The One Who Controls the Information Appropriates Its Rents*

Péter Eső†
Northwestern University, Kellogg MEDS
Balázs Szentes‡
University of Chicago, Department of Economics

April 2003

Abstract

We analyze a Principal-Agent model where the Principal can influence the precision of the Agent's private information by releasing, without observing, additional signals that refine the Agent's initial private type-estimate. We derive the Principal's optimal contract, whose terms incorporate an information disclosure policy, and characterize its properties. We show that in the optimal contract, the Principal always releases the information that she controls. Moreover, we show that the Principal is able to implement the same allocation and obtain the same utility as if she could observe the realizations of the additional signals.

Keywords: Adverse Selection, Information Disclosure, Mechanism Design
JEL Classification: C72, D82, D83

*We thank Yeon-Koo Che, Kim-Sau Chung, Eddie Dekel, Drew Fudenberg, Peter Klibanoff, Phil Reny, Jim Schummer, Nancy Stokey, seminar participants at the London School of Economics, Northwestern University, Princeton University, the University of Chicago, the University of Pennsylvania, the Virginia Polytechnical Institue, and the 2002 Midwest Theory meetings at Notre Dame for comments.

†2001 Sheridan Rd, Evanston, IL 60208. Email: eso@northwestern.edu.

‡1126 E. 59th St, Chicago, IL 60637. Email: szentes@uchicago.edu.

1 Introduction

We analyze a Principal-Agent relationship where the Principal can influence the precision of the Agent's original private information by controlling the release of additional signals to the Agent. These signals (like the Agent's original type-estimate) are only observable to the Agent. We derive the Principal's optimal contract, whose terms incorporate an information disclosure policy, and characterize several of its properties.

Besides the theoretical interest in this problem, the model is motivated by several applications. Suppose, for example, that the Department of Defense (Army, for short, acting as the Principal) wants to procur large quantities of a new vaccine from a commercial pharmaceutical company (the supplier, or Agent). The Army's laboratory, the vaccine's developer, has extensive information regarding the vaccine (e.g., its chemical properties), which might be important for the supplier in determining the costs of production. However, the Army has no expertise in the mass production of vaccines, therefore, it cannot evaluate the effect of its own information on the supplier's costs. If the Army discloses what it knows to the supplier, then all this information becomes the supplier's private knowledge, which it can use at its own advantage. The question is then how the Army should structure the contract with the supplier, including whether and how it should disclose the information that it controls but cannot evaluate, in order to maximize its objective function.

Another example could be that of the owner of a proprietary operating system (the Principal) hiring another software company (the Agent) to develop an application, say, a program synchronizing the computer and other devices via the internet. The inner workings of this new program are privately known by the Agent. The Principal may release information to the Agent regarding the operating system (e.g., its communication standards), which would help the Agent better gauge the difficulty of adapting its application to the Principal's system. Lacking the Agent's know-how, however, the Principal does not know whether these revelations would make the Agent's task harder or easier. Therefore, one can model the situation as if the Principal could disclose, without observing, the realizations of shocks that modify the Agent's private cost estimate.[1] The question is then whether and how the Principal should disclose these signals, and who will be better off due to the availability of more information.

These two examples can be generalized in a model where the Principal (she) is procuring goods or services from the Agent (he). Before they contract, the Agent has private information about his productivity, which can be made more precise by the Principal by providing a better job description, or by releasing other clues that the Agent can use to evaluate how the job matches his expertise. However, the Principal does not know how this additional information affects the Agent, that is, she cannot observe the realization of the signals she can disclose.[2] Our goal is then to find the optimal contract that the Principal can offer, which governs their bilateral trade *and* the Principal's information disclosure.

The Principal can influence how much the Agent can learn about his own type in other contractual situations as well. The producer of a durable good can offer a trial period to her buyer during which the buyer can learn the uses of, and therefore his own valuation for, the good. Another interesting example is that of an employee whose productivity depends both on his privately known talent and the quality or matching interests of his colleagues (e.g., a researcher). The potential employer can bring him in for a round of interviews, during which the candidate can learn more about the workplace and hence his potential productivity in that environment.[3]

[1] This situation is quite common in outsourcing financial, legal, or IT services: the buyer often does not know whether some of her proprietary information makes the provider's job easier or more difficult, because that judgment would already require the provider's know-how.

[2] For example, the Principal may be able to give a sample job to the Agent. In this case, we should assume that only the Agent (and not the Principal) can reliably evaluate the Agent's productivity based on the sample job, say, because the Agent can exert hidden effort that substitutes for ability on this job.

[3] Another purpose of flying in a candidate is to find out his qualities. However, note that the candidate (unlike the department) has powerful incentives to exert effort and look like a perfect fit.

In all these examples, it is interesting to ask whether and how the Principal should release her clues (the realizations of the additional signals) to the Agent. Starting from a situation with asymmetric information (recall that at the outset, the Agent already has a private type-estimate), the disclosure of information may or may not lead to more social surplus, but nevertheless, it will provide the Agent with private knowledge that he may be able to draw rents upon. Therefore, we are interested in finding the optimal contract between the Principal and the Agent (the mechanism that maximizes the Principal's utility). This contract will not only specify transfers contingent on the contractible variable, but it will also incorporate an information disclosure policy.[4]

In this paper, we characterize the contract that maximizes the Principal's utility. Our main result is that in this mechanism, all available additional information is disclosed by the Principal. Moreover, we show that in the optimal mechanism, the Principal's utility is the same as if she could observe the realizations of the signals whose release she can control. In other words, although the Agent enjoys information rents due to his original private information (his initial type-estimate), in the optimal mechanism, the Principal appropriates all rents for the additional signals that she controls. That is, in a second-best world, the one who controls the flow of information appropriates the rents of information.[5]

This economic lesson is valid in other environments as well. In Eső and Szentes (2002), we show that similar results hold in a special multi-agent adverse selection model, specifically, in an auction for a single good with risk neutral buyers and independent private valuations. There, we exhibit a simple mechanism, dubbed the "handicap auction," which is optimal when the seller controls the release of information that refines the buyers' original value-estimates. The properties of the optimal

[4]For example, outsourcing contracts may specify a trial (learning) period at the end of which the provider has the option to renew at a pre-specified lower price.

[5]This may be the reason why Newman (the mean postal worker) exclaims in one of the Seinfeld episodes, "If you control the mail, you control — information!"

handicap auction are similar to those of the optimal contract in the present model. For example, the seller discloses all information she controls, implements the same allocation rule as if she could observe the additionally released signals, and extracts all rents for them. However, the specification of the optimal handicap auction is considerably simpler than that of the optimal contract in the present model due to the special form of the payoffs in the monopolist's selling problem.

Information disclosure has been studied in the agency literature mostly in the context of the monopolist's selling problem. Milgrom and Weber (1982) investigate the Principal's (seller's) incentives to disclose public (as opposed to private) information in the standard auction formats in an environment with interdependent values. Ottaviani and Prat (2001) show that a monopolist is always better off by committing to reveal public signals affiliated to the buyer's information, provided she can adjust the selling mechanism in response to the additional signals. Another fruitful area has been the study of the buyer's incentives to acquire private information in different auction formats (see Persico (2000), Compte and Jehiel (2001) and the references therein). In contrast to these models, here, the Principal will decide whether to disclose (without observing) private information to the Agent. There are very few papers discussing the value for the monopolist of the private information of the buyer (see Lewis and Sappington (1994), and Bergemann and Pesendorfer (2002)). The two crucial differences between these papers and ours is that in our model, the Agent already has private information when the game starts, and that we allow the Principal to fully incorporate the rules of information disclosure into the optimal contract (in the two papers mentioned above, the seller first chooses a disclosure policy and then optimizes the selling mechanism without prior commitment).

An interesting strand of contract theory (see Caillaud et al. (1992) and the references therein) investigates adverse selection models where the Principal can contract on a noisy estimator of the Agent's action instead of the action itself. It is shown that (under certain conditions) the Principal can achieve the same allocation with noisy observation as in the case when the action is verifiable without error. The key in these models is that the optimal mechanism can be implemented via a menu of contracts where the transfer is linear in the Agent's action. Since the Agent is risk neutral, the optimal contract will be robust to any (unbiased) noise in the observation of the action. Note that our problem is very different as the signals that the Principal can reveal pertain to the type of the Agent, and actions are contractible without noise.

The paper is structured as follows. In the next section, we outline the model and introduce the necessary notation. In Section 3, we first derive the optimal contract for the case when the Principal can observe the additional signals that refine the Agent's type-estimate. Then, we show that the outcome can be attained even if the Principal cannot directly observe the additional signals, and further characterize the optimal contract. We conclude in Section 4.

2 The Model

Assume that two parties, a Principal and an Agent, are about to enter a contractual relationship. The contractible decision (which could be the Agent's production, the Principal's or the two party's joint activity) is represented by a real number, x, belonging to a compact interval $[\underline{x}, \bar{x}]$. The Agent has a characteristic, called type or ability, which is the sum of two independent random variables, θ and s, and influences the utilities of both parties. Here θ, called the type-estimate, is the Agent's private information, while s, called the shock, is initially not observable to him. However, the Principal can disclose, without observing, the realization of s to the Agent.[6]

The type-estimate, θ, is distributed on $[0, 1]$ according to a cumulative distribution function, F, with a positive density, f. The monotone hazard rate condition is satisfied,

[6] We allow, but not require, that the Principal can release (without observing) other signals that are correlated with s as well.

that is, the hazard rate, $H = (1 - F)/f$, is weakly decreasing. We do not make any assumption regarding the distribution of the shock except that s is independent of θ.

Utilities are quasilinear and perfectly transferable.[7] Denoting the transfer from the Principal to the Agent by t, we can write the Agent's utility as

$$U = u(x, \theta + s) + t,$$

and the Principal's utility as

$$V = v(x, \theta + s) - t.$$

We will use subscripts to refer to partial derivatives, for example, $u_1 = \partial u/\partial x$, and assume that both u and v thrice differentiable.

We assume that u and v are concave in x, that is, the decision variable has a weakly decreasing marginal effect on the utilities of both parties (they have increasing marginal costs or decreasing marginal utilities from the decision). We also assume that the Agent's utility is monotonic in his ability, and normalize it to be "good," that is, $u_2 > 0$. We assume that u_2 is uniformly continuous.

We impose the following additional conditions on the functions u and v. First, we assume that the Spence–Mirrlees (or single-crossing) condition holds: the Agent's marginal utility from x is strictly increasing in his ability, $u_{12} > 0$. In other words, the single-crossing condition implies that an increase in the decision variable (activity) is more beneficial (or less costly) for an Agent with a higher type. We require that $u_{112} \geq 0$, that is, this "advantage" of a high-type Agent (compared to a low type) from increasing the decision is greater if x, the level of the decision variable, is higher. We also require that $u_{122} \leq 0$, that is, the marginal utility of the decision is concave in the Agent's type (it is increasing at a weakly decreasing rate). Finally, we assume that

[7] Our results go through with quasilinear but imperfectly transferable utilities as well.

$v_{12} \geq 0$, that is, the Principal's marginal utility from the decision is non-decreasing in the Agent's ability. These conditions on the utility functions are standard in Principal-Agent models and hold in most applications.[8]

In order to see that these assumptions are reasonable (and hold) in the applications mentioned in the Introduction as well, let us interpret them in the context of our leading examples. In the procurement or outsourcing application, x can be thought of as the amount of goods or services provided by the Agent to the Principal. The type of the Agent, $\theta + s$, is a parameter influencing the supplier's production costs. The Agent initially only knows θ, and the Principal can disclose s to him. Write the utilities for the Agent and Principal as

$$\begin{aligned} U &= t - c(x, \theta + s), \\ V &= v(x) - t, \end{aligned}$$

respectively. Here, $c(x, \theta + s)$ is the Agent's cost function, $v(x)$ is the Principal's utility from consuming quantity x, and t is monetary transfer from the Principal to the Agent.

The assumptions on the derivatives of the utility functions apply as follows. The Agent's type is normalized so that cost is decreasing in type, $c_2 < 0$. The single-crossing condition becomes $c_{12} < 0$, that is, the higher the Agent's type, the lower his marginal cost of production. The Principal's utility function is concave in x (her marginal utility from consuming the good is weakly decreasing). The Agent's utility is concave in x as well, meaning that he has weakly increasing marginal cost of production (for a given type). The conditions on the third derivatives of the Agent's utility function hold if $c_{112} \leq 0$, the marginal cost function is "less convex" for higher types (not only that the marginal cost is lower for higher types, it also increases more slowly in x), and $c_{122} \geq 0$, the marginal cost decreases in the Agent's type at a decreasing rate (the Agent's type

[8] See the discussion of assumptions A1–A10 in Chapter 7 of Fudenberg and Tirole (1991).

exhibits decreasing returns at reducing the marginal cost).[9] Finally, the Principal's utility does not depend on the Agent's type, so $v_{12} \geq 0$ holds.

An even more specialized version of this model would be one where the Principal offers an indivisible good or contract to the Agent, and x is interpreted as the probability of trade. The good or contract, if awarded to the Agent, is worth $(\theta + s)$ to him, less the monetary transfer (if any) he has to make to the Principal. For simplicity, let the monetary value of the good or contract be zero for the Principal (the same as the value of no trade). Under these assumptions, $U = (\theta + s)x + t$ and $V = -t$. This is essentially a model of auctioning a single item to a single buyer, where the Principal is the seller and the Agent is the buyer.

3 Results

In Subsection 3.1, we first look at the case where the Principal can observe the realization of the shock (the additional signal, s). We solve for the Principal's optimal contract in this benchmark case. Then, in Subsection 3.2, we return to our original model, where the Principal can decide whether to release, but cannot actually observe, the realization of the shock. We characterize the mechanism maximizing the Principal's utility by relating it to the optimal contract found in the benchmark case.

3.1 The Optimal Mechanism When the Principal Can Observe the Shock

Assume, in this subsection and for benchmarking purposes only, that the Principal can actually observe the shock, s, while the Agent cannot. Importantly, the Principal commits to the contract (mechanism) before observing the shock.

[9] A simple form of the cost function that satisfies these conditions would be $c(x, \theta + s) = xe^{-(\theta+s)}$.

By the Revelation Principle, without loss of generality, we can restrict our attention to incentive compatible direct mechanisms where the Agent reports to the Principal his type-estimate, θ, and then the Principal sets the decision and the transfer as a function of the reported type-estimate and the realization of the shock, $x(\theta, s)$ and $t(\theta, s)$, respectively. Incentive compatibility of a mechanism (x, t) means that

$$\mathbf{E}_s[u(x(\theta', s), \theta + s) + t(\theta', s)] \leq \mathbf{E}_s[u(x(\theta, s), \theta + s) + t(\theta, s)] \quad \text{for all } \theta \text{ and } \theta'. \quad (1)$$

The problem of characterizing incentive compatible mechanisms and then finding the optimal contract can be solved by using techniques of Bayesian mechanism design.[10] The (mainly technical) complication is that incentive compatibility, (1), should hold in expectation with respect to s, and the order of taking expectation and applying the utility function cannot be reversed. Incentive compatible allocation rules may not be fully monotonic (in θ and s), and hence it is not immediate that the Agent's utility (together with the Principal's objective function) can be written in the familiar integral form.

In the following lemma, we provide a necessary and a (stronger) sufficient condition for incentive compatiblity of a mechanism in the benchmark case. Subsequently, in Theorem 1, we characterize the Principal's optimal mechanism.

Lemma 1 *Assume that the Principal can observe the shock after having committed to a contract with the Agent. If a mechanism (x, t) is incentive compatible then*

$$U(\theta) = U(0) + \int_0^\theta \mathbf{E}_s[u_2\left(x\left(z, s\right), z + s\right)]dz, \quad (2)$$

where $U(\theta)$ is the expected payoff of the Agent with type-estimate θ. Moreover, if (2) holds and $x(\theta, s)$ is weakly increasing in θ, then the mechanism is incentive compatible.

[10] For the origins of these techniques, see Mirrlees (1971). For a textbook treatment, see, for example, chapter 7 of Fudenberg and Tirole (1991).

Proof. Suppose that (1) holds. The Agent's expected payoff with type-estimate θ if he reported θ' is

$$\begin{aligned}\mathbf{E}_s[u(x(\theta',s),\theta+s)+t(\theta',s)] &= \mathbf{E}_s[u(x(\theta',s),\theta'+s)+t(\theta',s)] \\ &\quad +\mathbf{E}_s[u(x(\theta',s),\theta+s)-u(x(\theta',s),\theta'+s)] \\ &= U(\theta')+\mathbf{E}_s[u(x(\theta',s),\theta+s)-u(x(\theta',s),\theta'+s)].\end{aligned}$$

Therefore, we can rewrite (1) as

$$U(\theta)\geq U(\theta')+\mathbf{E}_s[u(x(\theta',s),\theta+s)-u(x(\theta',s),\theta'+s)] \quad \text{for all } \theta \text{ and } \theta'. \tag{3}$$

Reversing the rules of θ and θ', (1) is equivalent to the following inequalities

$$\begin{aligned}\mathbf{E}_s[u(x(\theta',s),\theta+s)-u(x(\theta',s),\theta'+s)] &\leq U(\theta)-U(\theta') \\ &\leq \mathbf{E}_s[u(x(\theta,s),\theta'+s)-u(x(\theta,s),\theta+s)].\end{aligned}$$

From these inequalities, it immediately follows that U is continuous and weakly increasing, since $u_2>0$. Therefore U' exists almost everywhere. Since u is continuously differentiable, by the Fundamental Theorem of Calculus, assuming that $\theta>\theta'$, we can rewrite the previous chain of inequalities as

$$\frac{\mathbf{E}_s\left[\int_{\theta'}^{\theta}u_2(x(\theta',s),z+s)dz\right]}{\theta-\theta'}\leq\frac{U(\theta)-U(\theta')}{\theta-\theta'}\leq\frac{\mathbf{E}_s\left[\int_{\theta'}^{\theta}u_2(x(\theta,s),z+s)dz\right]}{\theta-\theta'}. \tag{4}$$

By letting θ converge to θ' and using that u_2 is absolute continuous, we conclude that $\mathbf{E}_s[u_2(x(\theta',s),\theta'+s)]\leq U'_+(\theta)$ whenever $U'_+(\theta)$ exists. Similarly, by taking θ' to θ, we have $U'_-(\theta)\leq\mathbf{E}_s[u_2(x(\theta,s),\theta+s)]$ whenever $U'_-(\theta)$ exists. Since U is weakly increasing, its derivative exists almost everywhere (see, for example, Royden (1967), Theorem 2), hence

$$U'(\theta) = \mathbf{E}_s[u_2(x(\theta, s), \theta + s)] \tag{5}$$

almost everywhere.

From (4) it follows that

$$\frac{U(\theta) - U(\theta')}{\theta - \theta'} \leq \sup_{z \in (\theta', \theta)} \mathbf{E}_s u_2(x(\theta, s), z + s) \leq \sup_{x,z} \mathbf{E}_s u_2(x, z + s).$$

Since u_2 is uniformly continuous, $\mathbf{E}_s u_2(x, z+s)$ is continuous, therefore $\sup_{x,z} \mathbf{E}_s u_2(x, z+s)$ is finite. We can conclude that U is Lipschitz continuous, which implies that it is also absolute continuous, and hence it can be recovered from its derivative (Royden (1967), Theorem 13). That is, for all $\theta \in [0, 1]$,

$$U(\theta) = U(0) + \int_0^\theta U'(z)\, dz.$$

From this together with (5) we can conclude that (2) is indeed satisfied.

Suppose now that (2) holds and x is weakly increasing in its first argument, and $\theta > \theta'$.Then from (2)

$$\begin{aligned} U(\theta) - U(\theta') &= \int_{\theta'}^{\theta} \mathbf{E}_s[u_2(x(z, s), z + s)]dz \\ &\geq \int_{\theta'}^{\theta} \mathbf{E}_s[u_2(x(\theta', s), z + s)]dz \\ &= \mathbf{E}_s[u(x(\theta', s), \theta + s)] - \mathbf{E}_s[u(x(\theta', s), \theta' + s)] \end{aligned}$$

where the inequality follows from x being weakly increasing in its first argument and the single-crossing condition, $u_{12} > 0$. But, this is just (3) which was seen to be equivalent to (1). If $\theta < \theta'$ then a similar argument can applied to show that there is no incentive to deviate upward. ■

By Lemma 1, in any incentive compatible mechanism, the Agent's expected utility can be written in the familiar integral form (where the integrand is a function of the decision rule), and this, together with a monotonicity condition on the decision rule, is sufficient for incentive compatibility.[11] Next, we use this lemma to derive the optimal mechanism in the benchmark case.

Theorem 1 *Assume that the Principal can observe the shock after having committed to a contract with the Agent. In the mechanism that maximizes the Principal's utility,* $x(\theta, s)$ *is weakly increasing in both of its arguments and maximizes*

$$v(x, \theta + s) + u(x, \theta + s) - H(\theta)\, u_2(x, \theta + s), \tag{6}$$

where $H = (1 - F)/f$ *is the hazard rate. The Agent's expected payoff is (2) with* $U(0) = 0$*. The Principal's expected utility is*

$$\mathbf{E}_s \left[\int_0^1 \left\{ v(x(\theta, s), \theta + s) + u(x(\theta, s), \theta + s) - H(\theta)\, u_2(x(\theta, s), \theta + s) \right\} dF(\theta) \right].$$

Proof. The Principal's expected surplus equals the total surplus minus the Agent's surplus, which, using (2), can be written as

$$\mathbf{E}_s \left[\int_0^1 \left\{ v(x(\theta, s), \theta + s) + u(x(\theta, s), \theta + s) - \int_0^\theta u_2(x(z, s), z + s) dz \right\} dF(\theta) \right] - U(0)\, .$$

Notice that

$$\begin{aligned} \int_0^1 \int_0^\theta u_2(x(z, s), z + s) dz dF(\theta) &= \int_0^1 \int_z^1 u_2(x(z, s), z + s) dF(\theta) dz \\ &= \int_0^1 (1 - F(z))\, u_2(x(z, s), z + s) dz. \end{aligned}$$

[11] Lemma 1 is not an equivalence statement due to the presence of the shock. However, this form is exactly what we need for characterizing the optimal mechanism in the benchmark case.

Therefore, the Principal's expected surplus is

$$\mathbf{E}_s\left[\int_0^1 \{v(x(\theta,s),\theta+s)+u(x(\theta,s),\theta+s)-H(\theta)u_2(x(\theta,s),\theta+s)\}\ dF(\theta)\right]-U\left(0\right).$$

We can maximize the integrand pointwise and set $U\left(0\right)$ equal to zero. That is, for all θ and s, $x\left(\theta,s\right)$ is defined as the maximizer of

$$v(x,\theta+s)+u(x,\theta+s)-H(\theta)u_2(x,\theta+s),$$

which is expression (6). By assumption, u and v are concave and u_2 is convex in their first argument, so $x\left(\theta,s\right)$ is well-defined. Furthermore, it is implicitly defined by

$$v_1(x\left(\theta,s\right),\theta+s)+u_1(x\left(\theta,s\right),\theta+s)-H(\theta)u_{12}(x\left(\theta,s\right),\theta+s)=0.$$

We now show that $x(\theta,s)$ is increasing in θ and s, hence Lemma 1 applies, and the mechanism is incentive compatible. By the Implicit Function Theorem, $x_1\left(\theta,s\right)$ exists and equals

$$-\frac{v_{12}(x,\theta+s)+u_{12}(x,\theta+s)-H'(\theta)u_{12}(x,\theta+s)-H(\theta)\,u_{122}(x,\theta+s)}{v_{11}(x,\theta+s)+u_{11}(x,\theta+s)-H(\theta)\,u_{112}(x,\theta+s)}$$

at $x=x(\theta,s)$. Since u_{12} is positive, v_{12}, u_{112} are non-negative, and v_{11}, u_{11}, u_{122} are non-positive, this expression is positive. Also, by the Implicit Function Theorem, $x_2\left(\theta,s\right)$ exists and equals

$$-\frac{v_{12}(x,\theta+s)+u_{12}(x,\theta+s)-H\left(\theta\right)\ u_{122}(x,\theta+s)}{v_{11}(x,\theta+s)+u_{11}(x,\theta+s)-H\left(\theta\right)\ u_{112}(x,\theta+s)}$$

at $x=x(\theta,s)$. Since u_{12} is positivie, v_{12}, u_{112} are non-negative, and v_{11}, u_{11}, u_{122} are non-positive, this expression is positive. ■

The expression (6) in Theorem 1 may be called the shock-adjusted virtual social surplus. Due to asymmetric information, in the second-best, the allocation rule is chosen to maximize (6) instead of the shock-adjusted social surplus, $v(x, \theta + s) + u(x, \theta + s)$. The Principal's utility equals the expected social surplus minus the rents paid for the Agent's private information (whose only source is θ).

The optimal decision rule in the benchmark case exhibits another interesting and useful property. Note that $x(\theta, s)$ is implicitly defined by

$$v_1(x(\theta, s), \theta + s) + u_1(x(\theta, s), \theta + s) - H(\theta)\, u_{12}(x(\theta, s), \theta + s) = 0,$$

which is the first-order condition of maximizing (6). Since H is weakly decreasing and $u_{12} > 0$, for $x = x(\theta, s)$,

$$v_1(x, \theta + s) + u_1(x, \theta + s) - H(\theta + z)\, u_{12}(x, \theta + s) \geq 0$$

if and only if $z \geq 0$. Observe that the left-hand side is weakly decreasing in x (since v_{11} and u_{11} are non-positive and u_{112} is non-negative). Hence, in order to make this expression zero, we must weakly increase x when $z \geq 0$, and similarly, we must weakly decrease x when $z \leq 0$. Therefore, $x(\theta, s) \leq x(\theta + z, s - z)$ if and only if $z \geq 0$. We state this observation in addition to the previous theorem.

Corollary 1 *In the optimal mechanism, for all θ, s, and all $z \in [-\theta, 1 - \theta]$,*

$$x(\theta, s) \leq x(\theta + z, s - z) \quad \textit{if and only if} \quad z \geq 0.$$

This means that in the optimal mechanism of the benchmark case, the Principal does not treat the Agent with the same actual type, but different initial type-estimate, the same way: the decision, x, will be set higher if the Agent's original type-estimate is higher. That is, given $\theta' + s' = \theta + s$, we have $x(\theta', s') \leq x(\theta, s)$ if and only if $\theta' \leq \theta$.

3.2 The Optimal Mechanism When the Principal Cannot Observe the Shock

We now return to the original model where the Principal cannot observe the shock but has the ability to disclose its realization (and, perhaps, the realizations of some other signals correlated with s) to the Agent. One difficulty with this model is that the Revelation Principle, in its standard form, does not apply. We will overcome this difficulty by considering a restricted class of mechanisms, and showing that even in this class, there exists a mechanism that guarantees the Principal the same utility as in the case when she could observe the shock. Since the Principal cannot achieve higher utility by using any mechanism than she did in the benchmark case, the mechanism that we will find is going to be optimal.

We will consider the class of sequentially incentive compatible two-stage direct mechanisms, defined as follows. In a two-stage direct mechanism, in the first stage, the Agent reports a type-estimate, θ, and the Principal gives him a transfer of $p(\theta)$. Then, the Principal allows the Agent observe the shock. In the second stage, the Agent reports back a value for the realization of the shock, s, and the Principal gives him a transfer of $q(\theta, s)$ and makes a decision of $x(\theta, s)$. (Equivalently, the Principal could specify the transfers in one payment, $t(\theta, s) = p(\theta) + q(\theta, s)$, paid at the end.) We will call a two-stage direct mechanism sequentially incentive compatible if it has a subgame perfect equilibrium where, in the first stage, the Agent reports his type-estimate truthfully, and then also reports the realization of the shock truthfully.[12]

Obviously, using only this type of mechanism, the Principal cannot attain higher utility than she did in the optimal mechanism when she could observe the shock. Our goal is to show that she can still attain the same utility and implement the same

[12] Two-stage mechanisms like these have been studied by Riordan and Sappington (1987). They derive the optimal incentive compatible contract between a regulator and a monopolist (chosen from several candidates according to the highest θ), where the monopolist has both current and future private information (i.e., a cost estimate and an eventual realization of marginal cost).

decision rule as in the benchmark case of the previous subsection. Hence, we restrict our attention to transfers and decision rules that are differentiable and such that x is weakly increasing in both of its arguments.

In the next lemma we derive the second-stage transfer (given the decision rule), which makes it optimal for the Agent to report the realization of the shock truthfully in the second stage once he had reported his type-estimate truthfully in the first stage. These are the incentive compatibility constraints expressing that it is optimal to tell the truth in the second round once the Agent had told the truth in the first round,

$$s = \arg\max_{s'} \{u(x(\theta, s'), \theta + s) + q(\theta + s')\} \quad \text{for all } \theta, s. \tag{7}$$

Lemma 2 *If $x(\theta, s)$ is weakly increasing in s, then (7) holds if and only if*

$$q(\theta, s) = \int_{-\infty}^{s} u_2\left(x\left(\theta, z\right), \theta + z\right) dz - u\left(x\left(\theta, s\right), \theta + s\right) + C_1\left(\theta\right), \tag{8}$$

where C_1 can be any function of θ.

Proof. Suppose first that (7) holds. The first order condition is

$$u_1(x\left(\theta, s\right), \theta + s)\, x_2(\theta, s) + q_2\left(\theta, s\right) = 0 \quad \text{for all } s. \tag{9}$$

Also observe that $u\left(x\left(\theta, s\right), \theta + s\right) + q\left(\theta + s\right)$ can be rewritten by the Fundamental Theorem of Calculus as

$$\int_{-\infty}^{s} \left[u_1(x\left(\theta, z\right), \theta + z) x_2(\theta, z) + u_2\left(x\left(\theta, z\right), \theta + z\right) + q_2\left(\theta, z\right)\right] dz + C_1\left(\theta\right),$$

where $C_1\left(\theta\right)$ equals $\lim_{z \to -\infty} \left\{u\left(x\left(\theta, z\right), \theta + z\right) + q\left(\theta + z\right)\right\}$. By plugging in the first order condition, we get

$$u\left(x\left(\theta,s\right),\theta+s\right)+q\left(\theta+s\right)=\int_{-\infty}^{s}u_{2}\left(x\left(\theta,z\right),\theta+z\right)dz+C_{1}\left(\theta\right),$$

which is exactly (8). Observe that $C_1(\theta)$ can be any constant, since $\lim_{z\to-\infty} q(\theta+z)$ can be chosen arbitrarily.

Suppose now that (8) holds. First, we show that there is no incentive to report $s' < s$. By (8)

$$\begin{aligned} u\left(x\left(\theta,s\right),\theta+s\right)+q\left(\theta+s\right)-u\left(x\left(\theta,s'\right),\theta+s'\right)-q\left(\theta+s'\right) \\ =\int_{s'}^{s}u_{2}\left(x\left(\theta,z\right),\theta+z\right)dz. \end{aligned} \tag{10}$$

Furthermore, by the Fundamental Theorem of Calculus,

$$\begin{aligned} u\left(x\left(\theta,s'\right),\theta+s\right)+q\left(\theta+s'\right)-u\left(x\left(\theta,s'\right),\theta+s'\right)-q\left(\theta+s'\right) \\ =\int_{s'}^{s}u_{2}\left(x\left(\theta,s'\right),\theta+z\right)dz. \end{aligned} \tag{11}$$

After subtracting (11) from (10) we get

$$\begin{aligned} u\left(x\left(\theta,s\right),\theta+s\right)+q\left(\theta+s\right)-u\left(x\left(\theta,s'\right),\theta+s\right)-q\left(\theta+s'\right) \\ =\int_{s'}^{s}\left[u_{2}\left(x\left(\theta,z\right),\theta+z\right)-u_{2}\left(x\left(\theta,s'\right),\theta+z\right)\right]dz. \end{aligned} \tag{12}$$

Note that the left-hand side is the negative of the Agent's gain from deviating to s'. Since $x(\theta,s)$ is weakly increasing in s and u possesses the single-crossing property, $u_{12} > 0$, we have

$$u_{2}\left(x\left(\theta,z\right),\theta+z\right)\geq u_{2}\left(x\left(\theta,s'\right),\theta+z\right)\quad\text{for all } z\geq s'.$$

Therefore, the integral in (12) is non-negative, and there is no incentive to deviate downward. Second, assume that $s' > s$. Then, by (8),

$$u\left(x\left(\theta,s\right),\theta+s\right)+q\left(\theta+s\right)-u\left(x\left(\theta,s'\right),\theta+s'\right)-q\left(\theta+s'\right)$$
$$=-\int_{s'}^{s}u_2\left(x\left(\theta,z\right),\theta+z\right)dz. \quad (13)$$

and by the Fundamental Theorem of Calculus,

$$u\left(x\left(\theta,s'\right),\theta+s\right)+q\left(\theta+s'\right)-u\left(x\left(\theta,s'\right),\theta+s'\right)-q\left(\theta+s'\right)$$
$$=-\int_{s}^{s'}u_2\left(x\left(\theta,s'\right),\theta+z\right)dz. \quad (14)$$

After subtracting (14) from (13) we get

$$u\left(x\left(\theta,s\right),\theta+s\right)+q\left(\theta+s\right)-u\left(x\left(\theta,s'\right),\theta+s\right)-q\left(\theta+s'\right)$$
$$=\int_{s}^{s'}\left[u_2\left(x\left(\theta,s'\right),\theta+z\right)-u_2\left(x\left(\theta,z\right),\theta+z\right)\right]dz.$$

The left-hand side is again the negative of the Agent's gain from deviating to s'. Since $x\left(\theta,s\right)$ is weakly increasing in s and the single-crossing property holds, this integral is non-negative. Therefore, there is no incentive to deviate upward either. ■

In order to analyze whether truth-telling is incentive compatible in the first stage, we need to find out what the Agent would report in the second round once he had lied in the first round. The following lemma establishes just this.

Lemma 3 *Assume that (7) holds. If the Agent with type-estimate θ reports θ' in the first stage and observes s, then, in the second stage, he will report $s' = s + \theta - \theta'$. Furthermore, the payoff of this Agent is the same as if he had type-estimate θ', observed shock s', and truthfully reported both.*

Proof. Observe that if $s' = s + \theta - \theta'$ then the Agent's second-stage payoff (his utility less the first-round transfer, $p(\theta')$) becomes

$$\begin{aligned} u\left(x\left(\theta', s'\right), \theta + s\right) + q\left(\theta' + s'\right) &= u\left(x\left(\theta', s'\right), \theta' + s'\right) + q\left(\theta' + s'\right) \\ &\geq u\left(x\left(\theta', s''\right), \theta' + s'\right) + q\left(\theta' + s''\right) \quad \text{for all } s''. \end{aligned}$$

The inequality follows from (7). Therefore, indeed s' should be reported. This Agent's utility is given by just the above expression plus $p(\theta')$, therefore, by the first equality, his overall utility is the same as if he had type-estimate θ', observed shock s', and truthfully reported both. ■

Now we turn to the first stage of the game. We want to describe the incentive compatibility constraints that guarantee that the Agent reports his type-estimate truthfully in the first round. By the previous lemma, if the Agent with type-estimate θ misreports θ' in the first round, then he will "correct" his lie and announce s' instead of s for the value of the shock, so that his ability is inferred correctly as $\theta' + s' = \theta + s$. His expected utility will then be

$$\mathbf{E}_s\left[u\left(x\left(\theta', s + \theta - \theta'\right), \theta + s\right) + p(\theta') + q\left(\theta', s + \theta - \theta'\right)\right]. \tag{15}$$

Therefore, the incentive compatibility constraints for the truthful revelation of θ at the beginning of the game are

$$\theta = \arg\max_{\theta'} \mathbf{E}_s\left[u\left(x\left(\theta', s + \theta - \theta'\right), \theta + s\right) + p(\theta') + q\left(\theta', s + \theta - \theta'\right)\right] \quad \text{for all } \theta. \tag{16}$$

In the following lemma, we characterize the first-stage transfer function (given the decision rule and a second-stage incentive compatible transfer scheme) such that the whole two-stage direct mechanism is sequentially incentive compatible.

Lemma 4 *Assume that (7) holds, and that for all θ, s, and all $z \in [-\theta, 1-\theta]$,*

$$x(\theta, s) \leq x(\theta + z, s - z) \quad \text{if and only if} \quad z \geq 0. \tag{17}$$

Then (16) is satisfied if and only if

$$p(\theta) = E_s \left[\int_0^\theta u_2\left(x\left(z, s\right), z + s\right) dz - u\left(x\left(\theta, s\right), \theta + s\right) - q\left(\theta, s\right) \right] + C_2, \tag{18}$$

where C_2 can be any constant.

Proof. Suppose that (16) holds. Then the corresponding first order condition is

$$E_s \left[u_1(x(\theta, s), \theta + s)(x_1(\theta, s) - x_2(\theta, s)) + q_1(\theta, s) - q_2(\theta, s) + p'\left(\theta\right)\right] = 0.$$

Since (7) is satisfied, we can use (9) and simplify it as

$$E_s \left[u_1(x(\theta, s), \theta + s)x_1(\theta, s) + q_1(\theta, s) + p'\left(\theta\right)\right] = 0.$$

From this it follows that

$$\frac{\partial}{\partial \theta} E_s \left[u\left(x\left(\theta, s\right), \theta + s\right) + p(\theta) + q\left(\theta, s\right)\right] = E_s \left[u_2\left(x\left(\theta, s\right), \theta + s\right)\right].$$

Using the Fundamental Theorem of Calculus we can conclude that

$$E_s \left[u\left(x\left(\theta, s\right), \theta + s\right) + p(\theta) + q\left(\theta, s\right)\right] = \int_0^\theta E_s \left[u_2\left(x\left(z, s\right), z + s\right)\right] dz + C_2,$$

where $C_2 = E_s \left[u\left(x\left(0, s\right), s\right) + p(0) + q\left(0, s\right)\right]$. This is equivalent to (18). Since $p\left(0\right)$ can be chosen arbitrarily, C_2 can be any constant.

Suppose now that (18) holds. We will show that the Agent with type-estimate θ has no incentive to deviate to θ' in the first round. Let $U(\theta)$ denote the expected payoff of the Agent with type-estimate θ. From (18), by rearranging,

$$U(\theta) = E_s\left[\int_0^\theta u_2(x(z,s), z+s)\, dz\right] + C_2. \tag{19}$$

By Lemma 3, the utility of the Agent with type-estimate θ reporting θ' is (15), which can be written using (8) as

$$E_s\left[\int_{-\infty}^{s+\theta-\theta'} u_2(x(\theta',z), \theta'+z)\, dz\right] + p(\theta') + C_1(\theta'). \tag{20}$$

For $\theta' < \theta$, rewrite (20) as

$$E_s\left[\int_{-\infty}^{s} u_2(x(\theta',z), \theta'+z)\, dz\right] + p(\theta') + C_1(\theta') + E_s\left[\int_{s}^{s+\theta-\theta'} u_2(x(\theta',z), \theta'+z)\, dz\right]$$
$$= U(\theta') + E_s\left[\int_{s}^{s+\theta-\theta'} u_2(x(\theta',z), \theta'+z)\, dz\right].$$

Hence, type-estimate θ has no incentive to deviate to θ', $\theta' < \theta$, if

$$U(\theta) - U(\theta') \geq E_s\left[\int_{s}^{s+\theta-\theta'} u_2(x(\theta',z), \theta'+z)\, dz\right],$$

which, using (19), is equivalent to

$$E_s\left[\int_{\theta'}^{\theta} u_2(x(z,s), z+s)\, dz\right] \geq E_s\left[\int_{s}^{s+\theta-\theta'} u_2(x(\theta',z), \theta'+z)\, dz\right].$$

After rewriting the right-hand side

$$E_s\left[\int_{\theta'}^{\theta} u_2(x(z,s), z+s)\, dz\right] \geq E_s\left[\int_{\theta'}^{\theta} u_2(x(\theta', s+z-\theta'), z+s)\, dz\right].$$

But this is indeed satisfied, since for all $z \geq \theta'$, $x(z,s) \geq x(\theta', s+z-\theta')$ by (17) and the single-crossing property.

For $\theta' > \theta$, rewrite (20) as

$$E_s\left[\int_{-\infty}^{s} u_2\left(x\left(\theta', z\right), \theta' + z\right) dz\right] + p\left(\theta'\right) + C_1(\theta') - E_s\left[\int_{s+\theta-\theta'}^{s} u_2\left(x\left(\theta', z\right), \theta' + z\right) dz\right]$$
$$= U\left(\theta'\right) - E_s\left[\int_{s+\theta-\theta'}^{s} u_2\left(x\left(\theta', z\right), \theta' + z\right) dz\right].$$

Hence, type-estimate θ has no incentive to deviate to θ', $\theta' > \theta$, if

$$U\left(\theta\right) - U\left(\theta'\right) \geq -E_s\left[\int_{s+\theta-\theta'}^{s} u_2\left(x\left(\theta', z\right), \theta' + z\right) dz\right],$$

which, using (19) and multiplying both sides by -1, is equivalent to

$$E_s\left[\int_{\theta}^{\theta'} u_2\left(x\left(z, s\right), z + s\right) dz\right] \leq E_s\left[\int_{s+\theta-\theta'}^{s} u_2\left(x\left(\theta', z\right), \theta' + z\right) dz\right].$$

After rewriting the right-hand side

$$E_s\left[\int_{\theta}^{\theta'} u_2\left(x\left(z, s\right), z + s\right) dz\right] \leq E_s\left[\int_{\theta}^{\theta'} u_2\left(x\left(\theta', s + z - \theta'\right), z + s\right) dz\right].$$

But this is indeed satisfied, since for all $z \leq \theta'$, $x\left(z, s\right) \leq x\left(\theta', s + z - \theta'\right)$ by (17) and the single-crossing property. ■

Finally, we are ready to state the main result of the paper. Using an incentive compatible two-stage direct mechanism, the Principal can attain the same outcome as in the benchmark case.

Theorem 2 *Assume that the Principal cannot observe the shock. In the optimal mechanism, the decision rule and the utility of the Principal are the same as in the optimal mechanism when she could observe the shocks (as in Theorem 1). The transfers are defined by (8) and (18) with* $C_1 \equiv 0$ *and* $C_2 = 0$.

Proof. The decision rule $x(\theta, s)$ that maximizes (6) clearly satisfies the assumptions of the previous lemmas: it is differentiable and weakly increasing in both arguments (Theorem 1), and for all θ, s and all $z \in [-\theta, 1-\theta]$, it satisfies (17) (Corollary 1). Use this decision rule to define the transfer functions, q and p, by (8) and (18), with $C_1 \equiv 0$ and $C_2 = 0$,

$$\begin{aligned} q(\theta, s) &= \int_{-\infty}^{s} u_2(x(\theta, z), \theta + z)dz, \\ p(\theta) &= \mathbf{E}_s \left[\int_0^{\theta} u_2(x(z, s), z + s)dz - \int_{-\infty}^{s} u_2(x(\theta, z), \theta + z)dz \right]. \end{aligned}$$

By Lemmas 2–4, this two-stage mechanism is sequentially incentive compatible.

From (18), the Agent's expected payoff after observing θ is

$$U(\theta) = E_s \left[\int_0^{\theta} u_2 \left(x \left(z, s \right), z + s \right) dz \right].$$

This payoff is identical to the one in the optimal mechanism of the previous subsection, (2) with $U(0) = 0$. Furthermore, note that the Agent prefers to participate in the mechanism. Since the decision rule is the same, the total surplus is also the same, therefore the Principal's payoff must also be the same as in the benchmark case. ■

The Agent faces the following trade-off in the first stage of the optimal contract. He can choose between contracts that come with a larger up-front transfer (from the Principal to the Agent, the sign of which may differ across applications) but reward higher ex-post types less in the second period, and contracts that carry a less attractive up-front transfer, but will be more generous for higher ex-post types in the second period. The Principal, by exposing the Agent to this type of trade-off, can effectively discriminate between Agents with higher and lower type-estimates. An Agent with a low type-estimate will choose a contract with a higher up-front transfer, because he thinks it is unlikely that he will have a high ex-post type in the second round. In contrast, an Agent with a high type-estimate stays away from this type of contract and instead chooses one that is less attractive in the first stage, but, for high ex-post types, is more rewarding in the second stage.

4 Conclusions

In this paper, we analyzed a Principal-Agent relationship with adverse selection and transferable utilities, where the Principal can influence the precision of the Agent's private information. In particular, we assumed that the Agent only had an initial estimate of his productivity parameter, however, the Principal could disclose to him, without observing, the difference between the his actual productivity and his initial estimate. We characterized the mechanism maximizing the Principal's utility. Interestingly, in this contract, the Principal always and fully reveals the information that she controls (but cannot directly observe); moreover, her utility and the implemented decision will be the same as if she could observe the signals released by her.

The optimal contract consists of two stages. In the first stage, the Agent, who only has an estimate of his type, is asked to pick a transfer and a corresponding second-period contract (consisting of an allocation rule and an additional transfer depending on his reported ex-post type) from a menu offered by the Principal. In the second stage, the Agent is asked to report his actual type. What is interesting in the optimal two-stage contract is that the first-period menu is chosen by the Principal so that the Agent will truthfully reveal his type-estimate and commit to reveal the additional signal (released but not observed by the Principal), and yet the Agent will only enjoy information rents for his original private information, the type-estimate. The Principal who controls the flow of information appropriates its rents.

References

[1] Bergemann, D., and M. Pesendorfer (2002), "Information Structures in Optimal Auctions," Cowles Foundation Discussion Paper No. 1323.

[2] Caillaud, B., R. Guesnerie, and P. Rey (1992): "Noisy Observation in Adverse Selection Models," *Review of Economic Studies*, 59:595-615.

[3] Compte, O., and P. Jehiel (2001): "Auctions and Information Acquisition: Sealed-bid or Dynamic Formats?," mimeo, C.E.R.A.S.

[4] Eső P. and B. Szentes (2002): "Optimal Information Disclosure in Auctions: The Handicap Auction," Northwestern University, CMS-EMS Discussion Paper #1361.

[5] Fudenberg, D., and J. Tirole (1991): *Game Theory*, MIT Press, Cambridge, USA

[6] Lewis, T., and D. Sappington (1994): "Supplying Information to Facilitate Price Discrimination," *International Economic Review*, 35:309-327.

[7] Milgrom, P., and R. Weber (1982): "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50:1089-1122.

[8] Ottaviani, M., and A. Prat (2001): "The Value of Public Information in Monopoly," *Econometrica*, 69:1673-1683.

[9] Persico, N. (2000): "Information Acquisition in Auctions," *Econometrica*, 68:135-148.

[10] Riordan, M. H., and D. E. M. Sappington (1987): "Awarding Monopoly Franchises," *American Economic Review*, 77:375-387.

[11] Royden, H. L. (1988): *Real Analysis,* New York, Macmillen Publishing Company.