Hidden-Information Agency

Bernard Caillaud   Benjamin E. Hermalin

March 2000

Contents

1 Introduction

Our purpose is to consider the problem of hidden information; that is, a game between two economic actors, one of whom possesses *mutually* relevant information that the other does not. This is a common situation: The classic example being the "game" between a monopolist, who doesn't know the consumer's willingness to pay, and the consumer, who obviously does. Within the realm of contract theory, relevant situations include a seller who is better informed than a buyer about the cost of producing a specific good; an employee who alone knows the difficulty of completing a task for his employer; a divisional manager who can conceal information about his division's investment opportunities from headquarters; and a leader with better information than her followers about the value of pursuing a given course of action. In each of these situations, having private information gives the player possessing it a potential strategic advantage in his dealings with the other player. For example, consider a seller who has better information about his costs than his buyer. By behaving as if he had high costs, the seller can seek to induce the buyer to pay him more than she would if she knew he had low costs. That is, he has an incentive to use his superior knowledge to capture an *"information rent."* Of course, the buyer is aware of this possibility; so, if she has the right to propose the contract between them, she will propose a contract that works to reduce this information rent. Indeed, how the contract proposer—the principal—designs contracts to mitigate the informational *dis*advantage she faces will be a major focus of this reading. INFORMATION RENT

Not surprisingly, given the many applications of the screening model, our coverage of it cannot hope to be fully original.[1] Indeed, while there are idiosyncratic aspects to our approach, our treatment is quite standard.

2 The Basics of Contractual Screening

Let us begin by broadly describing the situation in which we are interested. We shall fill in the blanks as we proceed through this reading.

- Two players are involved in a strategic relationship; that is, each player's well being depends on the play of the other player.
- One player is better informed (or will become better informed) than the other; that is, he has *private information* about some state of nature relevant to the relationship. As is typical in information economics, we refer to the player with the private information as the *informed player* and the player without the private information as the *uninformed player*. PRIVATE INFORMATION INFORMED PLAYER UNINFORMED PLAYER
- Critical to our analysis of these situations is the bargaining game that determines the contract. We will refer to the contract proposer as the

[1] The books by Laffont and Tirole (1993), Salanié (1997), and Macho-Stadler and Pérez-Castrillo (1997) include similar chapters, although the emphasis varies widely among them. Surveys have also appeared in journals (see, *e.g.,* Caillaud et al., 1988).

principal and the player who receives the proposal as the *agent.* Moreover, we assume contracts are proposed on a take-it-or-leave-it basis: The agent's only choices are to accept or reject the contract proposed by the principal. Rejection ends the relationship between the players. A key assumption is that the principal is the *un*informed player. Models like this, in which the uninformed player proposes the contract, are referred to as *screening models.* In contrast, were the informed player the contract proposer, we would have a type of *signaling model.* PRINCIPAL AGENT SCREENING SIGNALING

- A contract can be seen as setting the rules of a secondary game to be played by the principal and the agent.

We presume that the *asymmetry* of information that exists in this game results because prior experience or expertise, location, or other factors give the agent free access to information about the state of nature; while the absence of expertise, different experience or location, or other factors exclude the principal from this information (make it prohibitively expensive for her to acquire it). For example, past jobs may tell a seller how efficient he is—and thus what his costs will be—while ignorance of these past jobs means the buyer has a less precise estimate of what his costs will be. We assume that the reason for this asymmetry of information is *exogenous.* In particular, the informed player is simply assumed to be endowed with his information for the purpose of the situation we wish to model. Here, we assume that only one player is better informed; that is, we are ruling *out* situations where *each* player has his or her own private information.[2]

Given this information structure, the two parties interact according to some specified rules that constitute the extensive form of a game. In this two-person game, the players must contract with each other to achieve some desired outcome. In particular, there is no ability to rely on some exogenously fixed and anonymous market mechanism. Our focus will be on instances of the game where the informed player can potentially benefit from his informational advantage (*e.g.,* perhaps inducing a buyer to pay more for a good than necessary because she fears the seller is high cost). But, because the informed player doesn't have the first move—the uninformed player gets to propose the contract—this informational advantage is not absolute: Through her design of the contract, the uninformed player will seek to offset the informed player's inherent advantage.

3 The Two-Type Screening Model

We will begin to formalize these ideas in as simple a model as possible, namely the *two-type model.* In the two-type model, the state of nature can take one of two possible values. As is common in this literature, we will refer to the realized state of nature as the agent's *type.* Given that there are only two possible state, TWO-TYPE MODEL TYPE

[2]Put formally, the uninformed player's information partition is coarser than the informed player's information partition.

the agent can have one of just two types.

Before proceeding, however, we need to emphasize that such simplicity in modeling is not without cost. The two-type model is "treacherous," in so far as it may suggest conclusions that seem general, but that are not. For example, the conclusion that we will shortly reach with this model that the optimal contract implies distinct outcomes for distinct states of nature—a result called *separation* —is not as general as it may seem. Moreover, the assumption of two types conceals, in essence, a variety of assumptions that must be made clear. It similarly conceals the richness of the screening problem in complex, more realistic, relationships. Our view is that few *economic* prescriptions and predictions should be reached from considering just the two-type model. Keeping this admonition in mind, we now turn to a simple analysis of private procurement in a two-type model. SEPARATION

3.1 A simple two-type screening situation

A large retailer (the principal) wishes to purchase units of some good for resale. Assume its size gives it all the bargaining power in its negotiations with the one firm capable of supplying this product (the agent). Let $x \in \mathbb{R}_+$ denote the units of this good and let $r(x)$ denote the retailer's revenues from x units.[3] Assume that $r(\cdot)$ is strictly concave and differentiable everywhere. Assume, too, that $r'(0) > 0$. (Because $r(\cdot)$ is a revenue function, $r(0) = 0$.)

The retailer is uncertain about the efficiency of the supplier. In particular, the retailer knows an *in*efficient supplier has production costs of $C_I(x)$, but an efficient supplier has production costs of $C_E(x)$. Let the retailer's prior belief be that the supplier is inefficient with probability f, where—reflecting its uninformed status—$0 < f < 1$. The supplier, in contrast, knows its *type*; that is, whether it is efficient or not.

Assume $C_t(\cdot)$ is increasing, everywhere differentiable, and convex for both types, t. (Because $C_t(\cdot)$ is a cost function, we know $C_t(0) = 0$.) Consistent with the ideas that the two types correspond to different levels of efficiency, we assume $C'_I(x) > C'_E(x)$ for all $x > 0$—the inefficient type's marginal-cost schedule lies above the efficient type's. Observe, necessarily then, that $C_I(x) > C_E(x)$ for all $x > 0$.

The retailer and the supplier have to agree on the quantity, x, of the good to trade and on a payment, s, for this *total* quantity. Please note that s is *not* the per-unit price, but the payment for *all* x units. Profits for retailer and supplier are, then, $r(x) - s$ and $s - C_t(x)$ respectively. The retailer makes a take-it-or-leave-it offer, which the supplier must either accept or refuse. If the supplier refuses the offer, there is no trade and each firm's payoff from this "transaction" is zero. This outcome, no agreement, is equivalent to agreeing to

[3]For convenience, we'll assume that the retailer incurs no costs other than those associated with acquiring the x units from the supplier. Alternatively, we could simply imagine that these other costs have been subtracted from revenues, so that $r(x)$ is profit gross of the cost of purchasing the x units.

trade 0 units for a 0 payment. Hence, it is without loss of generality to assume that the parties must reach some agreement in equilibrium.

We begin our analysis with the *benchmark* case of symmetric or full information. That is, for the moment, we'll assume the retailer knows the supplier's type (*i.e.,* $f = 0$ or $f = 1$). We may immediately characterize the Pareto optimal allocation: x_t^F units are traded, where

$$x_t^F = \arg\max_{x \geq 0} \{r(x) - C_t(x)\}.$$

Since the problem is otherwise uninteresting if trade is never desirable, let's assume that $r'(0) > C_E'(0)$ so that, with an efficient supplier at least, some trade is desirable. Pareto optimality, also referred to as *ex post efficiency* , then reduces to EX POST EFFICIENCY

$$r'(x_E^F) = C_E'(x_E^F) \text{ and } \left[r'\left(x_I^F\right) - C_I'\left(x_I^F\right)\right] x_I^F = 0$$

(where we take the larger non-negative root of the second equation). Because it is optimal to produce a greater amount when marginal costs are lower, our assumptions about $C_t'(\cdot)$ imply $0 \leq x_I^F < x_E^F$. In making its contract offer, the retailer sets $x = x_t^F$ and it offers a payment, s_t^F, no larger than necessary to induce the supplier to accept; that is, s_t^F satisfies

$$s_t^F - C_t(x_t^F) = 0.$$

3.2 Contracts under incomplete information

This symmetric information solution collapses when the retailer is uninformed about the state of nature. To see why, suppose that the retailer offered the supplier its choice of $\langle x_E^F, s_E^F \rangle$ or $\langle x_I^F, s_I^F \rangle$ with the expectation that the supplier would choose the one appropriate to its type (*i.e.,* the first contract if it were efficient and the second if it were not). Observe that the retailer is relying on the supplier to honestly disclose its type. Suppose, moreover, that the true state of nature is E. By truthfully revealing that the state of nature is E, the supplier would just be compensated for its cost of supplying x_E^F units; that is, it would earn a profit of $s_E^F - C_E(x_E^F) = 0$. On the other hand, if the supplier pretends to have high costs—claims the state of nature is I—it receives compensation s_I^F, while incurring cost $C_E(x_I^F)$ for supplying x_I^F units. This yields the supplier a profit of

$$\begin{aligned} s_I^F - C_E\left(x_I^F\right) &= \\ C_I(x_I^F) - C_E(x_I^F) &> 0 \end{aligned}$$

(recall $s_I^F = C_I\left(x_I^F\right)$). Clearly, then, the efficient-type supplier cannot be relied on to honestly disclose its type.

This difference or profit, $C_I(x_I^F) - C_E(x_I^F)$, which motivates the supplier to lie, is called an *information rent*. This is a loss to the retailer but a gain to the supplier. There is, however, an additional loss suffered by the retailer that INFORMATION RENT is *not* recaptured by the supplier: Lying means inefficiently little is produced; that is, a real deadweight loss of

$$\left[r\left(x_E^F\right) - C_E\left(x_E^F\right)\right] - \left[r\left(x_I^F\right) - C_E\left(x_I^F\right)\right]$$

is suffered.

Given this analysis, it would be surprising if the retailer would be so naïve as to rely on the supplier to freely reveal its type. In particular, we would expect the retailer to seek a means of improving on this ex post *in*efficient outcome by devising a more sophisticated contract. What kind of contracts can be offered? The retailer does *not* know the supplier's level of efficiency, so it may want to delegate the choice of quantity to the supplier under a *payment schedule* that implicitly rewards the supplier for not pretending its costs are high when they are truly low. This payment schedule, $S(\cdot)$, specifies what payment, $s = S(x)$, is to be paid the supplier as a function of the units, x, *it chooses* to supply. Wilson (1993) provides evidence that such payment schedules are common in real-world contracting.

PAYMENT SCHEDULE

If the supplier accepts such a contract, the supplier's choice of quantity, x_t, is given by

$$x_t \in \arg\max_{x\geq 0} \left\{S(x) - C_t(x)\right\}. \tag{1}$$

Assume for the moment that this program has a unique solution. Let u_t denote the value of this maximization program and let $s_t = S(x_t)$ be the supplier's payment under the terms of the contract. By definition,

$$u_t = s_t - C_t(x_t).$$

Observe that this means we can write the equilibrium payment, s_t, as

$$s_t = u_t + C_t\left(x_t\right).$$

We also define

$$R\left(\cdot\right) = C_I\left(\cdot\right) - C_E\left(\cdot\right)$$

as the *information-rent function.* Our earlier assumptions imply that $R\left(\cdot\right)$ is positive for $x > 0$, zero for $x = 0$, and strictly increasing.

RENT FUNCTION

Revealed preference in the choice of x necessarily implies the following about x_I and x_E:

$$\begin{aligned} u_E &= s_E - C_E(x_E) \geq s_I - C_E(x_I) = u_I + R(x_I) & (2)\\ u_I &= s_I - C_I(x_I) \geq s_E - C_I(x_E) = u_E - R(x_E). & (3) \end{aligned}$$

These inequalities are referred by many names in the literature: *incentive-compatibility* constraints, *self-selection* constraints, *revelation* constraints, and *truth-telling* constraints. Regardless of name, they simply capture the requirement that (x_I, s_I) and (x_E, s_E) be the preferred choices for the supplier in states I and E, respectively.

INCENTIVE CONSTRAINT
SELF-SELECTION
REVELATION
TRUTH-TELLING

What can we conclude from expressions (2) and (3)? First, rewriting them as

$$R(x_I) \leq u_E - u_I \leq R(x_E), \tag{4}$$

it follows that

$$x_I \leq x_E, \tag{5}$$

because $R(\cdot)$ is strictly increasing. Observe, too, that expression (2) implies $u_E > u_I$ (except if $x_I = 0$, in which case we only know $u_E \geq u_I$). Finally, expressions (2) and (5) implies $s_E > s_I$ (unless $x_E = x_I$, in which case (2) and (3) imply $s_E = s_I$).

Of course the contract—payment schedule $S(\cdot)$—must be acceptable to the supplier, which means

$$u_L \geq 0; \text{ and} \tag{6}$$

$$u_H \geq 0. \tag{7}$$

If these did not both hold, then the contract would be rejected by one or the other or both types of supplier. The constraints (6) and (7) are referred to as the agent's *participation* or *individual-rationality* constraints. They simply state that, without any bargaining power, the supplier accepts a contract if and only if accepting does not entail suffering a loss.

PARTICIPATION
INDIVIDUALLY RATIONAL

The retailer's problem is to determine a price schedule $S(\cdot)$ that maximizes its *expected* profit ("expected" because, recall, it knows only the probability that a give type will be realized). Specifically, the retailer seeks to maximize

$$f \times [r(x_I) - s_I] + (1 - f) \times [r(x_E) - s_E] \text{; or, equivalently,}$$
$$f \times [r(x_I) - C_I(x_I) - u_I] + (1 - f) \times [r(x_E) - C_E(x_E) - u_E],$$

where (x_t, u_t) are determined by the supplier's optimization program (1) in response to $S(\cdot)$.

Observe that only two points on the whole price schedule enter the retailer's objective function: (x_I, s_I) and (x_E, s_E); or, equivalently, (x_I, u_I) and (x_E, u_E). The maximization of the principal's objectives can be performed with respect to just these two points provided that we can recover a general payment schedule afterwards such that the supplier would accept this schedule and choose the appropriate point for its type given this schedule. For this to be possible, we know that the self-selection constraints, (2) and (3), plus the participation constraints, (6) and (7), must hold.

In fact, the self-selection constraints and the participation constraints on (x_I, s_I) and (x_E, s_E) are necessary and sufficient for there to exist a payment schedule such that the solution to (1) for type t is (x_t, s_t). To prove this assertion, let (x_I, s_I) and (x_E, s_E) satisfy those constraints and construct the rest of the payment schedule as follows:

$$\begin{aligned} S(x) &= 0 && \text{if } 0 \leq x < x_I \\ &= s_I && \text{if } x_I \leq x < x_E \\ &= s_E && \text{if } x_E \leq x, \end{aligned}$$

when $0 < x_I < x_E$.[4] Given that $C_t(\cdot)$ is increasing in x, no supplier would ever choose an x other than 0, x_I, or x_E (the supplier's marginal revenue is zero except at these three points). The participation constraints ensure that (x_t, s_t) is (weakly) preferable to $(0, 0)$ and the self-selection constraints ensure that a type-t supplier prefers (x_t, s_t) to $(x_{t'}, s_{t'})$, $t \neq t'$. That is, we've shown that faced with this schedule, the type-I supplier's solution to (1) is (x_I, s_I)—as required—and that the type-E supplier's solution to (1) is (x_E, s_E)—as required.

The retailer's problem can thus be stated as

$$\max_{\{x_I, x_E, u_I, u_E\}} f \times [r(x_I) - C_I(x_I) - u_I] + (1 - f) \times [r(x_E) - C_E(x_E) - u_E] \quad (8)$$

subject to (2), (3), (6), and (7). Solving this problem using the standard Lagrangean method is straightforward, albeit tedious. Because, however, such a mechanical method provides little intuition, we pursue a different, though equivalent, line of reasoning.

- One can check that ignoring the self-selection constraints (treating them as *not* binding) leads us back to the symmetric-information arrangement; and we know that at least one self-selection constraint is then violated. We can, thus, conclude that in our solution to (8) at least one of the self-selection constraints is binding.

- The self-selection constraint in state E implies that: $u_E \geq R(x_I) + u_I \geq u_I$. Therefore, if the supplier accepts the contract in state I, it will also accept it in state E. We can, thus, conclude that constraint (7) is slack and can be ignored.

- It is, however, the case that (6) must be binding at the optimum: Suppose not, then we could lower both utility terms u_L and u_H by some $\varepsilon > 0$ without violating the participation constraints. Moreover, since the two utilities have been changed by the same amount, this can't affect the self-selection constraints. But, from (8), lowering the utilities raises the principal's profits—which means our "optimum" wasn't optimal.

- Using the fact that (6) is binding, expression (4)—the pair of self-selection constraints—reduces to

  $$R(x_I) \leq u_E \leq R(x_E).$$

  Given a pair of quality levels (x_I, x_E), the retailer wants to keep the supplier's rent as low as possible and will, therefore, choose to pay him the smallest possible information rent; that is, we can conclude that $u_E = R(x_I)$. The self-selection constraint (2) is, thus, slack, provided the necessary monotonicity condition (5) holds.

[4] If $x_I = 0$, then $s_I = 0$. If $x_I = x_E$, then $s_I = s_E$.

Plugging our findings, $u_I = 0$ and $u_E = R(x_I)$, into the retailer's objectives yields the following reduced program:

$$\max_{\{(x_I,x_E)|\ x_I \leq x_E\}} \{f \times [r(x_I) - C_I(x_I)] + (1-f) \times [r(x_E) - C_E(x_E) - R(x_I)]\}.$$

The solution is

$$x_E = x_E^F = \arg\max_{x \geq 0} \{r(x) - C_E(x)\} \quad (9)$$

$$x_I = x_I^*(f) \equiv \arg\max_{x \geq 0} \left\{ r(x) - C_I(x) - \frac{1-f}{f} R(x) \right\}. \quad (10)$$

The only step left is to verify that the monotonicity condition (5) is satisfied for these values. If we consider the last two terms in the maximand of (10) to be cost, we see that the effective marginal cost of output from the inefficient type is

$$C_I'(x) + \frac{1-f}{f} R'(x) > C_I'(x) > C_E'(x)$$

for $x > 0$.[5] The greater the marginal-cost schedule given a fixed marginal-revenue schedule, the less is traded; that is, it must be that $x_I^*(f) < x_E^F$—the monoticity condition (5) is satisfied.

It is worth summarizing the nature and properties of the optimal price schedule for the retailer to propose:

Proposition 1 *The optimal (non-linear) payment schedule for the principal induces two possible outcomes depending upon the state of nature such that:*

- *the supplier trades the ex post efficient quantity, x_E^F, when it is an efficient producer, but trades* less than *the efficient quantity when it is an inefficient producer (i.e., $x_I^*(f) < x_I^F$);*
- *an inefficient supplier makes no profit ($u_I = 0$), but an efficient supplier earns an information rent of $R[x_I^*(f)]$;*
- *the revelation constraint is binding in state E, slack in state I;*
- *the participation constraint is binding in state I, slack in state E;*
- *$x_I^*(f)$ and $R[x_I^*(f)]$ are non-decreasing in the probability of drawing an inefficient producer (i.e., are non-decreasing in f);*
- *and, finally, $\lim_{f\downarrow 0} x_I^*(f) = 0$, $\lim_{f\uparrow 1} x_I^*(f) = x_I^F$, $\lim_{f\downarrow 0} R[x_I^*(f)] = 0$, and $\lim_{f\uparrow 1} R[x_I^*(f)] = R(x_I^F)$.*

[5] Since $x_E^F > 0$, this is the relevant domain of output to consider.

To see that the last two points hold, note first that the effective marginal cost of production from an inefficient supplier,

$$C_I'(x) + \frac{1-f}{f} R'(x),$$

is falling in f. By the usual comparative statics, this means that $x_I^*(f)$ is non-decreasing. Since $R(\cdot)$ is an increasing function, $R[x_I^*(f)]$ must be similarly non-decreasing. As $f \downarrow 0$, this effective marginal cost tends to $+\infty$ for $x > 0$, which means the optimal level of trade falls to zero. As $f \uparrow 1$, this effective marginal cost tends to the symmetric-information marginal cost, hence $x_I^*(f)$ tends to the symmetric-information level, x_I^F.

Intuition for these results can be gained from Figure 1. This figure shows one indifference curve for an inefficient (type-I) supplier and three indifference curves for an efficient (type-E) supplier in output-payment space. The type-I indifference curve is that type's zero-profit curve (hence, by necessity, it passes through the origin). Correspondingly, the lowest and darkest of the type-E indifference curves is that type's zero-profit curve. The faint dash-dot lines are iso-profit curves for the retailer (to minimize clutter in the figure, they're sketched as straight lines, but this is not critical for what follows). Observe that an iso-profit curve is tangent to type-I's zero-profit indifference curve at point A. Likewise, we have similar tangency for type-E at point B. Hence, *under symmetric information*, points A and B would be the contracts offered. Under asymmetric information, however, contract B is *not* incentive compatible for type-E: Were it to lie and claim to be type-I (*i.e.,* move to point A), then it would be on a higher (more profitable) indifference curve (the highest of its three curves). Under asymmetric information, an incentive compatible pair of contracts that induce the symmetric-information levels of trade are A and C. The problem with this solution, however, is that type-E earns a large information rent, equal to the distance between B and C. The retailer can reduce this rent by distorting downward the quantity asked from a type-I supplier. For example, by lowering quantity to $x_I^*(f)$, the retailer significantly reduces the information rent (it's now the distance between B and E). How much distortion in quantity the retailer will impose depends on the likelihood of the two types. When f is small, the expected savings in information rent is large, while the expected cost of too-little output is small, so the downward distortion in type-I's output is big. Conversely, when f is large, the expected savings are small and the expected cost is large, so the downward distortion is small. The exact location of point D is determined by finding where the expected marginal cost of distorting type-I's output, $f \times [r'(x_I) - C_I'(x_I)]$, just equals the expected marginal reduction in type-E's information rent, $(1-f) \times R'(x_I)$.

From Figure 1, it is clear that the retailer loses from being uninformed about the supplier's type: Point D lies on a worse iso-profit curve than does point A and point E lies on a worse iso-profit curve than does point B.[6] Put another way, if the retailer draws a type-I supplier, then it gets a non-optimal (relative

[6]Since the retailer likes more output (in the relevant range) and smaller payments to the

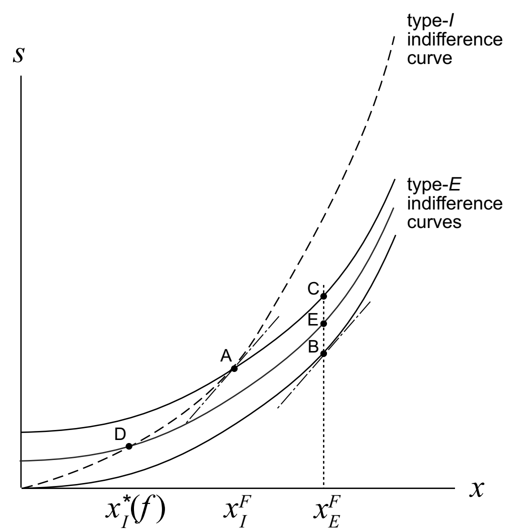

Figure 1: The symmetric-information contracts, points A and B, are *not* incentive compatible. The symmetric-information quantities, x_E^F and x_I^F, are too expensive because of the information rent (the distance from B to C). Consequently, with asymmetric information, the principal trades off a distortion in type-I's output (from x_I^F to $x_I^*(f)$) to reduce type-E's information rent (from BC to BE).

to symmetric information) quantity of the good. While if it draws a type-E supplier, then it pays more for the optimal quantity (again, relative to symmetric information). Part—but only part—of the retailer's loss is the supplier's gain. In expectation, the supplier's profit has increased by $(1-f)\,R\left[x_I^*(f)\right]$. But part of the retailer's loss is also deadweight loss: Trading $x_I^*(f)$ units instead of x_I^F units is simply inefficient. In essence, our retailer is a monopsonist and, as is typical of monopsony, there is a deadweight loss. Moreover, the deadweight loss arises here for precisely the same reason it occurs in monopsony: Like a monopsonist, our retailer is asking the payment (price) schedule to play two roles. First, it asking it to allocate goods and, second, it is asking it to preserve its rents from having all the bargaining power. Since only the first role has anything to do with allocative efficiency, giving weight to the second role can only create allocative *in*efficiency. As often happens in economics, a decision maker has one instrument—here, the payment schedule—but is asking it to serve multiple roles. Not surprisingly then, the ultimate outcome is less than first best.

Since the first best is *not* achieved, it is natural to ask whether the retailer

supplier, its utility (profits) are greater on iso-profit curves toward the southeast of the figure and less on iso-profit curves toward the northwest.

could improve on the outcome in Proposition 1 by using some more sophisticated contract? The answer is no and the proof is, as we will see later, quite general. Whatever sophisticated contract the retailer uses, this contract will boil down to a pair of points, (x_I, s_I) and (x_E, s_E), once it is executed; that is, a final quantity traded and a final payment for each possible state of nature. Consequently, whatever complicated play is induced by the contract, both parties can see through it and forecast that the equilibrium outcomes correspond to these two points. Moreover, by mimicking the strategy it would play in state t, the supplier can generate either of the two outcomes regardless of the true state. In addition, if it can't profit from (x_t, s_t) in state t, it can simply not participate. Necessary equilibrium conditions are, then, that the supplier choose (x_t, s_t) in state t rather than $(x_{t'}, s_{t'})$, $t \neq t'$, and that it choose to participate anticipating that it will choose (x_t, s_t) in state t. But these are precisely the revelation and participation constraints (2), (3), (6), and (7). Therefore, whatever the contractual arrangement, the final outcome can always be generated by a simple (non-linear) payment schedule like the one derived above. We've, thus, established that the outcome described in Proposition 1 cannot be improved on by using more sophisticated or alternative contracts.[7]

Finally, note that we don't need an entire payment schedule, $S(\cdot)$. In particular, there is a well-known alternative: a *direct-revelation contract* (mechanism). In a direct-revelation contract, the retailer commits to pay the supplier s_E for x_E or s_I for x_I depending on the supplier's announcement of its type. Failure by the supplier to announce its type (*i.e.,* failure to announce a $\hat{t} \in \{E, I\}$) is equivalent to the supplier rejecting the contract. Finally, if, after announcing its type, the supplier produces a quantity other than $x_{\hat{t}}$, the supplier is punished (*e.g.,* paid nothing). It is immediate that this direct-revelation contract is equivalent to the optimal payment schedule derived above. It is also simpler, in that it only deals with the relevant part of the payment schedule. Admittedly, it is not terribly realistic,[8] but as this discussion suggests we can transform a direct-revelation contract into a more realistic contract (indeed, we will formalize this below in Proposition 3). More importantly, as we will see, in terms of determining what is the optimal feasible outcome, there is no loss of generality in restricting attention to direct-revelation contracts.

direct revelation

4 General Screening Framework

The two-type screening model yielded strong results. But buried within it is a lot of structure and some restrictive assumptions. If we are really to use the screening model to understand economic relationships, we need to deepen our understanding of the phenomena it unveils, the assumptions they require, and

[7] This is not to say that *another* contract couldn't do as well. This is rather obvious: For instance, suppose $\tilde{S}(x) = 0$ for all x except x_E and x_I, where it equals s_E or s_I, respectively. We've merely established that no other contract can do strictly better than the $S(\cdot)$ derived in the text.

[8] Although see Gonik (1978) for a real-life example of a direct-revelation mechanism.

the robustness of its conclusions. Our approach in this section is, thus, to start from a very general formalization of the problem and to motivate or discuss the assumptions necessary for making this model "work."

A principal and an agent are involved in a relationship that can be characterized by an *allocation* $x \in \mathcal{X}$ and a real-valued monetary transfer $s \in \mathcal{S} \subset \mathbb{R}$ between the two players. A transfer-allocation pair, (x, s), is called an *outcome*. The space of possible allocations, $\mathcal{X}$, can be quite general: Typically, as in our analysis of the retailer-supplier problem, it's a subspace of $\mathbb{R}$; but it could be another space, even a multi-dimensional one. In what follows, we assume that outcomes are verifiable. ALLOCATION OUTCOME

The agent's information is characterized by a parameter $\theta \in \Theta$. As before, we'll refer to this information as the agent's *type*. The *type space*, Θ, can be very general. Typically, however, it is either a discrete set (*e.g.,* as in the retailer-supplier example where we had $\Theta = \{I, E\}$) or a compact interval in $\mathbb{R}$. Nature draws θ from Θ according to a commonly known probability distribution. While the agent learns the value of θ perfectly, the principal only knows that it was drawn from the commonly known probability distribution. TYPE SPACE

Both players' preferences are described by von Neumann-Morgenstern utility functions, $\mathcal{W}(x, s, \theta)$ for the principal and $\mathcal{U}(x,s,\theta)$ for the agent, where both are defined over $\mathcal{X} \times \mathcal{S} \times \Theta$. Since we interpret s as a transfer *from* principal *to* agent, we assume that $\mathcal{U}$ increases in s, while $\mathcal{W}$ decreases in s. For convenience, we assume these utility functions are smooth; more precisely, that they are three-times continuously differentiable.[9]

To have a fully general treatment, we need to extend this analysis to allow for the possibility that the actual outcome is chosen randomly from $\mathcal{X} \times \mathcal{S}$. To this end, let σ denote a generic element of the set of probability distributions, $\Delta(\mathcal{X} \times \mathcal{S})$, over the set of possible outcomes, $\mathcal{X} \times \mathcal{S}$. We extend the utility functions to $\Delta(\mathcal{X} \times \mathcal{S})$ through the expectation operator:

$$\mathcal{W}(\sigma, \theta) \equiv \mathbb{E}_\sigma \left[\mathcal{W}(x, s, \theta)\right] \text{ and } \mathcal{U}(\sigma, \theta) \equiv \mathbb{E}_\sigma \left[\mathcal{U}(x, s, \theta)\right].$$

By adding an element to $\mathcal{X}$ if necessary, we assume that there exists a no-trade outcome $(x_0, 0)$; that is, $(x_0, 0)$ is the outcome if no agreement is reached (*e.g.,* in the retailer-supplier example this was $(0, 0)$). The values of both $\mathcal{W}$ and $\mathcal{U}$ at this no-agreement point play an important role in what follows, so we give them special notation:

$$W_R(\theta) \equiv \mathcal{W}(x_0, 0, \theta) \text{ and } U_R(\theta) \equiv \mathcal{U}(x_0, 0, \theta).$$

These will be referred to as the Reservation utilities of the players (alternatively, their individual rationality payoffs). Obviously, an agent of type θ accepts a contract if and only if his utility from doing is not less than $U_R(\theta)$.

It is convenient, at this stage, to offer a formal and general definition of what a contract is:

[9] One could relax this smoothness assumption, but the economic value of doing so is too small to warrant time on this issue.

Table 1: The Retailer-Supplier Example in our General Notation

Description	General Notation	Specific Value
Allocation space	$\mathcal{X}$	$\mathbb{R}_+$
Transfer space	$\mathcal{S}$	$\mathbb{R}$
Outcome function	σ	All mass on $(x, S(x))$
Principal's strategy space	$\mathcal{N}$	$\{\text{pay } S(x)\}$
Agent's strategy space	$\mathcal{M}$	$\mathcal{X} = \mathbb{R}_+$

Definition 1 *A contract in the static contractual screening model is a game form, $\langle \mathcal{M}, \mathcal{N}, \sigma \rangle$, to be played by the principal and the agent, $\mathcal{M}$ denotes the agent's strategy set, $\mathcal{N}$ the principal's strategy set, and σ an outcome function that maps any pair of strategies (m, n) to a probability mapping on $\mathcal{X} \times \mathcal{S}$. That is, $\sigma : \mathcal{M} \times \mathcal{N} \to \Delta(\mathcal{X} \times \mathcal{S})$.*

To make this apparatus somewhat more intuitive, consider Table 1, which "translates" our retailer-supplier example into this more general framework. Observe that, in the example, the contract fixes a trivial strategy space for the principal: She has no discretion, she simply pays $S(x)$.[10] Moreover, there is no randomization in that example: σ simply assigns probability one to the pair $(x, S(x))$. As this discussion suggests, generality in notation need not facilitate understanding—the generality contained here is typically greater than we need. Fortunately, we will be able to jettison much of it shortly.

A *direct mechanism* is a mechanism in which $\mathcal{M} = \Theta$; that is, the agent's action is limited to making announcements about his type. The physical consequences of this announcement are then built into the outcome function, σ. For instance, as we saw at the end of the previous section, we can translate our retailer-supplier contract into a direct mechanism: Now $\mathcal{M} = \{E, I\}$, $\mathcal{N}$ is a singleton (so we can drop n as an argument of σ), and DIRECT MECHANISM

$$\sigma(m) = \begin{cases} (x_I, s_I) = (x_I^*(f), C_I[x_I^*(f)]) & \text{if } m = I \\ (x_E, s_E) = \left(x_E^F, R[x_I^*(f)] + C_E\left(x_E^F\right)\right) & \text{if } m = E \end{cases}.$$

A *direct-revelation mechanism* (alternatively, a direct truthful mechanism) is a direct mechanism where it is an equilibrium strategy for the agent to tell the truth: Hence, if $m(\cdot) : \Theta \to \Theta$ is the agent's strategy, we have $m(\theta) = \theta$ in equilibrium for all $\theta \in \Theta$. That is, for any θ and θ' in Θ, DIRECT-REVELATION

$$\mathcal{U}(\sigma(m[\theta]), \theta) \geq \mathcal{U}(\sigma(m[\theta']), \theta).$$

Note that not every direct mechanism will be a direct-revelation mechanism. Being truthful in equilibrium is a property of a mechanism; that is, it depends on $\sigma(\cdot)$.

[10] We could, alternatively, expand her strategy space to $\mathcal{S}$—she can attempt to pay the agent whatever she wants. But, then, the outcome function would have to contain a punishment for not paying the agent appropriately: That is, $\sigma(x, s) = (x, S(x))$ if $s = S(x)$ and equals (x, ∞) if $s \neq S(x)$ (where ∞ is shorthand for some large transfer sufficient to deter the principal from not making the correct payment).

Observe that the design of a contract means choosing $\mathcal{M}$, $\mathcal{N}$, and σ. In theory, the class of spaces and outcome functions is incomprehensibly large. How can we find the optimal contract in such a large class? Indeed, given the inherent difficulties in even characterizing such a large class, how can we ever be sure that we've found the optimal contract? Fortunately, two simple, yet subtle, results—*the revelation principle* and *the taxation principle*—allow us to avoid these difficulties.[11] From the revelation principle, the search for on optimal contract reduces *without loss of generality* to the search for the optimal direct-revelation mechanism. Moreover, if the outcome in the direct-revelation mechanism is a *deterministic* function of the agent's announcement, then, from the taxation principle, we may further restrict attention to a payment schedule that is a function of the allocation x (as we did in the retailer-supplier example).

Proposition 2 ***(The revelation principle)***[12] *For any general contract* $(\mathcal{M}, \mathcal{N}, \sigma)$ *and associated Bayesian equilibrium, there exists a direct-revelation mechanism such that the associated truthful Bayesian equilibrium generates the same equilibrium outcome as the general contract.*

Proof: The proof of the revelation principle is standard but informative. A Bayesian equilibrium of the game $(\mathcal{M}, \mathcal{N}, \sigma)$ is a pair of strategies $(m(\cdot), n)$.[13] Let us consider the following direct mechanism: $\hat{\sigma}(\cdot) = \sigma(m(\cdot), n)$. Our claim is that $\hat{\sigma}(\cdot)$ induces truth-telling (is a direct-*revelation* mechanism). To see this, suppose it were not true. Then there must exist a type θ such that the agent does better to lie—announce some $\theta' \neq \theta$—when he is type θ. Formally, there must exist θ and $\theta' \neq \theta$ such that

$$\mathcal{U}(\hat{\sigma}(\theta'), \theta) > \mathcal{U}(\hat{\sigma}(\theta), \theta).$$

Using the definition of $\hat{\sigma}(\cdot)$, this means that

$$\mathcal{U}(\sigma[m(\theta'), n], \theta) > \mathcal{U}(\sigma[m(\theta), n], \theta);$$

but this means the agent prefers to play $m(\theta')$ instead of $m(\theta)$ in the *original* mechanism against the principal's equilibrium strategy n. This, however, can't be since $m(\cdot)$ is an equilibrium best response to n in the original game. Hence, truthful revelation must be an optimal strategy for the agent under the constructed direct mechanism. Finally, when the agent truthfully reports the state of nature in the direct truthful mechanism, the same outcome $\hat{\sigma}(\theta) = \sigma(m(\theta), n)$

[11] It is unfortunate that these two fundamental results are called principles, since they are not, as their names might suggest, premises or hypotheses. They are, as we will show, deductive results.

[12] The revelation principle is often attributed to Myerson (1979), although Gibbard (1973) and Green and Laffont (1977) could be identified as earlier derivations. Suffice it to say that the revelation principle has been independently derived a number of times and was a well-known result before it received its name.

[13] Observe that the agent's strategy can be conditioned on θ, which he knows, while the principal's cannot be (since she is ignorant of θ).

is implemented in equilibrium. ■

An intuitive way to see the revelation principle is imagine that before he plays some general mechanism, the agent could delegate his play to some trustworthy third party. There are two *equivalent* ways this delegation could work. One, the agent could tell the third party to play m. Alternatively, if the third party knows the agent's equilibrium strategy—the mapping $m : \Theta \to \mathcal{M}$—then the agent could simply reveal (announce) his type to the third party with the understanding that the third party would choose the appropriate actions, $m(\theta)$. But, since we can build this third party into the design of our direct-revelation mechanism, this equivalence means that there is no loss of generality in restricting attention to direct-revelation mechanisms.

The taxation principle requires a little more structure: It assumes that there is a possibility of punishing the agent so as to deter him to violate the contractual rules. More specifically, let us consider the following assumption:

A0 *(Existence of a punishment): There exists an $\underline{s} \in \overline{\mathbb{R}}$ such that:*

$$\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \mathcal{U}(x,\underline{s},\theta) \leq \inf_{(s,x,\theta)\in\mathcal{X}\times\mathcal{S}\times\Theta} \mathcal{U}(x,s,\theta).$$

In other words, there exists a punishment so severe that the agent would always prefer not to suffer it.

With this assumption, one can construct a payment schedule (in $\overline{\mathbb{R}}$) that generates the same outcome as any *deterministic* direct-revelation mechanism and is, therefore, as general as any contract in this context.

Proposition 3 ***(The taxation principle)*** *Under Assumption A0, the equilibrium outcome under any deterministic direct-revelation mechanism, $\sigma(\cdot) = (x(\cdot), s(\cdot))$, is also an equilibrium outcome of the game where the principal proposes the payment schedule $S(\cdot)$ defined by*

$S(x) = s(\theta)$, *when $\theta \in x^{-1}(x)$ (i.e., such that $x = x(\theta)$ for some $\theta \in \Theta$)*
$S(x) = \underline{s}$, *otherwise.*

Proof: Let's first establish that $S(\cdot)$ is unambiguously defined: Suppose there existed θ_1 and θ_2 such that

$$x = x(\theta_1) = x(\theta_2),$$

but $s(\theta_1) \neq s(\theta_2)$. We're then free to suppose that $s(\theta_1) > s(\theta_2)$. Then, because $\mathcal{U}$ is increasing in s,

$$\mathcal{U}(x(\theta_1), s(\theta_1), \theta_2) > \mathcal{U}(x(\theta_2), s(\theta_2), \theta_2).$$

But this means the agent would prefer pretending that the state of nature is θ_2 when it's actually θ_1; the mechanism would *not* be truthful. Hence, if

$\theta_1, \theta_2 \in x^{-1}(x)$, we must have $s(\theta_1) = s(\theta_2)$; that is, the payment schedule $S(\cdot)$ is unambiguously defined.

Now, the agent's problem when faced with the payment schedule $S(\cdot)$ is simply to choose the allocation x that maximizes $\mathcal{U}(x, S(x), \theta)$. Given the severity of the punishment, the agent will restrict his choice to $x \in x(\Theta)$. But since our original mechanism was a direct-revelation mechanism, we know

$$\mathcal{U}(x(\theta), s(\theta), \theta) \geq \mathcal{U}(x(\theta'), s(\theta'), \theta)$$

for all θ and θ'. So no type θ can do better than to choose $x = x(\theta)$. ■

The economic meaning of the taxation principle is straightforward: When designing a contract, the principal is effectively free to focus on "realistic" compensation mechanisms that pay the agent according to his achievements. Hence, as we argued above, there is no loss of generality in our solution to the retailer-supplier problem.

Although payment schedules involve no loss of generality and are realistic, the fact that they are often nonlinear means that they can be difficult to work with.[14] In particular, when looking for the optimal non-linear price schedule, one must be able to compute the functional mapping that associates to each schedule $S(\cdot)$ the action choice $x(\theta)$ that maximizes $\mathcal{U}(x, S(x), \theta)$. Even assuming $S(\cdot)$ is differentiable—which is not ideal because one's not supposed to make assumptions about *endogenous* variables—solving this problem can be difficult. Direct-revelation mechanisms, on the other hand, allow an easier mathematical treatment of the problem using standard convex analysis: The revelation constraints simply consist of writing $\theta' = \theta$ is a maximum of $\mathcal{U}(x(\theta'), s(\theta'), \theta)$ and writing that a given point is a maximum is easier that characterizing an unknown maximum. For this reason, much of the mechanism-design literature has focussed on direct-revelation mechanisms over optimal payment schedules despite the latter's greater realism.

5 The Standard Framework

As we've already hinted, the general framework introduced in the previous section is more general than what is commonly used. In this section, we introduce a *"standard"* framework, within which much of the contractual screening literature can be placed. We begin by defining this framework and contrasting it to the more general framework introduced above. At the end of this reading, we offer less conventional views on the screening model, which require departing from the standard framework. In order to provide a road map to the reader, we spell out each assumption in this section and suggest what consequences would arise from alternative assumptions.

[14] Admittedly, if the derived payment schedule is sufficiently nonlinear, one could begin to question its realism, as real-world contracts are often linear or, at worst, piecewise linear. Although see footnote 8 for an example of where reality "matches" theory.

In the standard framework, the allocation space, $\mathcal{X}$, is $\mathbb{R}_+$.[15] The type space, Θ, is $[\theta_L, \theta_H] \subset \mathbb{R}$, where both bounds, θ_L and θ_H, are finite.[16] The most critical assumptions in going from the general framework to the standard framework involve the utility functions. Henceforth, we assume they are additively separable in the transfer and the allocation. Moreover, we assume the marginal value of money is type *in*dependent (*i.e.,* $\partial\left(\partial\mathcal{U}/\partial s\right)/\partial\theta = \partial\left(\partial\mathcal{W}/\partial s\right)/\partial\theta = 0$). These assumptions simplify the analysis by eliminating income effects from consideration. Formally, the agent's and principal's utility functions are

$$\begin{aligned}\mathcal{U}(x,s,\theta) &= s + u(x,\theta) \text{ and} \\ \mathcal{W}(x,s,\theta) &= w(x,\theta) - s,\end{aligned}$$

respectively. Essentially for convenience, we take $u(\cdot,\cdot)$ and $w(\cdot,\cdot)$ to be three-times continuously differentiable. The aggregate (full-information) surplus is defined as:

$$\Omega(x,\theta) = w(x,\theta) + u(x,\theta).$$

We also assume, mainly for convenience, that

1. **Some trade is desirable:** For all $\theta \in (\theta_L, \theta_H]$, $\partial\Omega\left(0,\theta\right)/\partial x > 0$.

2. **There can be too much of a good thing:** For all $\theta \in [\theta_L, \theta_H]$ $\exists\bar{x}\left(\theta\right)$ such that $\Omega\left(x,\theta\right) \leq \Omega\left(0,\theta\right)$ for all $x > \bar{x}\left(\theta\right)$.

Observe these two assumptions entail that $\Omega\left(x,\theta\right)$ has an interior maximum for all $\theta \in (\theta_L, \theta_H]$. If the first assumption didn't hold for at least some types, then trade—contracting—would be pointless. Extending the desirability of trade to almost all types saves from the bookkeeping headache of distinquishing between types with which trade is efficient and those with which it is not. The second assumption is just one of many ways of expressing the sensible economic idea that, beyond some point, welfare is reduced by trading more.

Observe that the standard framework is restrictive in several potential important ways:

- The type space is restricted to be one-dimensional. In many applications, such as the retailer-supplier example, this is a natural assumption. One can, however, conceive of applications where it doesn't fit: Suppose, *e.g.,* the retailer cared about the quantity and quality of the goods received and the supplier's type varied on both an efficiency dimension and a conscientiousness-of-employees dimension (the latter affecting the cost of providing quality). Not surprisingly, restricting attention to one dimension is done for analytic tractability: Assuming a single dimension

[15] That the space be bounded below at 0 is not critical—any lower bound would do. Alternatively, by appropriate changes to the utility functions, we could allow the allocation space to be unbounded. Zero is simply a convenience.

[16] This constitutes no loss of *economic* (as opposed to mathematical) generality, since we can set the bounds as far apart as we need.

make the order properties of the type space straightforward (*i.e.,* greater and less than are well-defined on the real line). As we will see, the order properties of the type space are critical to our analysis.

- The set of possible allocations is one-dimensional. Again, this is sufficient for some applications (*e.g.,* the quantity supplied to the retailer), but not others (*e.g.,* when the retailer cares about both quantity and quality). The difficulty in expanding to more than one dimension arise from difficulties in capturing how the agent's willingness to make tradeoffs among the dimensions (including his income) varies with his type. The reader interested in this extension should consult Rochet and Choné (1998).

- As noted, the utility functions are separable in money and allocation; the marginal utility of income is independent of the state of nature; and the marginal utility of income is constant, which means both players are risk neutral with respect to gambles over money. The gains from these assumptions are that we can compute the transfer function $s(\cdot)$ in terms of the allocation function $x(\cdot)$, which means our optimization problem is a standard optimal-control problem with a unique control, $x(\cdot)$. In addition, risk neutrality insulates us from problems that *exogenously* imposed risk might otherwise create (*e.g.,* the need to worry about mutual insurance). On the other hand, when the agent is risk averse, the ability to threaten him with *endogenously* imposed risk (from the contract itself) can provide the principal an additional tool with which to improve the ultimate allocation. For a discussion of some of these issues see Edlin and Hermalin (1997). Note we still have the flexibility to endogenously impose risk over the *allocation* (the x), we discuss the desirability of doing so below. See also Maskin (1981).

Continuing with our development of the standard framework, we assume that nature chooses the agent's type, θ, according to the distribution function $F(\cdot) : [\theta_L, \theta_H] \to [0,1]$. Let $f(\cdot)$ be the associated density function, which we assume to be continuous and to have full support (*i.e.,* $f(\theta) > 0$ for all $\theta \in [\theta_L, \theta_H]$). Assuming a continuum of types and a distribution without mass points is done largely for convenience. It also generalizes our analysis from just two types. Admittedly, we could have generalized beyond two types by allowing for a finite number of types greater than two. The conclusions we would reach by doing so would be economically similar to those we'll shortly obtain with a continuum of types.[17] The benefit of going all the way to the continuum is it allows us to employ calculus, which streamlines the analysis.

Recall that, at its most general, a direct-revelation mechanism is a mapping from the type space into a *distribution* over outcomes. Given the way that money enters both players utility functions, we're free to replace a distribution over payments with an expected payment, which means we're free to assume that

[17] See, *e.g.,* Caillaud and Hermalin (1993), particular §3, for a finite-type-space analysis under the standard framework.

the payment is fixed deterministically by the agent's announcement.[18] What about random-*allocation* mechanisms? The answer depends on the risk properties of the two players' utilities over allocation. If we postulate that $w(\cdot,\theta)$ and $u(\cdot,\theta)$ are concave for all $\theta \in [\theta_L, \theta_H]$, then, *absent incentive concerns*, there would be no reason for the principal to prefer a random-allocation mechanism; indeed, if at least one is strictly concave (*i.e.,* concave, but not affine), then she would strictly prefer not to employ a random-allocation mechanism absent incentive concerns: Her expected utility is greater with a deterministic mechanism and, since the agent's expected utility is greater, her payment to him will be less (a benefit to her). Hence, we would only expect to see random-allocation mechanisms if the randomness somehow relaxed the incentive concerns. Where does this leave us? At this point, consistent with what is standardly done, we will assume that both $w(\cdot,\theta)$ and $u(\cdot,\theta)$ are concave, with one at least being strictly concave, for all $\theta \in [\theta_L, \theta_H]$ (note this entails that $\Omega(\cdot,\theta)$, the social surplus function, is also strictly concave). Hence, absent incentive concerns, we'd be free to ignore random-allocation mechanisms. For the time being, we'll also ignore random-allocation mechanisms with incentive concerns. Later, we'll consider the circumstances under which this is appropriate. Since we're ignoring random mechanisms, we'll henceforth write $\langle x(\cdot), s(\cdot)\rangle$ instead of $\sigma(\cdot)$ for the mechanism.

Within this framework, the route that we follow consists of two steps. First, we will characterize the set of direct-revelation contracts; that is, the set of contracts $\langle x(\cdot), s(\cdot)\rangle$ from $[\theta_L, \theta_H]$ to $\mathbb{R} \times \mathbb{R}_+$ that satisfy truthful revelation. This truthful-revelation—or incentive compability—condition can be expressed as:

$$s(\theta) + u\left[x(\theta), \theta\right] \geq s(\tilde{\theta}) + u\left[x(\tilde{\theta}), \theta\right] \tag{11}$$

for all $(\theta, \tilde{\theta}) \in [\theta_L, \theta_H]^2$. After completing this first step, the second step is identifying from within this set of incentive-compatible contracts the one that maximizes the principal's expected utility subject to the agent's participation. Whether the agent participates depends on whether his equilibrium utility exceeds his reservation utility, $U_R(\theta)$. Observe that the requirement that the agent accept the contract imposes an additional constraint on the principal in designing the optimal contract. As before, we refer to this constraint as the *participation* or *individual-rationality* (IR) constraint. Recall that $\mathcal{X}$ is assumed to contain a no-trade allocation, x_0, $s = 0$ is a feasible transfer, and $U_R(\theta) \equiv \mathcal{U}(x_0, 0, \theta)$. Hence there is no loss of generality in requiring the principal to offer a contract that is individually rational for all types (although the "contract" for some types might be no trade).

We assume that the agent acts in the principal's interest when he's otherwise indifferent. In particular, he accepts a contract when he is indifferent between accepting and rejecting it and he tells the truth when indifferent between being honest and lying. This is simply a necessary condition for there to exist an

[18] That is, the mechanism that maps θ to a distribution $G(\theta)$ over payments is equivalent to a mechanism that maps θ to the deterministic payment $\hat{s}(\theta) = \mathbb{E}_{G(\theta)}\{s\}$.

equilibrum and, as such, should not be deemed controversial.

In our treatment of the retailer-supplier example, we assumed that both types of supplier had the same reservation utility (*i.e.,* recall, $U_R(\theta) = \mathcal{U}(0,0,\theta) = -C_\theta(0) = 0$). It is possible, however, to imagine models in which the reservation utility varies with θ. For instance, suppose that an efficient supplier could, if *not* employed by the retailer, market its goods directly to the ultimate consumers (although, presumably, not as well as the retailer could). Suppose, in fact, it would earn a profit of $\pi_E > 0$ from direct marketing. Then we would have $U_R(E) = \pi_E$ and $U_R(I) = 0$. A number of authors (see, *e.g.,* Lewis and Sappington, 1989, Maggi and Rodriguez-Clare, 1995, and Jullien, 1996) have recently studied the role of such type-*dependent* reservation utilities in contractual screening models. Type dependence can, however, greatly complicate the analysis. We will, therefore, adopt the more standard assumption of type-*in*dependent reservation utilities; that is, we assume—as we did in our retailer-supplier example—that

$$U_R(\theta) = U_R \text{ and } W_R(\theta) = W_R$$

for all $\theta \in [\theta_L, \theta_H]$. As a further convenience, we will interpret $x = 0$ as the no-trade allocation. Observe that these last two assumptions imply

$$\begin{aligned} u(0,\theta) &= u(0,\theta') = U_R \text{ and} \\ w(0,\theta) &= w(0,\theta') = W_R \end{aligned}$$

for all θ, $\theta' \in [\theta_L, \theta_H]$. Given these assumptions, we can express the agent's participation constraint as

$$s(\theta) + u[x(\theta), \theta] \geq U_R. \tag{12}$$

Although our treatment of reservation utilities and no trade is standard, we have, nevertheless added to the assumptions underlying the standard framework.

At last, we can state the problem that we seek to solve: Find the optimal contract $\langle x(\cdot), s(\cdot)\rangle$ that maximizes

$$\int_{\theta_L}^{\theta_H} \left(w\left[x(\theta),\theta\right] - s(\theta)\right) f(\theta) d\theta \tag{13}$$

subject to (11) and (12) holding.

Before solving this program, it's valuable to consider the full (symmetric) information benchmark: *Ex post* efficiency corresponds to adopting the allocation

$$x^F(\theta) \in \arg\max_{x \in \mathbb{R}_+} \Omega(x,\theta)$$

for each θ (recall $\Omega(x,\theta)$ is the aggregate surplus from the relationship). Our earlier assumptions ensure that $\Omega(\cdot,\theta)$ is strictly concave for each θ; so the *ex post* efficient allocation is uniquely defined by the first-order condition:

$$\frac{\partial \Omega}{\partial x}(x^F(\theta),\theta) = \frac{\partial w}{\partial x}(x^F(\theta),\theta) + \frac{\partial u}{\partial x}(x^F(\theta),\theta) = 0.$$

Our earlier assumptions also entail that $x^F(\cdot)$ is uniformly bounded from above. Observe that any sharing of the surplus can, then, be realized by the appropriate transfer function.[19] It follows that, if the principal knows θ—that is, the parties are playing under full (symmetric) information—then the contracting game can be easily solved: In equilibrium, the principal offers a contract $\langle x^F(\cdot), s^F(\cdot)\rangle$ such that the agent's utility is exactly equal to his reservation utility; *i.e.,*

$$s^F(\theta) = U_R - u(x^F(\theta), \theta).$$

In other words, the principal captures the entire surplus—a consequence of endowing her with all the bargaining power—leaving the agent at his outside (non-participation) option.

5.1 The Spence-Mirrlees Assumption

Before we proceed to solve (13), we need to introduce one more assumption. Given this assumption's importance, it is worth devoting a short section to it.

In order to screen types, the principal must be able to exploit differences across the *tradeoffs* that different types are willing to make between money and allocation. Otherwise a strategy, for instance, of decreasing the x expected from the agent in exchange for slightly less pay wouldn't work to induce one type to reveal himself to be different than another type. Recall, for instance, because the marginal cost of output differed between the efficient and inefficient types in our retailer-supplier example, we (the buyer) could design a contract to induce revelation. Different willingness to make tradeoffs means we require that the indifference curves of the agent in x-s space differ with his type. In fact, we want, for any point in x-s space, that these slopes vary monotonically with whatever natural order applies to the type space. Or, when, no natural order applies, we want it to be possible to define an order, $\succ$, over the types so that $\theta \succ \theta'$ if and only if the slope θ's indifference curve is greater (alternatively less) than the slope of θ''s indifference at *every* point $(x, s) \in \mathcal{X} \times \mathcal{S}$. Such a monotonicity-of-indifference-curves condition is known as a *Spence-Mirrless condition* and, correspondingly, the assumption that this condition is met is known as the Spence-Mirrlees assumption. SPENCE-MIRRLEES

The slope of the indifference curve in x-s space is equal to $-\partial u/\partial x$. Hence, we require that $-\partial u/\partial x$ or, equivalently and more naturally, $\partial u/\partial x$ vary monotonically in θ. Specifically, we assume:

A1 (Spence-Mirrlees Assumption): *For all* $x \in \mathcal{X}$,

$$\frac{\partial u\,(x, \theta)}{\partial x} > \frac{\partial u\,(x, \theta')}{\partial x}$$

[19] Note that this is an instance of where we're exploiting the additive separability (lack of income effects) assumed of the utility functions. Were there income effects, we couldn't define $x^F(\cdot)$ independently of the transfers.

if $\theta > \theta'$ *(note* $\Theta \subset \mathbb{R}$*).*[20]

That is, if $\theta > \theta'$—θ is a *higher* type than θ'—then -1 times the slope of type θ's indiffernce curve is, at any point, greater than -1 times the slope of type θ''s indifference curve. Observe that a consequence of Assumption A1 is that a given indiffence curve for one type can cross a given indifference curve of another type at most once. For this reason, Assumption A1 is sometimes called a *single-crossing condition.* Figure 2 illustrates. SINGLE-CROSSING

If, as assumed in the standard framework, $\Theta = [\theta_L, \theta_H]$ and $u(\cdot, \cdot)$ is three-times differentiable, then Assumption A1 is equivalent to

A1 (Standard framework Spence-Mirrlees Assumption): *For all* $(x, \theta) \in \mathbb{R}_+ \times [\theta_L, \theta_H]$,

$$\frac{\partial^2 u(x, \theta)}{\partial \theta \partial x} > 0.$$

Economically, the Spence-Mirrlees assumption tells us that the agent's marginal benefit from increasing x is increasing in his type. As an example of all this, recall our retailer-supplier model. There, the efficient, E, type was the higher type (*i.e.,* $E \succ I$). Recall too that $u(x, \theta) = -C_\theta(x)$. Recall as well our assumption (definition) that the efficient type had the lower marginal cost. Putting this all together, we see that Assumption A1 holds for retailer-supplier model.

It is important to understand that the Spence-Mirrlees assumption is an assumption about order. Consequently, differentiability of u with respect to either x or θ is not necessary. Nor, in fact, is it necessary that $\mathcal{U}$ be additively separable as we've been assuming. At its most general, then, we can state the Spence-Mirrlees assumption as

A1′ (General Spence-Mirrlees Assumption): *There exists an order* $\succ_\theta$ *on* Θ *such that if* $\theta' \succ_\theta \theta''$*, then*

$$\mathcal{U}(x', s', \theta'') \geq \mathcal{U}(x'', s'', \theta'') \Longrightarrow \mathcal{U}(x', s', \theta') > \mathcal{U}(x'', s'', \theta'),$$

whenever $x' \succ_x x''$ *(where* s'*,* $s'' \in \mathcal{S}$ *and* $\succ_x$ *completely orders* $\mathcal{X}$*).*

[20] The assumption that $\Theta \subset \mathbb{R}$ is not critical. For suppose that $t \in \mathcal{T}$ were some natural or intuitive definition of type, where $\mathcal{T}$ had no natural order or was ordered by something other than $>$. Suppose that for any pairs of types t and $t' \in \mathcal{T}$ that

$$\left(\frac{\partial u(x, t)}{\partial x} - \frac{\partial u(x, t')}{\partial x}\right) \times \left(\frac{\partial u(x', t)}{\partial x} - \frac{\partial u(x', t')}{\partial x}\right) > 0$$

for any pair of allocations x and $x' \in \mathcal{X}$. Then observe that by picking a specific x, we can define

$$\theta(t) = \frac{\partial u(x, t)}{\partial x}.$$

This new type space, $\Theta \equiv \theta(\mathcal{T}) \subset \mathbb{R}$ is thus isomorphic to our original type space plus which it meets the Spence-Mirrlees assumption.

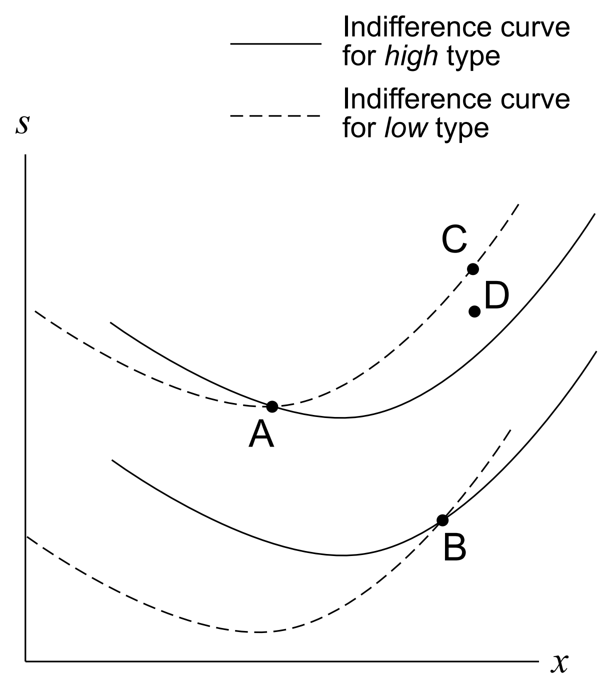

Figure 2: The *Spence-Mirrlees* Assumption: Through any point (*e.g.*, A or B), the indifference curve through that point for the high type cross the indifference curve through that point for the low point from *above.*

This generalized Spence-Mirrlees assumption states that we can order the types so that if a low type (under this order) prefers, at least weakly, an outcome with more x (with "more" being defined by the order $\succ_x$) than a second outcome, then a higher type must strictly prefer the first outcome to the second. Figure 2 illustrates: Since the low type prefers point C to A (weakly), the high type must strictly prefer C to A, which the figure confirms. Similarly, since the low type prefers C to B (strictly), the high type must also strictly prefer C to B, which the figure likewise confirms.[21] See Milgrom and Shannon (1994) for

[21]Observe, as shown in Figure 2, that the area above the higher type's indifference curve to the right of given point (A or B) is larger than the area above the lower type's indifference curve to the right of that point. This suggests an alternative statement of A1$'$. For a given point (x_0, s_0), define $\mathcal{H}_0 = \{(x,s)\,|x \succsim_x x_0\}$ (*i.e.*, $\mathcal{H}_0$ is the right half-plane defined by the vertical line $x = x_0$) and define

$$\mathcal{P}_0(\theta) = \{(x,s)\,|\mathcal{U}(s,x,\theta) \geq \mathcal{U}(s_0,x_0,\theta)\}$$

(*i.e.*, $\mathcal{P}_0(\theta)$ are the outcomes preferred by a type-θ agent to (x_0, s_0)). Then A1$'$ is equivalent to

$$\mathcal{P}_0(\theta') \cap \mathcal{H}_0 \subset \mathcal{P}_0(\theta) \cap \mathcal{H}_0$$

for any point (x_0, s_0) when $\theta \succ_\theta \theta'$.

a more complete discussion of the relationship between Assumption A1′ and Assumption A1.

As suggested at beginning of this sub-section, the consequence of the Spence-Mirrlees assumption (stated either as A1 or A1′) is that it is possible to *separate* any two types; by which we mean it is possible to find two outcomes (x_1, s_1) and (x_2, s_2) such that a type-θ_1 agent prefers (x_1, s_1) to (x_2, s_2), but a type-θ_2 agent has the opposite preferences. For instance, in Figure 2, let point A be (x_1, s_1) and let D be (x_2, s_2). If θ_2 is the high type and θ_1 is the low type, then it is clear that given the choice between A and D, θ_1 would select A and θ_2 would select D; that is, this pair of contracts separates the two types. Or, for example, back in Figure 1, contracts D and E separate the inefficient and efficient types of supplier.

5.2 Characterizing the Incentive-Compatible Contracts

Our approach to solving the principal's problem (13) is a two-step one. First, we will find a convenient characterization of the set of incentive-compatible mechanisms (*i.e.,* those that satisfy (11)). This is our objective here. Later, we will search from *within* this set for those that maximize (13) subject to (12).

Within the standard framework it is relatively straightforward to derive the necessary conditions implied by the self-selection constraints. Our approach is standard (see, *e.g.,* Myerson, 1979, among others). Consider any direct-revelation mechanism $\langle x(\cdot), s(\cdot)\rangle$ and consider any pair of types, θ_1 and θ_2, with $\theta_1 < \theta_2$. Direct revelation implies, among other things, that type θ_1 won't wish to pretend to be type θ_2 and *vice versa.* Hence,

$$\begin{aligned} s(\theta_1) + u\left[x(\theta_1), \theta_1\right] &\geq s(\theta_2) + u\left[x(\theta_2), \theta_1\right] \text{ and} \\ s(\theta_2) + u\left[x(\theta_2), \theta_2\right] &\geq s(\theta_1) + u\left[x(\theta_1), \theta_2\right]. \end{aligned}$$

As is often the case in contract theory, it is easier to work with utilities than payments. To this end, define

$$v(\theta) = s(\theta) + u\left[x(\theta), \theta\right].$$

Observe that $v(\theta)$ is the type-θ agent's *equilibrium* utility. The above pair of inequalities can then be written as:

$$\begin{aligned} v(\theta_1) &\geq v(\theta_2) - u\left[x(\theta_2), \theta_2\right] + u\left[x(\theta_2), \theta_1\right] \text{ and} \\ v(\theta_2) &\geq v(\theta_1) - u\left[x(\theta_1), \theta_1\right] + u\left[x(\theta_1), \theta_2\right]. \end{aligned}$$

Or, combining these two inequalities, as

$$\int_{\theta_1}^{\theta_2} \frac{\partial u\left[x(\theta_1), \theta\right]}{\partial \theta} d\theta \leq v(\theta_2) - v(\theta_1) \leq \int_{\theta_1}^{\theta_2} \frac{\partial u\left[x(\theta_2), \theta\right]}{\partial \theta} d\theta. \tag{14}$$

This double inequality has two consequences. First, ignoring the middle term, it implies

$$\int_{\theta_1}^{\theta_2}\int_{x(\theta_1)}^{x(\theta_2)}\frac{\partial^2 u}{\partial x\partial\theta}(x,\theta)dxd\theta \geq 0.$$

The Spence-Mirrlees assumption, A1, means the integrand is positive. Given $\theta_1 < \theta_2$, this means the integral can be non-negative only if $x(\theta_1) \leq x(\theta_2)$. Since this is true for any $\theta_1 < \theta_2$, we may conclude that *the allocation function* $x(\cdot)$ *is non-decreasing.* Note that this necessarily implies that $x(\cdot)$ is almost everywhere continuous.

The second consequence of (14) is as follows: By fixing one end point and letting the other converge towards it, we see that $v(\cdot)$ is absolutely continuous with respect to Lebesgue measure and is, thus, almost everywhere differentiable.[22] This derivative is

$$\frac{dv(\theta)}{d\theta} = \frac{\partial u(x(\theta),\theta)}{\partial\theta}$$

almost everywhere.[23] Consequently, one can express equilibrium utility as

$$v(\theta) = v(\theta_L) + \int_{\theta_L}^{\theta}\frac{\partial u}{\partial\theta}(x(t),t)dt. \tag{15}$$

Expression (15) and the monotonicity of the allocation function $x(\cdot)$ are, thus, necessary properties of a direct-revelation mechanism. In particular, we've just proved that a necessary condition for an allocation function $x(\cdot)$ to be implementable is that

$$s(\theta) = v_L - u(x(\theta),\theta) + \int_{\theta_L}^{\theta}\frac{\partial u}{\partial\theta}(x(t),t)dt.$$

where v_L is an arbitrary constant.

It turns out that these properties are also *sufficient*:

Theorem 1 ***(Characterization of direct-revelation mechanisms)*** *Within the standard framework and under Assumption A1, a direct mechanism* $\langle x(\cdot), s(\cdot)\rangle$ *is truthful if and only if there exists a real number* v_L *such that:*

$$s(\theta) = v_L - u(x(\theta),\theta) + \int_{\theta_L}^{\theta}\frac{\partial u}{\partial\theta}(x(t),t)dt \tag{16}$$

$$\text{and } x(\cdot) \text{ is non-decreasing.} \tag{17}$$

Consequently, an allocation function $x(\cdot)$ *is implementable if and only if it is non-decreasing.*

[22] To be precise, one should make assumptions to ensure that $\partial u(x(\cdot),\cdot)/\partial\theta$ is integrable on Θ. This can be done by requiring that $\mathcal{X}$ be bounded or that $\partial u(\cdot,\cdot)/\partial\theta$ be bounded on $\mathcal{X}\times\Theta$. Both assumptions are natural in most economic settings and simply extend the assumptions that bound $x^F(\cdot)$. Henceforth, we assume $\partial u/\partial\theta$ is bounded.

[23] Note the important difference between $\frac{\partial u[x(\theta),\theta]}{\partial\theta}$ and $\frac{du[x(\theta),\theta]}{d\theta}$. The former is the *partial* derivative of u with respect to its second argument evaluated at $(x(\theta),\theta)$, while the latter is the total derivative of u.

Proof: Since we established necessity in the text, we need only prove sufficiency here. Let $\langle x(\cdot), s(\cdot)\rangle$ satisfy (16) and (17). Consider the agent's utility when the state of nature is θ, but he claims that it is $\theta' > \theta$:

$$s(\theta') + u(x(\theta'),\theta) = \overbrace{v_L - u\left(x\left(\theta'\right),\theta'\right) + \int_{\theta_L}^{\theta'} \frac{\partial u}{\partial \theta}(x(t),t)dt}^{s(\theta')} + u(x(\theta'),\theta)$$

$$= \underbrace{v_L - u\left(x\left(\theta\right),\theta\right) + \int_{\theta_L}^{\theta} \frac{\partial u}{\partial \theta}(x(t),t)dt}_{s(\theta)} + u(x(\theta),\theta) \tag{18}$$

$$+ \left[u(x(\theta'),\theta) - u\left(x\left(\theta'\right),\theta'\right)\right] + \left[\int_{\theta_L}^{\theta'} \frac{\partial u}{\partial \theta}(x(t),t)dt - \int_{\theta_L}^{\theta} \frac{\partial u}{\partial \theta}(x(t),t)dt\right] \tag{19}$$

$$= v\left(\theta\right) + \int_{\theta}^{\theta'} \left[\frac{\partial u}{\partial \theta}(x(t),t) - \frac{\partial u}{\partial \theta}\left(x\left(\theta'\right),t\right)\right]dt$$

Where the second equality (beginning of (18)) derives from adding and subtracting $s(\theta)$. In the last line, the first term, $v(\theta)$, is (18) and the second term, the integral, is (19). Since we've assumed (17), $x(t) \leq x(\theta')$ for $t \in [\theta, \theta']$. Moreover, A1 implies $\partial u/\partial\theta$ is increasing in x. Hence, the integral in the last line is non-positive; which means we may conclude

$$s(\theta') + u(x(\theta'),\theta) \leq v\left(\theta\right).$$

That is, under this mechanism, the agent does better to tell the truth than exaggerate his type. An analogous analysis can be used for $\theta' < \theta$ (*i.e.,* to show the agent does better to tell the truth than understate his type). Therefore, the revelation constraints hold and the mechanism is indeed truthful. ■

This characterization theorem is, now, a well-known result and can be found, implicitly at least, in almost every mechanism design paper. Given its importance, it is worth understanding how our assumptions drive this result. In particular, we wish to call attention to the fact that neither the necessity of (15) nor (16) depends on the Spence-Mirrlees assumption. The Spence-Mirrlees assumption's role is to establish that a monotonic allocation function is necessary and that, if $x(\cdot)$ is monotonic, then (16) is sufficient to ensure a truth-telling equilibrium.

To further illustrate these points let's consider an alternative approach to the revelation constraints inspired by Guesnerie and Laffont (1984). The standard framework continues to apply. Assume, however, that $x(\cdot)$ is piecewise twice continuously differentiable.[24] The revelation constraint, expressed as (14), shows this property extends to the transfer function, $s(\cdot)$. Now, let $U(\theta', \theta)$ denote the agent's utility when he claims to be type θ' but is really type θ. The revelation constraint can, thus, be written:

$$v(\theta) = \max_{\theta' \in \Theta} U(\theta', \theta) = U(\theta, \theta). \tag{20}$$

Applying the envelope theorem to this program yields

$$\frac{dv}{d\theta}(\theta) = U_2(\theta', \theta) \mid_{\theta'=\theta} = \frac{\partial u}{\partial \theta}(x(\theta), \theta)$$

(for all but a finite number of θ).[25] The integral expressions (15) and (16) follow; hence, as claimed, the relationship between the transfer function and the allocation function in a direct-revelation mechanism does not depend upon the Spence Mirrlees assumption. In fact, these integral expressions are simply deduced from (20)'s first-order condition, $U_1(\theta, \theta) = 0$. Of course, we also need to pay attention to the the *second*-order conditions. Given our differentiability assumptions, the second-order conditions reduce to $U(\theta', \theta)$ being locally concave in θ' around the point $\theta' = \theta$, that is:

$$U_{11}(\theta, \theta) \leq 0$$

(for all but a finite number of θ). Observe that, differentiating the first-order condition with respect to θ',

$$U_{11}(\theta, \theta) + U_{12}(\theta, \theta) = 0;$$

hence the local concavity condition is, therefore, equivalent to:

$$0 \leq U_{12}(\theta, \theta) = \frac{\partial^2 u(x(\theta), \theta)}{\partial x \partial \theta} \cdot x'(\theta)$$

(for all but a finite number of θ). The specific role played by the Spence-Mirrlees assumption now emerges: The assumption allows one to translate the second-order condition implicit in the revelation program (20) into a monotonicity condition on the function $x(\cdot)$.

This discussion also demonstrates a point that was implicit in our earlier discussion of the Spence-Mirrlees assumption: What is critical is not that $\frac{\partial^2 u}{\partial \theta \partial x}$

[24] "Piecewise" means that the property is true except at a finite number of points. The approach in Guesnerie and Laffont (1984) actually requires just that $x(\cdot)$ be piecewise continuously differentiable. Note that, since $x(\cdot)$ is ultimately *endogenous*, making any assumptions about it is less than ideal. If, however, we show that the optimal (second-best) $x(\cdot)$ has these properties, then there's no harm done.

[25] When working with $U(\cdot, \cdot)$ it is helpful to use the notation U_i to denote the partial derivative with respect to the ith argument. U_{ii} denotes the second partial derivative with respect to the ith argument and U_{ij} denotes the cross partial derivative.

be positive, but rather that it keep a constant sign over the relevant domain. If, instead of being positive, this cross-partial derivative were negative everywhere, then our analysis would remain valid, except that it would give us the inverse monotonicity condition: $x(\cdot)$ would need to be non-increasing in type. But with a simple change of the definition of type, $\tilde{\theta} = -\theta$, we're back to our original framework. Since, as we argued above, the definition of type is somewhat arbitrary, we see that our conclusion of a non-*de*creasing $x(\cdot)$ is simply a consequence of the assumption that different types of agent have different marginal rates of substitution between money and allocation and that an ordering of these marginal rates of substitution by type is invariant to which point in $\mathcal{X} \times \mathcal{S}$ we're considering.

What if the Spence-Mirrlees assumption is violated (*e.g.*, $\frac{\partial^2 u}{\partial x \partial \theta}$ changes sign)? As our discussion indicates, although we still have necessary conditions concerning incentive-compatible mechanisms, we no longer have any reason to expect $x(\cdot)$ to be monotonic. Moreover—and more critically if we hope to characterize the set of incentive-compatible mechanisms—we have no sufficiency results. It is not surprising, therefore, that little progress has been made on the problem of designing optimal contracts when the Spence-Mirrlees condition fails.

5.3 Optimization in the standard framework

The previous analysis has given us, within the standard framework at least, a complete characterization of the space of possible (incentive-compatible) contracts. We can now concentrate on the principal's problem of designing an optimal contract.

Finding the optimal direct-revelation mechanism for the principal means maximizing the principal's expected utility over the set of mechanisms that induce truthful revelation of the agent's type and full participation. From page 20, the participation constraint is (12); while, from the previous section, truthful revelation is equivalent to (16) and (17).[26] We can, thus, express the principal's problem as

$$\max_{x(\cdot),s(\cdot)} \int_{\theta_L}^{\theta_H} \left[w(x(\theta),\theta) - s(\theta)\right] f(\theta) d\theta$$
$$\text{subject to (12), (16), and (17).}$$

Once again, it's more convenient to work with $v(\cdot)$ than $s(\cdot)$. Observe that

$$\begin{aligned} w\left[x(\theta),\theta\right] - s(\theta) &= w\left[x(\theta),\theta\right] - v(\theta) + u\left[x(\theta),\theta\right] \\ &= \Omega(x(\theta),\theta) - v(\theta). \end{aligned}$$

Moreover, (16) can be used to compute $v(\theta)$ using only $x(\cdot)$ and a number v_L.

[26] Assuming A1 holds, which we will do henceforth.

Hence, we're free to write the principal's problem as

$$\begin{aligned}\int_{\theta_L}^{\theta_H}[w(x(\theta),\theta)-s(\theta)]\,f(\theta)d\theta &= \int_{\theta_L}^{\theta_H}[\Omega(x(\theta),\theta)-v(\theta)]\,f(\theta)d\theta \\ &= \int_{\theta_L}^{\theta_H}\left[\Omega(x(\theta),\theta)-\int_{\theta_L}^{\theta}\frac{\partial u}{\partial\theta}(x(t),t)dt\right]f(\theta)d\theta - v_L.\end{aligned}$$

Integration by parts (or Fubini's theorem)[27] implies

$$-\int_{\theta_L}^{\theta_H}\left[\int_{\theta_L}^{\theta}\frac{\partial u}{\partial\theta}(x(t),t)dt\right]f(\theta)d\theta = -\int_{\theta_L}^{\theta_H}[1-F(\theta)]\,\frac{\partial u}{\partial\theta}(x(\theta),\theta)d\theta,$$

which allows us to further transform the principal's objective function:

$$\int_{\theta_L}^{\theta_H}[w(x(\theta),\theta)-s(\theta)]\,f(\theta)d\theta = \int_{\theta_L}^{\theta_H}\left[\Omega\left(x\left(\theta\right),\theta\right)-\frac{1-F\left(\theta\right)}{f\left(\theta\right)}\frac{\partial u}{\partial\theta}\left(x\left(\theta\right),\theta\right)\right]f\left(\theta\right)d\theta - v_L.$$

Remark 1 *From this last expression, we can see that it is unreasonable to expect to achieve the first best: The principal's objective function differs from the first-best objective function,* $\max\mathbb{E}_\theta\left\{\Omega\left(x\left(\theta\right),\theta\right)\right\}$, *by*

$$-\int_{\theta_L}^{\theta_H}\left[1-F\left(\theta\right)\right]\frac{\partial u\left[x\left(\theta\right),\theta\right]}{\partial\theta}d\theta.$$

Consequently, since the principal wishes to maximize something other than expected social surplus and the principal proposes the contract, we can't expect the contract to maximize social surplus.

Define

$$\Sigma(x,\theta)\equiv\Omega(x,\theta)-\frac{[1-F(\theta)]}{f(\theta)}\frac{\partial u}{\partial\theta}(x,\theta).$$

Observe that our earlier assumptions ensure that $\Sigma(x,\theta)$ is bounded and at least twice-differentiable. We will refer to $\Sigma\left(x,\theta\right)$ as the *virtual surplus* (this follows Jullien, 1996).[28]

The principal's problem can now be restated in a tractable and compact form:

$$\max_{x(\cdot),v_L}\left\{\int_{\theta_L}^{\theta_H}\Sigma(x(\theta),\theta)f(\theta)d\theta - v_L\right\} \tag{21}$$

$$\text{subject to } v_L+\int_{\theta_L}^{\theta}\frac{\partial u}{\partial\theta}(x(t),t)dt\geq U_R$$

$$\text{and } x(\cdot)\text{ is non-decreasing.}$$

---

[27] Both are equivalent in the present case, but Fubini's theorem is perhaps more appropriate since it extends to multi-dimensional frameworks as well. Note too our reliance on $\partial u/\partial\theta$ being bounded (see footnote 22).

[28] Guesnerie and Laffont (1984) and Caillaud et al. (1988) use the term *surrogate welfare function*.

Before solving this program generally (a problem we return to in Section 5.5), we will first solve the problem under three additional assumptions:

1. For all x, $\partial u/\partial\theta \geq 0$ (*i.e.,* utility is non-decreasing in type);
2. $\Sigma(\cdot,\theta)$ is strictly quasi-concave for all $\theta \in [\theta_L, \theta_H]$; and
3. $\partial\Sigma/\partial x$ is non-decreasing in θ for all x.

What are the consequences of these assumptions? The first entails that the participation constraint holds for all types if it holds for the lowest type, θ_L. Consequently, we can ignore this constraint for all but the lowest type. Moreover, for this type, the constraint reduces to $v_L \geq U_R$. Since v_L is a direct transfer to the agent without incentive effects, we know the principal will set it as low as possible. That is, we can conclude that, optimally, $v_L = U_R$. Note, too, this means the participation constraint is binding for the lowest type (similarly to what we saw in the retailer-supplier example). To summarize:

Lemma 1 *If utility is non-decreasing in type for all allocations (i.e., $\partial u/\partial\theta \geq 0$ for all x), then (i) $v_L = U_R$; (ii) the participation constraint is binding for the lowest type, θ_L; and (iii) the participation constraint holds trivially (is slack) for all higher types (i.e., for $\theta > \theta_L$).*

In light of this lemma, we can be emboldened to try the following solution technique for (21): Ignore the monotonicity constraint and see if the *un*constrained problem yields a monotonic solution. The solution to the unconstrained problem, $x^*(\cdot)$, is to solve (21) pointwise; that is, to set $x^*(\theta) = X(\theta)$, where

$$X(\theta) \equiv \arg\max_x \Sigma(x,\theta).$$

Note that the second assumption means $X(\cdot)$ is uniquely defined. Finally, the third assumption—the marginal-benefit schedule, $\partial\Sigma/\partial x$, is non-decreasing in θ—means the point at which $\partial\Sigma(x,\theta)/\partial x$ crosses zero is non-decreasing in θ. But this point is $X(\theta)$; hence, monotonicity is ensured. To conclude:

Proposition 4 *If*

- *for all x, $\partial u/\partial\theta \geq 0$;*
- *$\Sigma(\cdot,\theta)$ is strictly quasi-concave; and*
- *$\partial\Sigma/\partial x$ is non-decreasing in θ;*

then the solution to (21) is $x^*(\theta) = X(\theta)$ and $v_L = U_R$.

How does the solution in Proposition 4 compare to the full-information benchmark? The answer is given by the following corollaries:

Corollary 1 *$x^*(\theta) < x^F(\theta)$ for all $\theta \in [\theta_L, \theta_H)$ and $x^*(\theta_H) = x^F(\theta_H)$.*

Proof: At $x = x^F(\theta)$,

$$\frac{\partial \Sigma}{\partial x} = -\frac{1 - F(\theta)}{f(\theta)} \times \frac{\partial^2 u}{\partial x \partial \theta}.$$

Since (i) $1 - F(\theta) > 0$ (except for $\theta = \theta_H$) and (ii) the cross-partial derivative is strictly positive by A1, the right-hand side is negative for all θ, except θ_H. Consequently, since $\Sigma(\cdot, \theta)$ is strictly quasi-concave, we can conclude that $x^*(\theta) < x^F(\theta)$ for all $\theta \in [\theta_L, \theta_H)$. For $\theta = \theta_H$, we've just seen that

$$\frac{\partial \Sigma\left[x^F(\theta_H), \theta_H\right]}{\partial x} = 0;$$

hence, the strict quasi-concavity of $\Sigma(\cdot, \theta_H)$ ensures that $x^F(\theta_H)$ is the maximum. ■

Corollary 2 *$v'(\theta) \geq 0$ and $v(\theta_L) = U_R$.*

Proof: We've already established the second conclusion. The first follows since

$$v(\theta) = v_L + \int_{\theta_L}^{\theta} \frac{\partial u}{\partial \theta}(x(t), t)dt;$$

so

$$v'(\theta) = \frac{\partial u}{\partial \theta}[x(\theta), \theta] \geq 0$$

by the first assumption. ■

In much of the literature, the three additional assumptions supporting Proposition 4 hold, so Proposition 4 and its corollaries might be deemed the "standard solution" to the contractual screening problem within the standard framework. These conclusions are sometimes summarized as

Remark 2 *Under the* standard solution, *there's a downward distortion in allocation (relative to full information) for all types but the highest, the lowest type earns* no *information rent, but higher types may.*

Observe the "may" at the end of the last remark becomes a "do" if

$$\frac{\partial u[X(\theta), \theta]}{\partial \theta} > 0$$

for all $\theta > \theta_L$.[29]

[29]There are, of course, many assumptions that will change the "may" to a "do."

5.4 The retailer-supplier example revisited

Before returning to our fairly abstract analysis of contractual screening, let's consider an extension of our earlier retailer-supplier example. Specifically, let's imagine that there are a continuum of efficiency types, which we normalize to be the interval $[1,2]$. Instead of $C_t(x)$, write the supplier's cost function as $C(x,\theta)$, and suppose that

$$\frac{\partial^2 C}{\partial\theta\partial x} < 0; \tag{22}$$

that is, higher (more efficient) types have lower marginal costs. Since

$$u(x,\theta) = -C(x,\theta),$$

(22) implies that the Spence-Mirrlees assumption is met. In addition, to these assumptions, we are maintaining all the assumptions from our earlier model. In particular, we assume the revenue function, $r(\cdot)$, is concave and bounded. Hence, since $C(\cdot,\theta)$ is strictly convex,

$$\lim_{x\to\infty} \Omega(x,\theta) = -\infty \text{ for all } \theta.$$

Assume, too, that $\partial\Omega(0,\theta)/\partial x > 0$ for all θ; *i.e.,* $x^F(\theta) > 0$ for all θ. It is readily checked that all the assumptions of the standard framework are, therefore, satisfied.

Since $C(\cdot,\theta)$ is a cost function, we necessarily have $C(0,\theta) = 0$ for all θ. Combined with (22), this entails that $C(x,\theta) > C(x,\theta')$ if $\theta < \theta'$.[30] Hence, we may conclude

$$\frac{\partial u}{\partial\theta}(x,\theta) > 0. \tag{23}$$

Observe that

$$\Sigma(x,\theta) = r(x) - C(x,\theta) - \frac{1-F(\theta)}{f(\theta)}\left[\frac{-\partial C}{\partial\theta}\right].$$

Hence,

$$\frac{\partial\Sigma(x,\theta)}{\partial x} = r'(x) - \frac{\partial C}{\partial x} - \frac{1-F(\theta)}{f(\theta)}\left[\frac{-\partial^2 C}{\partial x\partial\theta}\right].$$

It is clear, therefore, that to take advantage of Proposition 4, we need to know something about the shape of $\partial C(\cdot,\theta)/\partial\theta$ (*e.g.,* is it at least quasi-concave) and how the *Mills ratio* [31] and the cross-partial derivative change with respect to θ. To this end, let's impose two frequently made assumptions: MILLS RATIO

[30]**Proof:** The initial condition, $C(0,t) = 0$ for all $t \in [1,2]$, means we can write

$$C(x,\theta') - C(x,\theta) = \int_0^x \int_\theta^{\theta'} \frac{\partial^2 C}{\partial\theta\partial x} d\theta dx;$$

the result follows from (22).

[31]The Mills ratio is the ratio of a survival function (here, $1-F(\theta)$) to its density function (here, $f(\theta)$). Because the Mills ratio is the inverse of the hazard rate, it is also known as the *inverse* hazard rate (this formal distinction between the inverse and "regular" hazard rate is not always respected by economic theorists).

- $C(x,\theta) = h(\theta)\,c(x)$, where $h(\cdot)$ is positive, strictly *de*creasing, and convex —higher types have lower marginal costs, but this marginal cost advantage may be less pronounced when moving up from one high type to another than it is when moving up from one low type to another. The function $c(\cdot)$ is strictly increasing, strictly convex, and $c(0) = 0$.
- Let $M(\theta)$ denote the Mills ratio. Assume $M'(\theta) \leq 1$.

Given these assumptions, we may conclude that $\Sigma(\cdot,\theta)$ is globally strictly concave and that

$$\begin{aligned}\frac{\partial^2 \Sigma(x,\theta)}{\partial\theta\partial x} &= -h'(\theta)\,c'(x) + M'(\theta)\,h'(\theta)\,c'(x) + M(\theta)\,h''(\theta)\,c'(x) \\ &\propto h'(\theta)\left[M'(\theta) - 1\right] + M(\theta)\,h''(\theta) \geq 0.\end{aligned}$$

That is, we may conclude that $\Sigma(\cdot,\theta)$ admits a unique maximum, which is nondecreasing with type. Combined with (23), this means we can apply Proposition 4.

For example, suppose $r(x) = x$, $h(\theta) = 1/\theta$, $c(x) = x^2/2$, and $F(\theta) = \theta - 1$ (*i.e.,* the uniform distribution on $[1,2]$). Both bullet points are met, so we can apply Proposition 4. This yields $x^*(\theta) = X(\theta)$, where

$$\begin{aligned}X(\theta) \text{ solves } 1 - \frac{x}{\theta} - (2-\theta)\frac{x}{\theta^2} &= 0; \\ \text{or } X(\theta) &= \frac{1}{2}\theta^2.\end{aligned}$$

From Proposition 4, we may set $v_L = U_R$, which is zero in this model. Noting that $\partial u/\partial\theta = x^2/2\theta^2$, we thus have

$$\begin{aligned}v(\theta) &= 0 + \int_{\theta_L}^{\theta} \frac{x^*(t)^2}{2t^2}dt \\ &= \int_1^{\theta} \frac{t^2}{8}dt \\ &= \frac{1}{24}\theta^3 - \frac{1}{24}.\end{aligned}$$

Hence, the *transfer* function is

$$\begin{aligned}s(\theta) &= v(\theta) - u(x(\theta),\theta) \\ &= \frac{1}{24}\theta^3 - \frac{1}{24} + \frac{\theta^3}{8} \\ &= \frac{\theta^3}{6} - \frac{1}{24}.\end{aligned}$$

Observe that

$$x^*(\theta) = \frac{1}{2}\theta^2 < \theta = x^F(\theta)$$

for all $\theta < 2$; that $x^*(2) = 2 = x^F(2)$; that $v(1) = 0$; and that $v(\theta) > 0$ for all $\theta > 1$—all consistent with the corollaries to Proposition 4.

Finally, although there's no need to do it, we can check that the incentive-compability constraints are indeed satisfied by the mechanism $\langle \frac{1}{2}\theta^2, \frac{1}{6}\theta^3 - \frac{1}{24} \rangle$:

$$\begin{aligned} \max_{\hat{\theta}} U\left(\hat{\theta}, \theta\right) &= \max_{\hat{\theta}} \frac{1}{6}\hat{\theta}^3 - \frac{1}{24} - \frac{\left[\frac{1}{2}\hat{\theta}^2\right]^2}{2\theta} \\ &\Longrightarrow \frac{\hat{\theta}^2}{2} - \frac{\hat{\theta}^3}{2\theta} = 0. \end{aligned}$$

Clearly, $\hat{\theta} = \theta$ satisfies the first-order condition (since $U(\cdot, \theta)$ is clearly concave, the second-order conditions are also met).

By the taxation principle (Proposition 3), an alternative to this direct-revelation contract is a payment schedule. Noting that $x^{*-1}(x) = \sqrt{2x}$, an optimal payment schedule, $S(x)$, is

$$S(x) = \begin{cases} s\left[x^{*-1}(x)\right] = \frac{x^{3/2}\sqrt{2}}{3} - \frac{1}{24} \text{ for } x \in \left[\frac{1}{2}, 2\right] \\ 0 \text{ for } x \notin \left[\frac{1}{2}, 2\right] \end{cases}.$$

5.5 General conditions for solving the principal's problem

We have seen that the principal's problem (21) has a simple and straightforward solution if $\partial u / \partial \theta$ is non-negative, $\Sigma(\cdot, \theta)$ is strictly quasi-concave, and $\partial \Sigma / \partial x$ is non-decreasing in θ (Proposition 4). In this section, we explore solving (21) under more general assumptions.

We begin, first, with the question of whether a solution to (21) exists.[32] Our assumptions on $w(\cdot, \theta)$ and $u(\cdot, \theta)$ ensure that $\Omega(\cdot, \theta)$ is concave and has an interior maximum. The latter conclusion means $x^F(\cdot)$ is bounded. This is relevant since, as Jullien (1996) shows, if $\frac{\partial u}{\partial \theta} \geq 0$, then a bounded $x^F(\cdot)$ implies that we can ignore any allocation function such that $x(\theta_H) > \max_{\theta \in \Theta}\left\{x^F(\theta)\right\}$. Hence, by the monotonicity condition, we can then conclude that the optimal $x(\cdot)$ is bounded. A bounded $x(\cdot)$ helps to ensure the contract space is compact, which is sufficient for existence of an optimal contract.

Theorem 2 ***(Jullien (1996))*** *Assume the standard framework, A1, and*

$$\frac{\partial u}{\partial \theta}(x, \theta) \geq 0 \quad \forall (x, \theta) \in \mathcal{X} \times \Theta, \tag{24}$$

then there exists an optimal contract and the optimal allocation is bounded.

[32]Some articles that deal with the existence question are Guesnerie and Laffont (1984); Page (1992); Jullien (1996) and Rochet and Choné (1998). Each considers slightly different assumptions and offer varying degrees of generality (including relaxing some of the standard framework assumptions). Here, we follow Jullien's approach.

Proof: See Jullien (1996). ■

As we saw in the previous section, (24) is a natural assumption in many contexts—indeed, as in the last example, it can be a *consequence* of the Spence-Mirrlees assumption. It is not, however, always so straightforward. For instance, suppose that, as in Lewis and Sappington (1989), we had

$$u(x,\theta) = -C(x,\theta) = \begin{cases} (\theta - k)\,x - K(\theta) & \text{if } x > 0 \\ 0 \text{ if } x = 0 \end{cases} ;$$

that is, there is a fixed (overhead) cost associated with production that is also a function of type.[33] An interpretation of this cost function is that there is a technology frontier that involves tradeoffs between marginal and fixed costs, with the consequence that a type that has low marginal costs has high fixed costs and *vice versa* (*i.e.,* $K'(\theta) > 0$). Hence, while the Spence-Mirrlees assumption is clearly met, (24) could fail to hold.[34]

Existence is one thing. Characterizing the solution and determining the *uniqueness* of the solution are another. If

$$\Sigma(\cdot,\theta) \text{ is strictly quasi-concave,} \tag{25}$$

then the solution must be unique. Moreover, making this assumption facilitates characterizing the solution, as we saw in Proposition 4. Indeed, we know of no research in which this assumption is not made. Like lemmings, we will follow the herd in this regard. Nonetheless, it is worth taking a moment to consider what this assumption entails. There are many economic justifications we can give to ensure strict quasi-concavity of the *true* surplus function, $\Omega(\cdot,\theta)$. But the *virtual* surplus function, $\Sigma(\cdot,\theta)$, differs from the true surplus function by an amount

$$-\frac{1-F(\theta)}{f(\theta)}\frac{\partial u(\cdot,\theta)}{\partial\theta}.$$

As a general proposition, it is difficult to argue from *economic* principles that $-\partial u(\cdot,\theta)/\partial\theta$ should be strictly quasi-concave.[35] In specific cases, admittedly, one can appeal to economic principles: If, as we did in our extended retailer-supplier example, one is willing to postulate u exhibits separability—

$$u(x,\theta) = b_1(x)\,h(\theta) + b_2(x) + K(\theta)\,,$$

[33] Note this cost function violates our earlier maintained assumption of continuity at $x = 0$.

[34] Note that we could move the $K(\theta)$ term into the reservation utility; *i.e.,* set $U_R(\theta) = K(\theta)$. Indeed, there is often an isomorphism between models in which (24) fails, but A1 holds, and models with type-*dependent* reservation utilities. Put loosely, given that we're "truly" in the standard framework, we can generally expect (24) to hold if A1 holds.

[35] Of course, it is not *necessary* that both components of $\Sigma(\cdot,\theta)$ be strictly quasi-concave for $\Sigma(\cdot,\theta)$ to be strictly quasi-concave. However, it is sufficient and certainly an easy way to verify $\Sigma(\cdot,\theta)$ is strictly quasi-concave.

with $h'(\theta) < 0$—then strict quasi-concavity of $-\partial u(\cdot, \theta)/\partial\theta$ is ensured by the same economic principles that justify u being strictly quasi-concave. But because separability is not a generic property, nor a natural economic property, it is worth appreciating that (25) need not be an innocuous assumption.

Given (24) and (25), we have two of the three assumptions (in addition to the standard framework and A1) required by Proposition 4. The third, that the marginal virtual surplus be non-decreasing in type—*i.e.,* that

$$\frac{\partial\Sigma}{\partial x} \text{ be non-decreasing in } \theta\text{—} \tag{26}$$

can be a fairly stringent condition. Writing out marginal virtual surplus,

$$\frac{\partial\Sigma}{\partial x} = \frac{\partial w}{\partial x} + \frac{\partial u}{\partial x} - \frac{1 - F(\theta)}{f(\theta)} \frac{\partial^2 u}{\partial x \partial\theta},$$

we see that only one of our maintained assumptions—Spence-Mirrlees—applies and it helps with only one term (the middle). Of course in many models, such as our extended retailer-supplier model, the principal's utility, w, is not directly a function of type, so $\partial w/\partial x$ is trivially non-decreasing in type. This leaves the last term. Using $M(\cdot)$ to denote the Mills ratio, a sufficient condition for the third term to be non-decreasing in type is

$$M'(\theta) \frac{\partial^2 u}{\partial x \partial\theta} + M(\theta) \frac{\partial^3 u}{\partial\theta^2 \partial x} \leq 0. \tag{27}$$

The Mills ratio must be positive and, by Spence-Mirrlees, the cross-partial derivative must also be positive. Hence, *sufficient* conditions for (27) to be valid are that $M'(\theta)$ and the third derivative be non-positive. It is difficult to tell a compelling economic story for why the third derivative should be non-positive; hence, this is a problematic assumption. Whether $M'(\theta) \leq 0$, known as the *monotone hazard rate property*, depends on the underlying distribution assumed for the types.[36] Of course, (26) is a only a *sufficient* condition. If in maximizing $\Sigma(x, \theta)$ we discovered that the function $X(\cdot)$ is non-decreasing, the result in Proposition 4 would still hold true. The problem is, except when working with specific functional forms, it is rather difficult to directly assess the monotonicity of $X(\cdot)$.

What if $X(\cdot)$ is not monotonic? Then, although the principal would like to impose a contract with $x(\theta) = X(\theta)$, we know that such an allocation function won't be incentive compatible. In this case, intuition suggests that the principal will try to design a contractual allocation function $x(\cdot)$ as close as possible to $X(\cdot)$, but subject to the condition that $x(\cdot)$ be non-decreasing. Put differently, the constraint that $x(\cdot)$ be non-decreasing must now bind over some interval(s) of types and, hence, we must pay explicit attention to it. But if this condition

[36] Some distributions that satisfy the monotone hazard rate property are the uniform, the exponential, the logistic, and the normal. An example of a distribution *not* satisfying this property is the Pareto distribution.

binds, then we must have $x'(\theta) = 0$ for some interval; that is, we lose the full separation across types we had before (*e.g.,* in our retailer-supplier models). Such non-separation is called *bunching*. BUNCHING

Deriving the optimal contract when $X(\cdot)$ is non-monotonic is a standard optimal-control problem. Note that before we can tackle this control problem, we need to be assured that we can look for the optimal $x(\cdot)$ within the class of absolutely continuous functions. This level of technicality is, however, beyond our scope here and the interested reader is directed to Jullien (1996). We will simply assume the optimal $x(\cdot)$ is absolutely continuous. Hence, the principal's program can be written as:

$$\max_{x(\cdot),y(\cdot)} \int_{\theta_L}^{\theta_H} \Sigma(x(\theta),\theta) f(\theta) d\theta$$

$$\begin{aligned} \text{s.t. } \frac{dx}{d\theta}(\theta) &= y(\theta), \text{ for all } \theta \\ \text{and } y(\theta) &\geq 0, \text{ for a.e. } \theta. \end{aligned}$$

In the language of optimal-control problems, $x(\cdot)$ is the state variable, $y(\cdot)$ is the control variable, and the program imposes a positivity constraint on the control. Introducing the *co-state variable* $\lambda(\cdot)$, we obtain the Hamiltonian for this problem:

$$\mathcal{H}\left[x(\cdot), y(\cdot), \lambda(\cdot), \theta\right] = \Sigma\left[x(\theta), \theta\right] f(\theta) + \lambda(\theta) y(\theta).$$

The necessary first-order conditions are then: $\lambda(\cdot)$ is absolutely continuous and non-positive, $\lambda(0) = \lambda(1) = 0$,

$$y(\theta) \in \arg\max_{y \geq 0} \left[\Sigma(x(\theta),\theta) f(\theta) + \lambda(\theta) y\right], \text{ and} \tag{28}$$

$$\frac{d\lambda}{d\theta} = -\frac{\partial \mathcal{H}}{\partial x} = -\frac{\partial \Sigma}{\partial x}(x(\theta),\theta) f(\theta). \tag{29}$$

Condition (28) is particularly trivial since it points towards simple results: if $\lambda(\theta) < 0$ over some interval, then $y(\theta)$ must be equal to 0 which means that the allocation function $x(\cdot)$ must be constant—there is bunching over this interval. And if $\lambda(\theta) = 0$ over some interval, then obviously $\frac{d\lambda}{d\theta} = 0$ and $x(\theta) = X(\theta)$ over this interval. It follows that the optimal allocation function $x^*(\cdot)$ is obtained by piecing together the increasing parts of $X(\cdot)$ and constants, so that $x^*(\cdot)$ is continuous. We may conclude that:

Proposition 5 ***(Characterization of the optimal contract with bunching)*** *Within the standard framework and assuming A1, (24), and (25), the optimal contract necessarily satisfies the following: The allocation function $x^*(\cdot)$ is continuous, bounded, and for almost all θ either*

- $x^*(\theta) = X(\theta)$*; or*

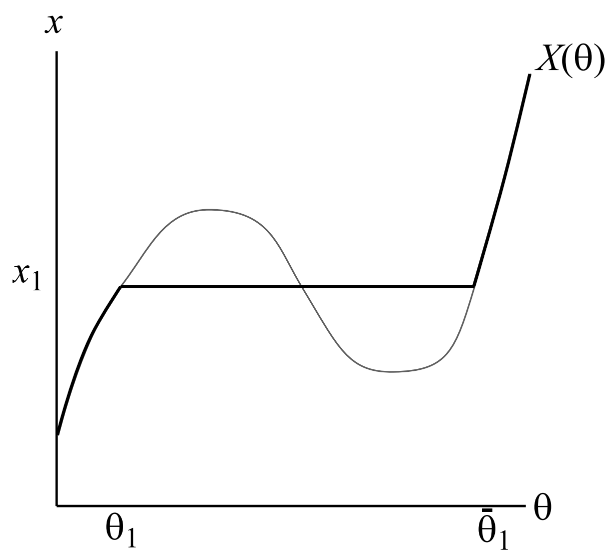

Figure 3: The light curve represents the function that maximizes virtual surplus, $X(\theta)$. Observe it is non-monotonic in type. Hence, the optimal incentive-compatible allocation function, represented by the heavy curve, exhibits *bunching*.

- *$x^*(\theta)$ is constant, equal to x_i, over some interval $\left(\underline{\theta}_i, \bar{\theta}_i\right)$ such that for all $\theta' \in \left(\underline{\theta}_i, \bar{\theta}_i\right)$:*

$$\int_{\theta'}^{\bar{\theta}_i} \frac{\partial \Sigma}{\partial x}(x_i, t)dt \leq 0 \text{ and } \int_{\underline{\theta}_i}^{\bar{\theta}_i} \frac{\partial \Sigma}{\partial x}(x_i, t)dt = 0.$$

Remark 3 *This proposition only provides necessary conditions, as the monotonicity condition may introduce non-convexities in the optimization problem. If, however, we assume that $\Sigma(\cdot, \theta)$ is strictly concave, then the above conditions are necessary and sufficient.*

Figure 3 illustrates.

5.6 Random-allocation mechanisms

Remember that we have heretofore ruled out random-allocation mechanisms by fiat; that is, without considering whether random-allocation mechanisms could be superior to deterministic mechanisms. Here, we briefly reconsider when this is appropriate. Recall from our earlier discussion that (i) additively separability plus risk-neutrality over money mean there is no point to consider random-*payment* mechanisms and (ii) *absent* incentive effects, there is no point to random-*allocation* mechanisms. We're not, however, in a world without incentive effects; hence, we might ask whether a random-allocation mechanism eases the truth-telling constraints for the principal enough to compensate for the risk imposed (recall both principal and agent are risk-averse—at least weakly—with respect to allocations).

If we suppose that $\partial u/\partial\theta$ is convex (at least weakly) in x—an assumption we've essentially made before to ensure the strict quasi-concavity of the *virtual* surplus function—then the answer is *no.* Given this assumption plus the assumed strict concavity of $\Omega(\cdot,\theta)$, we see that

$$\Sigma(\cdot,\theta) = \Omega(\cdot,\theta) - \frac{1-F(\theta)}{f(\theta)}\frac{\partial u}{\partial\theta}$$

is a concave function. Let $x(\cdot)$ be any monotonic *deterministic*-allocation function and let $\tilde{x}(\cdot)$ be any random-allocation function with the property that $\mathbb{E}\{\tilde{x}(\theta)\} = x(\theta)$. Then Jensen's inequality implies

$$\mathbb{E}_{\tilde{x}(\theta)}\{\Sigma[\tilde{x}(\theta),\theta]\} < \Sigma[x(\theta),\theta]$$

for all θ. That is, taking into account the incentive constraint, the principal's expected utility is less, type by type. Hence, randomizing based on a feasible deterministic-allocation mechanism cannot improve the principal's expected utility. See Maskin (1981) for an analysis when $\partial u/\partial\theta$ is not convex.

6 The Hidden-Knowledge Model

In this section, we consider a model that, at first, seems quite different than contractual screening, but which ultimately shares many similarities to it. In particular, we consider the *hidden-knowledge model.* In this model, unlike above, the principal and agent are *symmetrically* informed at the time they enter into a contractual arrangement. *After* contracting, the agent acquires private information (his hidden knowledge). As an example, suppose the principal employs the agent to do some task—for instance, build a well on the principal's farm—initially, both parties could be symmetrically informed about the difficulty of the task (*e.g.,* the likely composition of the rock and soil, how deep the water is, etc.). However, once the agent starts, he may acquire information about how hard the task really is (*e.g.,* he alone gains information that better predicts the depth of the water). HIDDEN-KNOWLEDGE

This well-digging example reflects a general problem. In many employment situations, the technological, organizational, market, and other conditions that an employee will face will become known to him only after he's been employed by the firm. This information will affect how difficult his job is, and, thus, his utility. Similarly, think of two firms that want to engage in a specific trade (*e.g.,* a parts manufacturer and an automobile manufacturer who contract for the former to supply parts meeting the latter's unique specifications). Before the contract is signed, the supplier may not know much about the cost of producing the specific asset and the buyer may have little knowledge about the prospects of selling the good to downstream consumers. These pieces of information will flow in during the relationship—*but after contracting*—and once again a hidden-knowledge framework is more appropriate for studying such a situation.

This difference in timing is reflected in the participation constraint for this problem. When considering a contract, $\langle x(\cdot), s(\cdot)\rangle$, the agent compares his *expected utility* if he accepts the contract,

$$\mathbb{E}\{u(\theta)\} = \mathbb{E}\{s(\theta) + u[x(\theta), \theta]\},$$

to his expected utility if he refuses the contract; that is, to $U_R^a \equiv \mathbb{E}\{U_R(\theta)\}$. Acceptation or refusal of the contract cannot depend upon the, as yet, unrealized state of nature: Unlike the screening model, the participation decision is *not* contingent on type.

Why is this discussion so important? Because it turns out that, at least in the standard framework, the hidden-knowledge model has an extremely simple solution. To see this, we invoke the assumptions of the standard framework, including the Spence-Mirrlees assumption, A1. In addition, we assume—consistent with the assumptions of the standard framework—that $\partial\Omega/\partial x$ is non-decreasing in θ. Consequently, we know the first-best allocation, $x^F(\cdot)$, is non-decreasing. We can then be sure from Theorem 1 that there exists a transfer function, $\hat{s}^F(\cdot)$, such that $\langle x^F(\cdot), \hat{s}^F(\cdot)\rangle$ is a direct-revelation mechanism.[37] Note that because $\hat{s}^F(\cdot)$ is defined by (16), it is defined up to a constant that can be chosen by the principal to ensure the agent meets his participation constraint. In particular, it can be chosen so that the agent's expected utility equals his non-participation expected utility:

$$\int_{\theta_L}^{\theta_H} \left[\hat{s}^F(\theta) + u\left(x^F(\theta), \theta\right)\right] f(\theta) d\theta = \int_{\theta_L}^{\theta_H} U_R(\theta) f(\theta) d\theta.$$

With this mechanism, the principal's expected utility becomes:

$$\begin{aligned}\int_{\theta_L}^{\theta_H} \left[w(x^F(\theta), \theta) - \hat{s}^F(\theta)\right] f(\theta) d\theta &= \int_{\theta_L}^{\theta_H} \left[\Omega(x^F(\theta), \theta) - U_R(\theta)\right] f(\theta) d\theta \\ &= \int_{\theta_L}^{\theta_H} \Omega(x^F(\theta), \theta) f(\theta)\, d\theta - U_R^a.\end{aligned}$$

Since the *ex post* efficient allocation—that is, $x^F(\cdot)$—maximizes the integrand in the last integral, the principal obtains the highest possible expected utility with this mechanism. Hence, $\langle x^F(\cdot), \hat{s}^F(\cdot)\rangle$ is optimal and we see, therefore, that the first-best allocation will be achieved with hidden-knowledge, in contrast to the less desirable equilibrium of the screening model:

[37] The transfer function $\hat{s}^F(\cdot)$ is *a priori* different from the full-information transfer function, $s^F(\cdot)$, because the latter is not subject to revelation constraints.

Proposition 6 *In a hidden-knowledge model, in which the agent fully commits to the contract* before *learning his type, which satisfies the assumptions of the standard framework, and in which $\partial\Omega/\partial x$ is non-decreasing in θ, then the equilibrium allocation is the* ex post *efficient allocation.*

To gain intuition for this result, return to our two-type example from Section 3, but suppose now that the agent doesn't learn his type until *after* contracting with the principal. Observe that we can reduce the agent's information rent *on average* by paying him less than his full-information payment if he announces he is the low type (*i.e.,* type I) and more than his full-information payment if he announces he is the high type (*i.e.,* type E): Now set the payments to be

$$\begin{aligned}\hat{s}_I^F &= C_I\left(x_I^F\right) - \gamma\text{; and}\\ \hat{s}_E^F &= C_E\left(x_E^F\right) + \eta.\end{aligned}$$

Since the agent doesn't learn his type until after contracting, the participation constraint is

$$\begin{aligned}f \times \left(\hat{s}_I^F - C_I\left(x_I^F\right)\right) + (1-f) \times \left(\hat{s}_E^F - C_E\left(x_E^F\right)\right) &= -f\gamma + (1-f)\,\eta\\ &\geq 0.\end{aligned}$$

We also need direct revelation, which, for type I, means

$$\begin{aligned}\hat{s}_I^F - C_I\left(x_I^F\right) &\geq \hat{s}_E^F - C_I\left(x_E^F\right)\text{; or}\\ -\gamma &\geq \eta + \left[C_E\left(x_E^F\right) - C_I\left(x_E^F\right)\right].\end{aligned}$$

Treating these two constraints as equalities, we can solve for γ and η:

$$\begin{aligned}\eta &= f \times \left[C_I\left(x_E^F\right) - C_E\left(x_E^F\right)\right]\text{; and}\\ \gamma &= (1-f) \times \left[C_I\left(x_E^F\right) - C_E\left(x_E^F\right)\right].\end{aligned}$$

Provided these also satisfy type E's revelation constraint, we're done. But they do, since

$$\begin{aligned}\hat{s}_E^F - C_I\left(x_E^F\right) &= \eta\\ &= -\gamma + C_I\left(x_E^F\right) - C_E\left(x_E^F\right)\\ &> -\gamma + C_I\left(x_I^F\right) - C_E\left(x_I^F\right)\\ &= \hat{s}_I^F - C_E\left(x_I^F\right)\end{aligned}$$

(the inequality follows because, recall, $C_I(\cdot) - C_E(\cdot)$ is increasing).

Note the phrase *"in which the agent fully commits to the contract"* that constitutes one of the assumptions in Proposition 6. Why this assumption? Well suppose that, after learning his type, the agent could quit (a reasonable assumption if the agent is a person who enjoys legal protections against slavery). If his payoff would be less than $U_R(\theta)$ if he played out the contract, he would do better to quit.[38] To keep the agent from quitting, the principal would have to design the contract so that $U(\theta, \theta) \geq U_R(\theta)$ for all θ. But then this is just the screening model again! In other words, the hidden-knowledge model reverts to the screening model—with all the usual conclusions of that model—if the agent is free to quit (in the parlance of the literature, if *interim participation constraints* must be met). Even if anti-slavery protections don't apply (*e.g.,* the agent is a firm), interim participation could still matter; for instance, in the last example, if γ is too big, then the agent may not have the financial resources to pay it (it would bankrupt him). Alternatively, in nations with an English law tradition, γ could be perceived as a penalty, and in many instances the courts will refuse to enforce contracts that call for one party to pay a penalty to another. In short, because *interim* participation constraints are often a feature of the real world, many situations that might seem to fit the hidden-knowledge model will ultimately prove to be screening-model problems instead.

INTERIM PARTICIPATION

7 Concluding Remarks

The screening model and variants, such as the hidden-knowledge model, are widely used models in economics. These models capture a fundamental tension in many contractual settings: One party has superior information about a state of nature relevant to both. Like any advantage, the party with the superior information will seek to capture some rents from this. In response, the other party, particularly if she has bargaining power to preserve, will seek to design a contract that limits the rents of the better-informed party. As we saw, in general, this will lead to distortions in physical allocations and, hence, create deadweight loss.

References

Caillaud, Bernard, Roger Guesnerie, Patrick Rey, and Jean Tirole, "Government Intervention in Production and Incentives Theory: A Review of Recent Contributions," *RAND Journal of Economics*, Spring 1988, *19* (1), 1–26.

Caillaud, Bernard and Benjamin E. Hermalin, "The Use of an Agent in a Signalling Model," *Journal of Economic Theory*, June 1993, *60* (1), 83–113.

Edlin, Aaron S. and Benjamin E. Hermalin, "Contract Renegotiation in Agency Problems," 1997. Working Paper, Department of Economics, University of California at Berkeley.

Gibbard, Allan, "Manipulation of Voting Schemes," *Econometrica*, July 1973, *41* (4), 587–601.

[38] Observe we're assuming that $U_R(\cdot)$ doesn't change—in particular, doesn't diminish—after the agent signs the contract; that is, his outside option remains the same over time.

Gonik, Jacob, "Tie Salesmen's Bonuses to Their Forecasts," *Harvard Business Review*, May-June 1978, pp. 116–122.

Green, Jerry and Jean-Jacques Laffont, "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods," *Econometrica*, March 1977, *45* (2), 427–438.

Guesnerie, Roger and Jean-Jacques Laffont, "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-managed Firm," *Journal of Public Economics*, December 1984, *25* (3), 329–369.

Jullien, Bruno, "L'impact des options exterieures sur les echanges en information asymetrique (The Impact of Outside Options on Exchange under Asymmetric Information)," *Revue Economique*, mai 1996, *47* (3), 437–446.

Laffont, Jean-Jacques and Jean Tirole, *A Theory of Incentives in Procurement and Regulation*, Cambridge, MA: MIT Press, 1993.

Lewis, Tracy R. and David E.M. Sappington, "Countervailing Incentives in Agency Problems," *Journal of Economic Theory*, December 1989, *49* (2), 294–313.

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

Maggi, Giovanni and Andres Rodriguez-Clare, "On the Countervailing Incentives," *Journal of Economic Theory*, June 1995, *66* (1), 238–263.

Maskin, Eric, "Randomization in Incentive Schemes," 1981. Mimeo, Harvard University.

Milgrom, Paul and Chris Shannon, "Monotone Comparative Statics," *Econometrica*, January 1994, *62* (1), 157–180.

Myerson, Roger, "Incentive Compability and the Bargaining Problem," *Econometrica*, 1979, *47*, 61–73.

Page, Frank H., "Mechanism Design for General Screening Problems with Moral Hazard," *Economic Theory*, April 1992, *2* (2), 265–281.

Rochet, Jean-Charles and Philippe Choné, "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, July 1998, *66* (4), 783–826.

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

Wilson, Robert B., *Nonlinear Pricing*, Oxford: Oxford University Press, 1993.