, Sanford J. and Oliver D. Hart**, "An Analysis of the Principal-Agent Problem," *Econometrica*, January 1983, *51* (1), 7–46.

**Hart, Oliver D. and Bengt Holmström**, "Theory of Contracts," in Truman Bewley, ed., *Advances in Economic Theory: Fifth World Congress*, Cambridge, England: Cambridge University Press, 1987.

**Hermalin, Benjamin E. and Michael L. Katz**, "Moral Hazard and Verifiability: The Effects of Renegotiation in Agency," *Econometrica*, November 1991, *59* (6), 1735–1753.

**Holmström, Bengt**, "Moral Hazard and Observability," *Bell Journal of Economics*, Spring 1979, *10* (1), 74–91.

**Jewitt, Ian**, "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, September 1988, *56* (5), 1177–1190.

**Macho-Stadler, Inés and David Pérez-Castrillo**, *An Introduction to the Economics of Information: Incentives & Contracts*, Oxford: Oxford University Press, 1997.

**Page, Frank H.**, "The Existence of Optimal Contracts in the Principal-Agent Model," *Journal of Mathematical Economics*, 1987, *16* (2), 157–167.

**Rabin, Matthew**, "Risk Aversion, Diminishing Marginal Utility, and Expected-Utility Theory: A Calibration Theorem," 1997. mimeo, Department of Economics, University of California at Berkeley.

**Rockafellar, R. Tyrrell**, *Convex Analysis*, Princeton, NJ: Princeton University Press, 1970.

**Rogerson, William P.**, "The First-Order Approach to Principal-Agent Problems," *Econometrica*, November 1985, *53* (6), 1357–1368.

**Salanié, Bernard**, *The Economics of Contracts: A Primer*, Cambridge, MA: MIT Press, 1997.

**Sappington, David E. M.**, "Limited Liability Contracts between Principal and Agent," *Journal of Economic Theory*, February 1983, *29* (1), 1–21.

**Shavell, Steven**, "Risk Sharing and Incentives in the Principal and Agent Relationship," *Bell Journal of Economics*, Spring 1979, *10* (1), 55–73.

**van Tiel, Jan**, *Convex Analysis*, New York: John Wiley and Sons, 1984.