

Endogenous Punishments in Agency with Verifiable Ex Post Information*

Anke S. Kessler Christoph Lülfesmann[†] Patrick W. Schmitz
Simon Fraser U. *Simon Fraser U.* *U. of Bonn*

This version: October 2003

Abstract. The paper studies an adverse selection model in which a contractible, but imperfect signal on the agent's type becomes available ex post. The agent is protected by limited liability which we model as an upper bound on total payments made to the principal. The maximum punishment level is then endogenous, and the following conclusions emerge. First, the qualitative effects of the signal can be unambiguously tied to the nature of the underlying problem (e.g., whether the agent is in a 'buyer' or a 'seller' position). Second, the joint surplus of the relationship may well fall, i.e., the distortions due to informational asymmetries can become more pronounced although more information is now available. Finally, the agent may benefit from releasing additional signals because of an associated increase in expected informational rents.

JEL Classification: D82

Keywords: Adverse Selection; Ex Post Information; Limited Liability

*We thank Fahad Khalil, Urs Schweizer, two anonymous referees, Nicola Persico (the Editor) and various participants at seminars in Zürich, Karlsruhe, and at the 2001 Verein für Socialpolitik meeting in Magdeburg for helpful comments and discussions. Remaining errors are our own.

[†]Corresponding Author. Address: University of Bonn, Department of Economics, Adenauer-allee 24-42, 53113 Bonn, Germany. E-mail: akessler@sfu.ca, clulfes@sfu.ca, schmitz@wipol.uni-bonn.de.

1 Introduction

Informational asymmetries in the form of adverse selection are prevalent in many economic transactions. At the time two contracting parties meet, one party possesses superior information about a parameter that is relevant to the transaction and, by extension, is of interest for the other party. To give just few examples, an employee may know more about his abilities than her employer, the creditworthiness of a debtor may be unknown to the bank, a customer may be better in assessing his risks than the insurance company, or a firm may sell products whose quality is unknown to the consumer. These and many more instances are formally analyzed in a well-known adverse selection framework in which a principal contracts with a privately informed agent. In this workhorse model, all agents other than the ‘worst’ type can command a rent. Also, because of this fact, the contracted decisions for all agents other than the ‘best’ type are distorted downwards, away from their first-best levels.

The present paper adds to a by now substantial body of literature which studies a natural extension of this standard principal-agent model (a literature survey is given below). Namely, we analyze a situation where information that is correlated with the agent’s type becomes available after a contractually specified action has been chosen. Note that additional ex-post information seems to be the rule rather than the exception in many empirically relevant circumstances. One set of prominent examples are those where the principal actively monitors the agent. In employment or procurement relationships, ex post audits reveal information on the agent’s true productivity or production costs. Tax authorities will conduct audits to verify taxpayers’ reports. Also, franchise contracts are usually contingent not only on sales data, but also on the franchisee’s behavior (cleanliness, product quality, friendly service) which is regularly monitored by the franchisor.¹ But further information is frequently available even without the principal taking any explicit action. If the agent sells a durable good to the principal, product failure or malfunctioning after

¹Baron and Besanko (1984) analyze the decision of government agencies to audit tax payers. For an assessment of franchise contracts, see Lafontaine and Slade (1996). In an empirical study of automobile franchising, Arruñada, Garicano, and Vázquez (2001) report that manufacturers condition their contracts on monitoring evidence related to the dealer’s sales, customer satisfaction, machinery, personnel, and financial performance, obtained by direct inspections and auditing.

the purchase will reveal additional information about the quality of the product. In credit (respectively, insurance) market contexts, changes in independent credit ratings (respectively, an accident) will constitute a verifiable ex post signal about a customer's true risk type on which his interest payment (respectively insurance premium) can be based. Or consider an upstream firm that contracts as a principal with a downstream firm as an agent over the supply of an intermediate good. The agent's willingness to pay for the input will then be related to the price which is later charged on the downstream market.² Finally, ex post information is common in situations where the decisions of other parties are based on information correlated with the agent's type as will be the case in the context of, e.g., auctions, team production, or product market competition.

There is one central finding in the previous literature on ex-post information: the availability of additional signals leads the principal to reduce the distortions which are imposed on less able types of agents, and thus serves to raise allocative efficiency. Even the first best is frequently attained as long as unbounded punishments can be imposed on an agent if the signal indicates he did not tell the truth. In line with many other papers, we analyze ex-post signals in a more realistic setting where the agent is protected by limited liability, which we interpret as a wealth constraint that limits total payments to the principal. The maximum feasible penalty then depends on the specific terms of contract and is therefore endogenous.

Perhaps surprisingly, our results are in striking contrast to those of the previous literature. We show that the total expected surplus of the relationship may *fall* as an informative signal becomes publicly available. As a second central result, we also demonstrate that the direction of the additional distortions that arise crucially depends on the nature of the underlying problem. In particular, if the agent's utility is increasing in the contracted action (e.g., the quantity of a good or service he buys from the principal), existing downward distortions will be amplified (the volume of trade further decreases) for signals that are not too informative. Conversely, if the agent's utility is decreasing in the contracted action (e.g., the amount of output he is to produce), the distortions may go in the opposite direction, i.e., the agent's

²A related point is made by Riley (1988) who argues that the seller of an oilfield will want to condition the buyer's payment on the quantity of oil extracted, which is readily observable and a noisy signal of oilfield profitability.

actions may exceed their first-best level. In this latter scenario, the agent might also command an even *larger* informational rent relative to a situation where no ex-post signal is available, another striking contrast to results of the existing literature. It implies that agents may have an incentive to ex-post reveal or generate signals about his privately held information to the principal.

Note that this dichotomy is not found in the standard model without ex post information,³ and it also cannot arise in settings with ex-post signals and exogenously given punishments. The distinction between ‘buyers’ and ‘sellers’ we identify arises for the following reason. By the maximum punishment principle, a penalty is optimally imposed and set to its maximal level whenever the signal ex post contradicts what the agent has claimed to be true. Consider first an agent in a ‘buyer’ position who is privately informed about his willingness to pay and has only a limited wealth. Now, the smaller the quantity that a low-valuation agent purchases from the principal, the lower the price he has to pay, and the more money is left in the pocket of an untruthful high-valuation agent that can serve as a penalty if the signal indicates non-compliance. Therefore, if the principal wants to increase potential penalties, she decreases the quantity traded beyond the level that would have been optimal in a situation where no ex post information is available.

The converse argument holds if the agent is in a ‘seller’ position: the higher the amount of output a high-cost agent is to produce, the more he receives in compensation and the more severely can an untruthful low-cost agent be punished if the signal indicates that he has been untruthful. Therefore, the principal optimally increases production and, as we will see, this effect may be sufficiently strong to yield production above the first-best level, so that total surplus may again fall.⁴ The

³In the standard model, both scenarios lead to less than efficient quantities traded for all types but the ‘top’ agent: what is important is that the single crossing property holds; i.e., absolute and marginal utility move in the same direction as we replace better types (high valuations or low costs) by worse types. This implies that a good-type agent suffers more (or benefits less) from a reduction of the quantity than a bad type, so that distorting the quantity downwards makes it less attractive to mimic worse types. Obviously, this logic does not depend on whether the agent is a ‘buyer’ or a ‘seller’.

⁴Importantly, this occurs even though we ignore any commitment problems. It is well known that additional information can be harmful if commitment not to use the information is ruled out [Riordan (1990), Dewatripont and Maskin (1995), or Crémer (1995)]. Upward distortions are also known from the literature on countervailing incentives where the agent’s reservation utility is type dependent. See Lewis and Sappington (1989) and Jullien (2000) for further references.

agent may then even enjoy a higher rent when ex post information about his type is available ex post. As we will see, this is because a low-cost agent is affected less than a high-cost agent if the quantity produced rises, i.e., it becomes more attractive for a low-cost type to mimic a high-cost type, *ceteris paribus*. Finally, note that the extent to which the principal actually wants to increase potential penalties depends on the informativeness of the ex post signal. If the signal is very precise, small penalties are sufficient and therefore the distortion will be weakened (i.e., additional information is welfare enhancing) in both cases.

The remainder of the paper is organized as follows. Section 2 surveys the existing body of literature on ex-post signals in adverse selection frameworks. Section 3 introduces the model. The optimal contract in the presence of a verifiable ex post signal is derived and discussed in Section 3. A final section concludes. All proofs are relegated to the Appendix.

2 Literature Review

Agency models with asymmetric information start out from the presumption that the principal possesses knowledge of the prior distribution according to which the agent's unknown type is distributed. The optimal contract then entails an output distortion for any than the most efficient type (Baron and Myerson, 1983, Laffont and Tirole, 1986). The literature on ex-post signals showed that this outcome can be significantly improved if the parties can make their contractual arrangement contingent on an observable signal that is realized after the agent reported his type. Under these circumstances, Nalebuff and Scharfstein (1987) show that the first best is attained if the agent's private information is perfectly revealed ex post with some positive probability. Using a sufficiently severe punishment whenever he did not report his type truthfully, the fear of being caught then prevents each agent from mimicking a less efficient type, and the first-best allocation can be implemented. Subsequently, it was shown that if agents are risk neutral as is commonly assumed in adverse selection settings, nor the accuracy nor the frequency of the signal matter. Any whatever noisy or infrequent signal is sufficient to eliminate the agent's rent and to ensure the first-best outcome, provided the signal can be contractually

employed and the agent can be subjected to unboundedly high penalties [Riordan and Sappington, 1988), Crémer and McLean (1988)].⁵ This is because even if the signal is very inaccurate or it is rarely available, it can help to slacken the incentive-compatibility constraints of higher-quality agents as long as the principal has access to unlimited penalties.⁶

Since unbounded penalties seem rather unrealistic in most situations, the subsequent literature addressed the question whether the favorable effect of additional information is preserved when the agent cannot be held unboundedly liable. Two basic forms of limited liability can be distinguished (see Laffont and Martimort, 2002). First, one might interpret limited liability as a bound on the maximum punishment which can be imposed on an agent whenever the signal indicated that he did not announce his true type. In this interpretation, the punishment is exogenous as it does not vary with the pecuniary provisions of the contract that is signed by principal and agent. Second, one can interpret limited liability in a way that the agent's total payments to the principal cannot exceed a certain amount.⁷ The maximum punishment is endogenous in this latter case as it depends on the contractual provisions, in particular, on the contractual payments which are made to or from the agent as a compensation for the transaction between principal and agent.⁸

A part of the existing literature has considered the case of exogenous punishments. Baron and Besanko (1984) derived the so-called maximum punishment principle according to which an agent that is caught lying should be made subject to the largest possible punishment. For the case of two types of agents, Kofman and Lawarree (1993) are able to derive stronger results. They fully characterize the optimal con-

⁵General conditions on the distribution of signals for which this result holds are derived by Riordan and Sappington (1988) for the standard agency model and by Crémer and McLean (1988) for auctions. See also McAfee and Reny (1992) for an extension beyond the auction setting.

⁶Note that in contrast to the case of perfect signals, penalties are realized in equilibrium as low-quality agents are also subjected to these penalties with positive probabilities in equilibrium. This explains why an inaccurate signal does not facilitate an efficient outcome when the agent is risk averse.

⁷The same representation of a limited liability constraint can also be found in Sappington (1983), Border and Sobel (1987) and Melumad and Mookherjee (1989), among others.

⁸An alternative way to model endogenous punishments would be to require that the agent's total utility - rather than his total payments - cannot fall below a certain boundary level. The problem about this approach is that an agent's utility depends on his true type which is unobservable. Consequently, the maximum feasible penalty need not be well defined off equilibrium.

tractual solution and show that the optimal contract depends on the accuracy of the signal: if the signal is imprecise, it leaves the optimal action of low-type agents unaffected while reducing the informational rent of the better type.⁹ Once one arrives at a precision level where the information rent is completely eliminated, a better signal precision now leads the principal to reduce the allocative distortion, that is, the equilibrium action of the low-type agent is raised relative to the second best. Finally, when the allocation becomes first best, the principal starts to reduce the frequency of the signal while leaving the first-best allocative outcome. The Kofman and Lawarree results show that for exogenous punishments, informative signals - at least weakly - raise the level of allocative efficiency, and at the same time lower the agent's information rents.

Ex-post signals in settings with endogenous punishments have been analyzed in Laffont and Tirole (1992; 1993, ch. 12), and in Demougin and Garvie (1991).¹⁰ In the model analyzed by Laffont and Tirole, the signal is perfectly revealing but observed only with a certain probability.¹¹ They argue that monitoring again reduces the agent's rent and mitigates the downward distortion. However, we show that the efficiency-improving effect arises only because Laffont and Tirole impose a certain ad-hoc constraint on the level of signal precision. If one removes this artificial constraint, our results indicate that additional ex post information can again render it optimal to raise output above the first-best level, so that total surplus is reduced. Finally, Demougin and Garvie (1991) analyze a framework which is closely related to our model. In a setting with a continuous type space, this paper concludes that a more informative signal always reduces informational rents, mitigates the

⁹This is because - with an exogenous punishment level - a more precise signal in this interval reduces the absolute informational rent, but not the marginal informational rent of good-type agents. Accordingly, the second-best distortion of the standard contract remains optimal.

¹⁰See also Gary-Bobo and Spiegel (2003) for a model where where additional signals represent real cost shocks, that is, where a signal has a direct allocative effect. Khalil (1997) analyzes a setting where the principal cannot commit to costly auditing. He finds that whenever auditing occurs with positive probability, the agent receives no rent, and there is an upward distortion which increases the probability that the agent complies with the contract. The driving force behind this result is that the principal must be given sufficient incentives to audit. While the mechanism at work is thus different, it also requires transfer-dependent penalties.

¹¹They also assume that the agent cannot be punished if no signal is observed. From what we said before, it should be clear that situations with accurate but infrequent signal are qualitatively identical to those with an inaccurate signal which is always available.

distortions, and is thus welfare improving. Although contradictory at first sight, their result is consistent with the findings laid out above as we will see below. In particular, their model may be seen as a special case of ours (but with a continuous type space) with specifications of the agent’s utility function ensuring that neither of the counterintuitive results we identify can occur.¹²

3 The Model

Consider the following principal-agent model with adverse selection. The utility functions of the principal (P) and the agent (A) are, respectively, $u_P = v(x, \theta) - t$ and $u_A = u(x, \theta) + t$, where $x \in [\underline{x}, \bar{x}] \subset \mathbb{R}$ is a contractible action, θ is a random parameter, and $t \in \mathbb{R}$ is a (possibly negative) transfer payment from P to A. The parameter θ is private information to the agent and can take only two values, $\theta \in \{\theta_h, \theta_l\}$.¹³ The ex ante probability that $\theta = \theta_h$ is common knowledge and denoted by $q \in (0, 1)$.

To fix ideas, it will be helpful in what follows to keep in mind two applications of the above model that have been extensively studied in the literature. The first example is a vertical relationship between an upstream firm (a manufacturer) and a downstream firm (a retailer), where the principal is the monopolistic supplier who sells a quantity x of some good or service to the agent at cost $-v(x, \theta) \equiv c(x) \geq 0$. The retailer’s gross profit is $u(x, \theta) \geq 0$. It depends on his intrinsic preference parameter θ (e.g., downstream market characteristics) and is strictly increasing in the quantity bought; i.e., $u'(x, \theta) > 0$, primes denoting partial derivatives with respect to x . The second example is a situation where the principal is a buyer who orders some product or service of non-contractible quality. The quality which is delivered depends on the seller’s intrinsic type, which may reflect his technological know how

¹²Demougin and Garvie (1991) assume that the agent’s utility function is linear in his output. For this special case and under an additional assumption on the elasticity of the (binary) signal with respect to agent type type, they show that the local first-order conditions for truth-telling imply global incentive compatibility. However, the validity of the first-order approach does extend to non-linear utility functions, which would render an analysis of continuous type spaces extremely difficult.

¹³The subscript h indicates a ‘high type’ and the subscript l a ‘low type’. Notice that the label ‘high-type’ (‘low-type’) refers to an agent with a *high* (low) valuation for a good or service if he is a buyer, and to an agent with *low* (high) production costs if he is a seller.

or his craftsmanship, and the high-quality agent may incur higher production costs (he uses more expensive inputs). The buyer's utility from ordering x units is then $v(x, \theta) \geq 0$, and the supplier's production cost is $-u(x, \theta) = c(x, \theta) \geq 0$. Total costs strictly increase in the quantity supplied, i.e., $u'(x, \theta) < 0$, and in the quality parameter θ .

Although the principal does not observe θ directly, she has access to a verifiable signal $s \in \{s_h, s_l\}$ which is realized after the action x has been taken, and which is imperfectly correlated with θ . In the manufacturer-retail example, s could be the final price that can be charged by the retailer as an intermediate on the downstream market.¹⁴ In the trade-quality example, the signal s could be a malfunctioning of the product, or some other verifiable indicator that the trade/service was not of acceptable quality. In both cases, the quality of information is related to the agent's cost parameter that the principal can infer from the characteristics of the finished product or obtain through conducting audits. Let π_{ij} be the probability that the signal $s = s_i$ is realized, conditional upon a parameter value $\theta = \theta_j$, $i, j \in \{h, l\}$ and assume $\pi_{ll} = \pi_{hh} \equiv \pi$ for simplicity. Hence, the signal is 'correct' with probability π and 'incorrect' with probability $1 - \pi$. In what follows, we use π as a measure of the informativeness of the signal s and without loss of generality let $\pi > \frac{1}{2}$.

Assumption 1. The functions $v(\cdot)$ and $u(\cdot)$ are twice continuously differentiable, monotone, and concave in x . Total surplus $S(x, \theta_i) \equiv v(x, \theta_i) + u(x, \theta_i)$ is strictly concave in x . Furthermore,

- a) $u(x, \theta_h) > u(x, \theta_l) \forall x \in (\underline{x}, \bar{x}]$,
- b) $u'(x, \theta_h) > u'(x, \theta_l) \forall x \in [\underline{x}, \bar{x}]$,
- c) $x_i^{FB} = \arg \max S(x, \theta_i)$ satisfies $x_l^{FB} \leq x_h^{FB}$ and $x_i^{FB} \in (\underline{x}, \bar{x})$, $i \in \{l, h\}$,

¹⁴Royalties to be paid by coal mining companies for operating on publicly owned land will generally vary with the price of coal from the same general area, its quality, and other factors the commissioner in charge deems 'relevant' [see, e.g., the Alaska Administrative Code]. More generally, contracts between a manufacturer and its exclusive retailers are often contingent on the retailers' behavior that supplements the information contained in the sales data [Lafontaine and Slade (1996)]. In franchising relationships, for instance, the franchisor usually takes the right to terminate the contract at will if the franchisee is not maintaining quality standards (which is subject to random monitoring). Since franchise contracts typically require franchisees to make highly specific investments in equipment, these investments are (at least partially) lost when the franchisor receives a bad signal [Dnes (1996)].

- d) $\hat{x}_l(\pi) = \arg \max S(x, \theta_l) - \frac{q}{1-q} \left[u(x, \theta_h) - \frac{1-\pi}{\pi} u(x, \theta_l) \right]$ is unique and satisfies $\hat{x}_l(\pi) \in (\underline{x}, \bar{x})$ for all values $\pi \in [\frac{1}{2}, 1)$.

Assumptions 1 a) and b) state that the agent's possible types θ have a natural ordering both in absolute and marginal utilities (single crossing property). Parts c) and d) are made for convenience and allow us to focus on first-order conditions.¹⁵

The timing is as follows. Before contracting takes place, nature chooses θ and the agent learns his type. Then, the principal proposes a contract to the agent which the latter can accept or reject. Focusing on the most interesting case, we assume throughout that the principal wants to contract with both types. If the agent accepts, the action x is taken in accordance with the contract. Next, the signal s is realized and the contractually specified transfer is paid. If the agent rejects, the parties obtain their reservation utilities. The sequence of events is summarized in Figure 1.

Assuming the principal can commit herself not to renegotiate the contract, we can invoke the Revelation Principle and confine attention to contracts $\{x(\hat{\theta}), t(\hat{\theta}, s)\}$ that specify an action x and a transfer t as a function of the agent's report $\hat{\theta}$ and the realization of the signal s . As future points of reference, we first characterize the cases where information is symmetric and where the signal is not available, respectively.

First Best: If θ is publicly observable, the principal maximizes her utility, $u_P = v(x, \theta_i) - t$, subject to the agent's participation constraint $u_A = u(x, \theta_i) + t \geq \bar{u}$, where \bar{u} denotes his reservation utility. For a type- θ_i agent, the first-best action is uniquely characterized by $S'(x_i^{FB}, \theta_i) = 0$ and the transfer payment is $t_i^{FB} = \bar{u} - u(x_i^{FB}, \theta_i)$.

Second Best: Next, suppose θ is known only to the agent but the signal is not available or, equivalently, completely uninformative ($\pi = \frac{1}{2}$). Let (x_i, t_i) denote the contract designed for an agent who claims to be of type θ_i . The individual rationality constraints are

$$u(x_i, \theta_i) + t_i \geq \bar{u} \quad i \in \{l, h\}. \quad (1)$$

¹⁵Note that $x_l^{FB} \leq x_h^{FB}$ is already implied by the single crossing property if the principal's utility $v(\cdot)$ does not depend on θ (Otherwise, $v'(x, \theta)$ increasing in θ is a sufficient condition). Also observe that for any given $q \in (0, 1)$ the objective function in Assumption 1d) is concave if $v(\cdot)$ is sufficiently concave. Moreover, it is sufficient for our analysis that Assumption 1d) is satisfied for $\pi \leq \pi^{FB}$, which is defined below.

In addition, truthful revelation requires

$$u(x_i, \theta_i) + t_i \geq u(x_j, \theta_i) + t_j \quad i, j \in \{l, h\}, i \neq j. \quad (2)$$

The principal maximizes her expected utility $q[v(x_h, \theta_h) - t_h] + (1 - q)[v(x_l, \theta_l) - t_l]$, subject to (1) and (2). It is easy to see that the incentive constraint for the low-type agent as well as the individual rationality constraint for the high-type agent are slack at the optimum and can be ignored. From the first-order conditions, the second best actions are determined by

$$\begin{aligned} S'(x_h^{SB}, \theta_h) = 0 &\Rightarrow x_h^{SB} = x_h^{FB}, \quad \text{and} \\ S'(x_l^{SB}, \theta_l) = \frac{q}{1 - q} \phi'(x_l^{SB}) &\Rightarrow x_l^{SB} < x_l^{FB}, \end{aligned}$$

where $\phi(x) \equiv u(x, \theta_h) - u(x, \theta_l) > 0$ with $\phi' > 0$. The corresponding transfers are $t_l^{SB} = \bar{u} - u(x_l^{SB}, \theta_l)$ and $t_h^{SB} = \bar{u} - u(x_h^{FB}, \theta_h) + \phi(x_l^{SB})$. Hence, under the no-signal second best contract, the action of the high-type agent is first-best and he earns an informational rent equal to $\phi(x_l^{SB})$. In contrast, the action of a low-type agent is distorted downward and he obtains his reservation utility. Evidently, this result does not depend on whether the agent is a buyer or a seller, i.e., if his payoff is increasing or decreasing in x .¹⁶

4 Optimal contracting with a Verifiable Signal

We now turn to the case where the principal can condition the transfers specified in the contract on the verifiable (and informative) signal s . Clearly, the principal may now want to lower the transfer when the realization of s contains evidence that contradicts the agent's claim $\hat{\theta}$. To analyze this situation, let us first assume that the agent disposes of unlimited wealth so that the size of possible penalties is unbounded. In this situation, we can make the following line of arguments which goes back to Nalebuff and Scharfstein (1987).¹⁷ Consider the incentive constraint of

¹⁶Note, though, that the conclusion is a matter of convention given the properties of the underlying problem as summarized in Assumption 1. For instance, we could use the transformation $y := \bar{x} - x$ which reversed the single crossing property. The optimal contract would then specify $y_l^{SB} > y_l^{FB}$ and our results in the remainder of the paper would change accordingly.

¹⁷In their original work, the signal is perfect but it is attained with less than full probability.

the high-type agent,

$$u(x_h, \theta_h) + t_h \geq u(x_l, \theta_h) + (1 - \pi)t_l(s_l) + \pi t_l(s_h). \quad (3)$$

In addition, consider the IR constraint of the bad-type agent,

$$u(x_l, \theta_l) + \pi t_l(s_l) + (1 - \pi)t_l(s_h) \geq \bar{u}, \quad (4)$$

and suppose first that the signal is completely uninformative, $\pi = 1/2$. Then, a binding (IR-1) requires that $dt_l(s_l)/dt_l(s_h) = -1$, i.e., any reduction in $t_l(s_h)$ must be accompanied by an identical raise in $t_l(s_l)$. Using this property in (3) reveals that the availability of the signal leaves $(IC - h)$ unaffected, implying there is no gain in using the signal in a contractual agreement. Next, suppose that $\pi > 1/2$. If $t_l(s_h)$ is lowered by one unit, $t_l(s_l)$ needs to increase by only less than one unit in order to keep the bad agent's individual-rationality constraint satisfied. Note that any decrease in $t_l(s_h)$ which is accompanied by some smaller increase in $t_l(s_l)$ relaxes the high-type agent's incentive constraint because the right-hand side of $(IC - h)$ becomes smaller. As a consequence, there must exist some transfer $t_l(s_h)$ and a corresponding payment $t_l(s_l)$ which satisfies the low-type agent's IR constraint, and at the same time makes the good agent's IC constraint slack for any choice of x_l ; in other words, it becomes possible to implement the efficient allocation x_l^* without any need to grant a costly informational rent to the high-type agent. Penalizing an agent who claimed to be of low type for contradictory evidence is sufficient to implement the first best, even if this information is very inaccurate. However, as the effective penalty is $(\pi - 1/2)t_l(s_h)$ converges to infinity as π converges to $1/2$, the 'penalty' payment increases without bounds which is feasible only if the agent's endowment is unlimited.

In the remainder, we rule out unboundedly high penalties:

Assumption 2. The agent is protected by limited liability,

$$t_i(s_j) \geq -W \quad \forall i, j \in \{l, h\}. \quad (\text{LL}_{ij})$$

Furthermore,

$$\bar{u} + W \geq \max \{u(x_h^{FB}, \theta_h), u(x_l^{SB}, \theta_l), u(x_l^{FB}, \theta_h)\}. \quad (5)$$

The first part of Assumption 2 imposes an exogenous lower bound on the feasible transfers between principal and agent, which we assume to be commonly known.¹⁸ We can interpret (LL_{ij}) as a wealth constraint where $W \geq 0$ denotes the initial wealth of the agent. Since this constraint restricts the total payments that an agent can make to the principal, the maximum feasible punishment depends on the size of transaction payments specified in a mechanism and is thus endogenous. Note that similar limits on transfers would arise from legal restrictions such as minimum wage laws (in which case $-W \geq 0$ is the minimum wage) or if the agent's preferences exhibit infinite risk aversion below a certain transfer level.

The second part of Assumption 2 ensures that the limited liability constraints do not bind at the first and second-best benchmark solutions determined previously, so that those optimal contracts are unaffected by (LL_{ij}) .¹⁹ The assumption serves to establish two benchmark outcomes which will be used for comparison with our subsequent results. Notice that if the agent's reservation utility \bar{u} is equal to zero, (5) is trivially satisfied if the agent is in a seller-position: even if the agent possesses no own wealth and $W = 0$, the first-best and second-best solutions remain feasible because the agent receives a positive payment under the optimal mechanism. This is not true if the agent is a buyer. Then, the principal will sell to the agent only if he receives a positive payment, which requires a positive endowment $W > 0$. Assumption 2 also allows us to focus attention on the interesting case where the agent's liability is large enough for the first-best to be implementable at $\pi = 1$, i.e., if the agent's type is perfectly revealed ex post.²⁰

Given the distribution of the signal as described above, the agent's participation constraints can be written as

$$\pi t_i(s_i) + (1 - \pi)t_i(s_j) + u(x_i, \theta_i) \geq \bar{u} \quad i, j \in \{l, h\}, i \neq j. \quad (\text{IR}_i)$$

¹⁸For an analysis of situations where the agent alone knows W , see Lewis and Sappington (2000, 2001). The relevance of limited liability constraints has long been stressed in the literature on moral hazard [see, e.g., Sappington (1983) on hidden information and Brander and Spencer (1989) on hidden action].

¹⁹Notice that an additional assumption $w + \bar{u} \geq u(x_h^{SB})$ is redundant here: since the high-type agent earns an informational rent under asymmetric information, our previous results imply that $t_h^{SB} < t_h^{FB}$.

²⁰To see this, suppose $\pi = 1$ and consider the first-best contract with the transfer equal to $-W$ in case the agent did not tell the truth. Under Assumption 2, this contract is incentive compatible because $\bar{u} \geq u(x_l^{FB}, \theta_h) - W$ and $\bar{u} \geq u(x_h^{FB}, \theta_l) - W$.

In addition, incentive compatibility now reads for $i, j \in \{l, h\}$, $i \neq j$,

$$\pi t_i(s_i) + (1 - \pi)t_i(s_j) + u(x_i, \theta_i) \geq \pi t_j(s_i) + (1 - \pi)t_j(s_j) + u(x_j, \theta_i). \quad (\text{IC}_i)$$

The principal's problem is to choose $\{x_i, t_i(s_j)\}$ so as to

$$\begin{aligned} \text{maximize} \quad & q \left[v(x_h, \theta_h) - \pi t_h(s_h) - (1 - \pi)t_h(s_l) \right] \\ & + (1 - q) \left[v(x_l, \theta_l) - \pi t_l(s_l) - (1 - \pi)t_l(s_h) \right] \\ \text{s.t.} \quad & (\text{IR}_i), (\text{IC}_i) \text{ and } (\text{LL}_{ij}). \end{aligned} \quad (\text{P})$$

As can easily be seen, the principal's return from the relationship is now strictly higher than under the optimal second-best contract where no informative signal is available. Also, if the signal is sufficiently precise, the possibility of contracting on s enables her to implement the first-best. The relevant cut-off value is given by (see the Appendix)

$$\pi^{FB} \equiv \frac{\bar{u} + W - u(x_l^{FB}, \theta_l)}{2 [\bar{u} + W - u(x_l^{FB}, \theta_l)] - \phi(x_l^{FB})} \in \left(\frac{1}{2}, 1 \right).$$

In the following, we first focus our investigation on how additional ex post information affects the feature of the optimal contract that concerns the induced action x , turning to the question of informational rents later. As the effect on the efficiency of the contractually induced allocation will crucially depend on whether the agent's utility is increasing or decreasing in x , we consider each of these two possibilities in turn.

Let us start with a situation where the agent is in a buyer position. For concreteness, let the principal be a monopolistic supplier who delivers his products to a retail store. The efficiency of the store depends on the sales skills of its owner or manager. The skill level is private knowledge of the agent, and θ_H refers to an agent with high skills.²¹ In this case, the agent's gross utility $u(\cdot)$ is increasing in x and the principal might make the agent's transfer payment (the price charged, a license fee, or a royalty) dependent on an ex-post available signal that is related to the retailer's productivity. This signal may be the retailer's profit which is a stochastic indicator for his sales skills.

²¹In the apparel industry, MORE

Proposition 1. *Suppose $u(x, \theta_i)$ is a monotonically increasing function of x . Under the optimal contract with a verifiable ex post signal, the action taken by the low-type agent, x_l^S , is continuous in the informativeness π of the signal and efficient for $\pi \geq \pi^{FB}$. Otherwise, we have $x_l^S \leq x_l^{FB}$ and there exists a value $\underline{\pi} \in (\frac{1}{2}, \pi^{FB})$ such that*

$$\pi < \underline{\pi} \Leftrightarrow x_l^S < x_l^{SB}.$$

The action taken by the high-type agent is efficient, i.e., $x_h^S = x_h^{FB}$ irrespective of π .

The main result of Proposition 1 is illustrated in Figure 2, which depicts the equilibrium action in state θ_l as a function of the precision of the signal, π .

We see that the induced distortions are alleviated relative to a situation where no signal is available *only if* the ex post signal is sufficiently precise. In particular, additional information reinforces the inefficiency whenever $\pi < \underline{\pi}$, so that the total surplus continues to fall over this range as the signal becomes more and more informative. This finding seems to lack intuition at first glance; after all, we know that the optimal contract balances the principal's concern for surplus maximization with her desire to extract rents. Since an informative signal alleviates the latter problem, the extent of this trade-off should change in favor of surplus maximization. This intuition is flawed, however, because it ignores the fact that additional information may be used most effectively to reduce the agent's rent by introducing further distortions. To see this more clearly, recall from the second-best benchmark that for an uninformative signal with $\pi = \frac{1}{2}$ only (IR_l) and (IC_h) bind and the informational rent is strictly positive. For π not too large, this will continue to be the case and we can use (IR_l) and (IC_h) to write the informational rent of a high-type agent as

$$\begin{aligned} R(x_l, \pi) &= u(x_h, \theta_h) + \pi t_h(s_h) + (1 - \pi)t_h(s_l) - \bar{u} \\ &= [u(x_l, \theta_h) + \pi t_l(s_h) + (1 - \pi)t_l(s_l)] - [u(x_l, \theta_l) + \pi t_l(s_l) + (1 - \pi)t_l(s_h)] \\ &= \phi(x_l) - (2\pi - 1)[t_l(s_l) - t_l(s_h)], \end{aligned} \tag{6}$$

where $\phi(x_l) = R(x_l, \frac{1}{2})$ is the corresponding rent in the no-signal case. The last term on the right hand side of (6) reflects the fact that a dishonest high-type agent faces a different probability distribution of the signal than a truthful low-type agent. In particular, since the probability of a signal $s = s_h$ is higher for the former, it is

optimal for the principal to set the agent's compensation in case of contradicting evidence as low as possible. Hence, we must have $t_l(s_h) = -W$ and (6) becomes

$$R(x_l, \pi) = \phi(x_l) - (2\pi - 1)[t_l(s_l) + W]. \quad (7)$$

The term in square brackets is the net expected punishment inflicted on an agent who untruthfully claims to be of type θ_l . It consists of the difference between the transfer $t_l(s_l)$ that the agent receives if the signal falsely indicates that he was honest and the minimum transfer when the signal (correctly) indicates non-compliance, $t_l(s_h) = -W$. If the transfer payment is negative as would be the case in a vertical (or credit) relationship with the agent buying (or borrowing) from the principal, we have $-t_l(s_h) = W > -t_l(s_l) > 0$.²² Inspection of (7) reveals that in order to reduce the rent, the principal will want to raise the payment $t_l(s_l)$ in comparison to the no-signal case. What does increasing $t_l(s_l)$ mean in terms of the action x_l ? From (IR_l), we have

$$t_l(s_l) = [\bar{u} - u(x_l, \theta_l) + (1 - \pi)W]/\pi, \quad (8)$$

i.e., $u'(x, \theta) > 0$ implies that $t_l(s_l)$ decreases in x_l . Hence, a *reduction* in x_l diminishes the agent's rent relative to the no-signal second-best case by the above argument. In our monopolistic supplier example, the smaller the quantity purchased from the principal, the lower the price a truthful agent has to pay, and the more money is left in the pocket of an untruthful agent that can serve as a penalty if the signal indicates non-compliance.

Consequently, the availability of additional information through s aggravates the welfare reduction caused by the downward distortion of x_l , provided π is relatively small. Indeed, this effect becomes stronger the more precise the signal is for values $\pi \leq \hat{\pi}$.²³ At $\hat{\pi}$, the agent's rent drops to zero and (IR_h) becomes binding. Then, the above effect is no longer operative and the principal optimally increases x_l until the (IC_h) constraint becomes slack at $\pi = \pi^{FB}$.

Next, we turn to the case where the agent is a seller, who delivers goods or services of unknown quality to some customer. In this case, the agent's utility is decreasing in x , and his production costs might be inferred from some ex-post signal that indicates

²²Note that the first inequality is a consequence of Assumption 2.

²³Formally, $R'(x_l, \pi) = \phi'(x_l) + \frac{2\pi-1}{\pi}u'(x_l, \theta_l)$, i.e., the principal's marginal benefit from reducing x_l , is increasing in π .

the quality of service.²⁴ Note that in our example, the index l refers to the high-cost and high-quality seller. Let

$$\bar{\pi} \equiv \frac{u'(x_l^{FB}, \theta_l)}{u'(x_l^{FB}, \theta_l) + u'(x_l^{FB}, \theta_h)} \in \left(\frac{1}{2}, 1\right).$$

Proposition 2. *Suppose $u(x, \theta_i)$ is a monotonically decreasing function of x . Under the optimal contract with a verifiable ex post signal, the low-type agent's action x_l^S is continuous in π and efficient for $\pi \geq \pi^{FB}$. Otherwise, $x_l^S \geq x_l^{SB}$ and*

- a) *if $\bar{\pi} \geq \pi^{FB}$, we have $x_l^S < x_l^{FB}$ for all values $\pi \geq \frac{1}{2}$;*
- b) *if $\bar{\pi} < \pi^{FB}$, we have $x_l^S > x_l^{FB} \Leftrightarrow \pi > \bar{\pi}$.*

The action taken by the high-type agent is efficient, i.e., $x_h^S = x_h^{FB}$ irrespective of π .

Again, the content of this proposition is best illustrated graphically. This is done in Figure 3, which displays x_l^S as a function of π .

According to the proposition, two possibilities have to be distinguished. First, if $\bar{\pi} \geq \pi^{FB}$, the downward distortion of x_l is simply alleviated so that total welfare increases as additional ex post information is present; i.e., the more precise the signal, the smaller is the distortion, as has been suggested by the literature discussed in the Introduction. This is case a), as indicated by the thin line. Importantly, case a) always applies if $\bar{u} = W = 0$ and $u(x, \theta) = -x\theta$, implying $\bar{\pi} = \pi^{FB}$ from the definitions of $\bar{\pi}$ and π^{FB} . This particular specification, which is completely innocuous in the standard model without ex post information, has been used in Demougin and Garvie (1991) as well as in Khalil (1997). Both papers, therefore, correctly conclude that a transfer dependent penalty by itself does not generate overproduction and the well-known result of downward distorted production continues to hold.²⁵ However,

²⁴For example, the car may break down shortly after a mechanic (the agent) repaired it. Note that $u'(x, \theta) < 0$ also holds in a variant of the adverse selection model which includes unobservable effort and is often encountered in the literature. For example, we could have $x = \theta + e$ where x is an output or cost related variable and e the effort exerted by the agent. Effort is costly, so that the agent's utility can be written as $u_A = -C(e) + t = t - C(x - \theta)$ with $u'(x, \theta) = -C_e < 0$. Some well-known applications of this framework are the regulation of a firm with unknown cost as formalized by Laffont and Tirole (1993), or the manager-shareholder model by Kofman and Lawarrée (1993).

²⁵For a model where overproduction may occur despite the normalization $\bar{u} = W = 0$, see Khalil and Lawarrée (2001). The authors consider a very different framework in which the principal ex post chooses between input and output monitoring, thereby taking advantage of the possibility that the agent can be caught on the wrong foot.

$\bar{\pi} < \pi^{FB}$ is easily possible if u is strictly concave (see the Appendix for a simple example) or if $W > 0$. This is case b), which is depicted by the thick line in the figure. If $\bar{\pi} < \pi^{FB}$, the availability of noisy ex post information on the agent's type will lead the principal to distort x_l in the *opposite* direction. To see why, recall that the informational rent of the high-type agent is given by (7), provided R is strictly positive. As before, we have $t_l(s_h) = -W$ and the principal wants to raise $t_l(s_l)$ above the amount that is optimal in the no-signal case in order to reduce the rent. In contrast to the previous case, however, $t_l(s_l)$ is now increasing in x_l due to $u'(x, \theta) < 0$ [see eq. (8)]. In our trade-quality example, the higher the amount of output a high-cost (and high-quality) agent is to produce, the more he receives in compensation and the more severely an untruthful agent can be punished if the signal indicates low quality. In particular, this punishment has the natural interpretation of a warranty which covers malfunctioning (a signal indicating low quality and thus low costs).²⁶

Therefore, the high-cost seller's output x_l always unambiguously exceeds x_l^{SB} , regardless of whether we are in case a) or case b). Also, as long as π is sufficiently small, the effect again becomes stronger the more informative the signal is.²⁷ But since the driving force behind raising x_l is the *marginal* – rather than the absolute – effect on the agent's rent, the effect may even lead the principal to raise x_l above x_l^{FB} .²⁸ Specifically, this situation emerges if the marginal effect on $-u(x_l^{FB}, \theta_l)$ and, hence, on $t_l(s_l)$ is sufficiently stronger than the marginal effect on $\phi(x_l^{FB}) = u(x_l^{FB}, \theta_h) - u(x_l^{FB}, \theta_l)$, i.e., for $\bar{\pi} < \pi^{FB}$. Again, if the agent's rent R drops to zero and (IR_h) becomes binding at $\hat{\pi}$, this effect vanishes. For high values of π , the (downward or upward) distortion is thus alleviated and eventually, the

²⁶See Emons (1989) for a survey of warranty contracts and the rationales behind them. A slightly different formulation of the problem would also allow the penalty to be interpreted as a money-back guarantee (with $W = 0$).

²⁷Formally, $R'(x_l, \pi)$ is now decreasing in π .

²⁸This result should be contrasted to Laffont and Tirole (1993, ch. 12), who conclude that “monitoring of effort [...] consequently leads to a smaller distortion of effort for the inefficient type” (p. 529). In their model, monitoring generates a signal which is perfectly revealing but observed with less than full probability, $z < 1$. The authors implicitly assume that i) the agent cannot be punished if no signal has been observed (no evidence and favorable evidence cannot be distinguished) and that ii) z is smaller than some threshold value z^* (which corresponds to $\bar{\pi}$ in our context). If one maintains i) but allows for signal frequencies $z > z^*$, a positive output distortion would emerge for the same reason as in our model. Also, if assumption i) is abolished, it is easy to see that the first-best can be achieved in their framework.

first-best is implemented for $\pi \geq \pi^{FB}$.

Having analyzed the allocative consequences of additional information for the two cases of a buyer-agent and a seller-agent, respectively, we can now explore how individual payoffs will be affected. Clearly, the principal's expected return never falls as an informative signal becomes available: because she could always offer the optimal second-best contract that does not condition on the signal, this observation follows trivially from revealed preferences. In light of the previous results, however, it is not *a priori* obvious how the agent's informational rent $R(x_l^S, \pi)$ is affected by additional ex post information.

Corollary. *If $u'(x, \theta_i) > 0$, $R(x_l^S, \pi)$ is weakly decreasing in the precision of the signal, π . If $u'(x, \theta_i) < 0$, $R(x_l^S, \pi)$ weakly decreases in π for all values $\pi \geq \bar{\pi}$, but it may increase in π for values $\pi < \bar{\pi}$. In particular, we may have $R(x_l^S, \pi) > R(x_l^{SB}, \frac{1}{2}) = \phi(x_l^{SB})$ for some $\pi \in (\frac{1}{2}, \bar{\pi})$.*

The first part of the corollary is straightforward. Consider Figure 2 where the agent is a buyer and $u'(x, \theta_i) > 0$. For values of $\pi < \hat{\pi}$, the distortion of x_l is aggravated and the total surplus decreases in π . Hence, a buyer-agent cannot be better off if additional ex post information about him becomes available. For $\pi \geq \hat{\pi}$, his rent is zero, so that a more precise ex post signal cannot help him either. But a similar reasoning need not apply for a seller-agent, $u'(x, \theta_i) < 0$, as is indicated in the second part of the corollary: the low-cost seller may actually *benefit* from an observable ex post signal about his type. From Figure 3, we see that the informational rent can only decrease for values $\pi \geq \bar{\pi}$. In case a), the seller-agent earns no rent for those values of π . In case b), x_l increases above x_l^{FB} , so that the total surplus again decreases (if $\pi \leq \hat{\pi}$) or the agent's rent is again zero (if $\pi > \hat{\pi}$). However, if $\pi < \bar{\pi}$ more precise ex post information reduces the distortion, so that total surplus increases. Then, not only the principal's profit, but also the agent's rent can increase. In particular, one can easily construct examples where the negative effect on $R(\cdot)$ due to a more precise signal is overcompensated by the positive effect on $R(\cdot)$ caused through the corresponding raise in x_l .²⁹ In our trade-quality scenario, this result implies that a seller-agent might actually gain when a buyer has the ability to prove a low quality of service.

²⁹See the Appendix for a simple example.

5 Conclusion

In this paper, we have studied a straightforward variant of a simple principal-agent adverse selection problem in which information that is imperfectly correlated with the agent's type becomes public ex post. Because the agent was protected by limited liability, the principal could not implement the first-best allocation with signals of arbitrary informativeness, which is in line with previous findings. The main results of our paper, though, are in stark contrast to findings in the previous literature. Perhaps most importantly, we show that the expected surplus from the relationship need not increase as the signal becomes available or more informative, respectively. Rather, there is a range of situations in which more contractually employable information leads to less efficiency and lower welfare overall. This is in contrast to the standard result where the allocative downwards distortion of low-type agents is mitigated as additional information of the agent's type becomes available (e.g., through audits).

As a second main result, we have also seen that the qualitative effects of additional information can unambiguously be tied to the nature of the underlying problem, namely, of whether the agent is in a 'buyer' or a 'seller' position. If the agent's utility is strictly increasing in the contractually specified action, we found that downward distortions will be strengthened if the signal is not too informative. This would imply that the equilibrium volume of trade drops if manufacturers use more information in their contractual relations with their retailers. Similarly, the equilibrium volume of loans (the agent is granted a credit of size x by the principal) may fall as banks make use of additional information on their customer's creditworthiness. In environmental regulation (the agent prefers to produce more output x , which is harmful for the environment) it might mean that the government may adopt less tight emission standards as more precise monitoring techniques become available. Conversely, additional information can result in upward (rather than downward) distortions if the agent's utility is strictly decreasing in the contractually specified action. In the trade-quality example we used in the text, this would mean that high-cost (and high-quality) producers deliver an inefficiently high amount of goods or provide an excessive amount of services to their customers if warranties or breach penalties can easier be enforced. In an employment context, the work effort of less

productive agents may even exceed the first-best level if the employer has access to a noisy monitoring technology. Finally, we showed that in all instances where the agent is in the seller position, he may also command a higher informational rent in the presence of an informative signal relative to the case where no additional information is available.

These results were derived in a framework where the agent's limited liability is modelled as a maximum bound on his total payments. Total payments are the sum of the regular transfer payment to be made in a bilateral transaction, and a possible punishment (a breach penalty) if the signal indicates the agent's non-compliance. An important feature of this interpretation is that the maximum punishment to be inflicted on an agent is endogenous, i.e., it is positively (agent-buyer) or negatively (agent-seller) related to the regular transaction payment and hence to the contractual terms of trade. In contrast, the literature often subscribes to an alternative understanding of limited liability, according to which the agent's maximum punishment is some exogenous level which is functionally unrelated to other contractual provisions between the parties. We think that depending on the specific situation one or the other of these alternative interpretations might be more realistic, and our results suggest that additional research in this area is highly desirable.

Although we have used a very simple framework to highlight the economic forces at work, several extensions of our model could be considered. In particular, we have treated the ex post signal itself as well as its informativeness as exogenously given. A natural extension of our model would allow for the information to be observed only if the principal conducts costly audits, or for an endogenous choice of the informativeness of the signal. As the principal's payoff is always higher if the signal is available and monotonically increasing in its precision, neither extension is likely to change our qualitative insights.³⁰ Another and perhaps more promising extension concerns the possibility that the agent himself can affect the signal that is received by the principal. In the manufacturer-retailer example, for instance, the

³⁰This argument notwithstanding, taking the choice of monitoring as exogenous may lead to wrong conclusions regarding the effects of variations in the informativeness of the signal on the variables of the model [see, e.g., Demougin and Fluet (2001)]. Also observe that since the agent is punished with positive probability in equilibrium, the principal might find it optimal to conduct costly audit even if she cannot commit to do so ex ante.

agent's private information may relate to demand conditions and he may be able to influence the price of final output that is produced using the principal's input. Then, the equilibrium signal itself will depend on the contracting terms and thus be chosen endogenously. More generally, suppose the agent can manipulate ex post information at some personal cost. Our findings then suggest that, surprisingly, total surplus could actually be enhanced if the agent's cost of manipulating ex post information decreases.³¹ Moreover, the agent may be better off if he can release some public information ex post, which may be an interesting topic for future research.

³¹See Maggi and Rodríguez-Clare (1995) for a related result in a model based on countervailing incentives.

Appendix

The optimal contract $\{(t_i^S(s_j), x_i^S)\}_{i,j=l,h}$ solves the program (P), the solution of which is characterized in the lemma below. Define $\Phi(x, \pi) \equiv u(x, \theta_h) - \frac{1-\pi}{\pi}u(x, \theta_l) - \frac{2\pi-1}{\pi}(\bar{u} + W)$ and note that

$$\hat{x}_l(\pi) = \arg \max_{x \in [\underline{x}, \bar{x}]} S(x, \theta_l) - \frac{q}{1-q} \Phi(x, \pi). \quad (9)$$

Lemma. *The optimal contract $\{x_i^S, t_i^S\}$ always prescribes $x_h^S = x_h^{FB}$. For values of $\pi \geq \pi^{FB}$, we have $x_l^S = x_l^{FB}$ and $\Phi(x_l^S, \pi) \leq 0$. For $\pi < \pi^{FB}$, x_l^S is characterized by*

a) $\Phi(x_l^S, \pi) = 0$ if $\Phi(\hat{x}_l(\pi), \pi) \leq 0$,

b) $x_l^S = \hat{x}_l(\pi)$ otherwise.

The transfers under this contract are $t_l^S(s_h) = -W$, $t_l^S(s_l) = [\bar{u} - u(x_l^S, \theta_l) + (1 - \pi)W]/\pi$ and $t_h^S(s_h) = t_h^S(s_l) = \bar{u} - u(x_h^{FB}, \theta_h) + \max\{\Phi(x_l^S, \pi), 0\}$.

Proof. Note first that by invoking the Maximum Punishment Principle [Baron and Besanko (1984)], we can set $t_l(s_h) = -W$ and let $t_l(s_l) \equiv t_l$ for brevity of exposition. Second, provided the (IC_l) constraint is slack, we can without loss of generality assume that $t_h(s_h) = t_h(s_l) \equiv t_h$. In what follows, we will ignore the (IC_l) constraint and later verify that it is indeed not binding in the optimum. Furthermore, since $t_h^S \geq \bar{u} - u(x_h^{FB}, \theta_h) \geq -W$ by (5), the wealth constraint for the θ_h -type agent will not be binding and can be ignored. Similarly, for values $\pi \geq \pi^{FB}$ we have $x_l^S = x_l^{FB}$, implying $t_l^S \geq -W$ by (5). An argument that the wealth constraint $t_l \geq -W$ is not binding for $\pi < \pi^{FB}$ is deferred to the proofs of Propositions 1 and 2 below. Rewriting the principal's payoff and the remaining constraints, the Lagrangian of the principal's problem is

$$\begin{aligned} \mathcal{L} = & q[v(x_h, \theta_h) - t_h] + (1-q)[v(x_l, \theta_l) - \pi t_l + (1-\pi)W] \\ & + \lambda_l \{\pi t_l - (1-\pi)W + u(x_l, \theta_l) - \bar{u}\} + \lambda_h \{t_h + u(x_h, \theta_h) - \bar{u}\} \\ & + \mu \{t_h + u(x_h, \theta_h) - (1-\pi)t_l + \pi W - u(x_l, \theta_h)\}. \end{aligned}$$

The first-order conditions are

$$\frac{\partial \mathcal{L}}{\partial x_h} = qv'(x_h, \theta_h) + (\lambda_h + \mu)u'(x_h, \theta_h) = 0 \quad (10)$$

$$\frac{\partial \mathcal{L}}{\partial t_h} = -q + \lambda_h + \mu = 0 \quad (11)$$

$$\frac{\partial \mathcal{L}}{\partial x_l} = (1 - q)v'(x_l, \theta_l) + \lambda_l u'(x_l, \theta_l) - \mu u'(x_l, \theta_h) = 0 \quad (12)$$

$$\frac{\partial \mathcal{L}}{\partial t_l} = -(1 - q)\pi + \lambda_l \pi - \mu(1 - \pi) = 0. \quad (13)$$

Inserting (11) into (10), using $S(x, \theta) = v(x, \theta) + u(x, \theta)$, yields

$$S'(x_h, \theta_h) = 0 \quad \Rightarrow \quad x_h^S = x_h^{FB}.$$

Next, (13) implies $\lambda_l = (1 - q) + \mu \frac{1 - \pi}{\pi} > 0$. Thus, (IR_l) is always binding. We can substitute λ_l in (12) to obtain

$$S'(x_l, \theta_l) = \frac{\mu}{1 - q} \left[u'(x_l, \theta_h) - \frac{1 - \pi}{\pi} u'(x_l, \theta_l) \right]. \quad (14)$$

Consider first $\mu = 0$. Then, $x_l = x_l^{FB}$ by (14) and $\lambda_h = q > 0$ from (11), so (IR_h) is satisfied with equality. Since (IR_l) is binding, in order for (IC_h) to hold in this case, we must have

$$u(x_l^{FB}, \theta_h) - \frac{1 - \pi}{\pi} [u(x_l^{FB}, \theta_l) - \bar{u}] - \frac{2\pi - 1}{\pi} W \leq \bar{u},$$

which is equivalent to $\Phi(x_l^{FB}, \pi) \leq 0$ or $\pi \geq \pi^{FB}$ as defined in the text. Conversely, for $\pi < \pi^{FB}$, $\mu = 0$ yields a contradiction, so (IC_h) is binding. There are two cases to distinguish:

a) $\lambda_h > 0$ implies that both (IR_h) and (IC_h) are binding at the optimum. x_l^S is then implicitly characterized by

$$\Phi(x_l^S, \pi) = u(x_l^S, \theta_h) - \frac{1 - \pi}{\pi} u(x_l^S, \theta_l) - \frac{2\pi - 1}{\pi} (\bar{u} + W) = 0.$$

b) $\lambda_h = 0$ yields $\mu = q$ from (11) and x_l^S can be recovered from equation (14),

$$S'(x_l^S, \theta_l) = \frac{q}{1 - q} \left[u'(x_l^S, \theta_h) - \frac{1 - \pi}{\pi} u'(x_l^S, \theta_l) \right], \quad (15)$$

implying $x_l^S = \hat{x}_l(\pi)$. Note that $\lambda_h = 0$ requires $\Phi(\hat{x}_l(\pi), \pi) \geq 0$.

To complete the proof, it remains to show that (IC_l) is not binding. Using (IR_l) and (IC_h) , this constraint can be written as

$$u(x_h^{FB}, \theta_l) - u(x_h^{FB}, \theta_h) + \max \{ \Phi(x_l^S, \pi), 0 \} \leq 0. \quad (16)$$

Observe first that if $\Phi(x_l^S, \pi) \leq 0$, inequality (16) is implied by Assumption 1 a). Thus, we can confine attention to case b) where (IR_h) is not binding and $x_l^S = \hat{x}_l(\pi)$. Next, note that for (16) to hold it suffices to prove that $\Phi(x_l^S, \pi) \leq \Phi(x_l^{FB}, \frac{1}{2})$, because $\Phi(x_l^{FB}, \frac{1}{2}) = u(x_l^{FB}, \theta_h) - u(x_l^{FB}, \theta_l)$, so (16) will again be satisfied by Assumptions 1 b) and c). Consider the following alternative contract: $x_i = x_i^{FB}$, $t_l(s_l) = t_l(s_h) = \bar{u} - u(x_l^{FB}, \theta_l)$ and $t_h = \bar{u} - u(x_h^{FB}, \theta_h) + \Phi(x_l^{FB}, \frac{1}{2})$. Since this contract satisfies all the constraints, the principal's expected utility under the optimal contract must weakly exceed her expected utility under the alternative contract. Hence,

$$S(x_l^S, \theta_l) - \frac{q}{1-q} \Phi(x_l^S, \pi) \geq S(x_l^{FB}, \theta_l) - \frac{q}{1-q} \Phi(x_l^{FB}, \frac{1}{2}),$$

by revealed preferences. The claim follows from $S(x_l^{FB}, \theta_l) \geq S(x_l^S, \theta_l)$ by definition of x_l^{FB} . \square

Proof of Proposition 1: As Lemma 1 already characterizes x_h^S as well as x_l^S for $\pi \geq \pi^{FB}$, we focus on how x_l^S varies with π for $\pi < \pi^{FB}$. Suppose $u'(x, \theta_i) > 0 \forall x \in [\underline{x}, \bar{x}]$ and recall from Section 2 that at $\pi = \frac{1}{2}$, $x_l^S = x_l^{SB} = \hat{x}_l(1/2)$ and $\Phi(x_l^{SB}, 1/2) > 0$. Since the objective function in (9) has decreasing marginal returns, $\hat{x}_l(\pi)$ is a strictly decreasing function of π [see, e.g., Theorem 1 in Edlin and Shannon (1998)].³² Also note that

$$\begin{aligned} \frac{d\Phi(\hat{x}_l(\pi), \pi)}{d\pi} &= \frac{\partial\Phi(\hat{x}_l(\pi), \pi)}{\partial\pi} + \frac{\partial\Phi(\hat{x}_l(\pi), \pi)}{\partial\hat{x}_l} \frac{\partial\hat{x}_l}{\partial\pi} \\ &= -\frac{1}{\pi^2} [\bar{u} + W - u(\hat{x}_l, \theta_l)] + \frac{1-q}{q} S'(\hat{x}_l, \theta_l) \frac{\partial\hat{x}_l}{\partial\pi}, \end{aligned} \quad (17)$$

where we have used the definition of $\Phi(\cdot)$ and the fact that \hat{x}_l satisfies the first-order condition (15) to program (9). Due to $\partial\hat{x}_l(\pi)/\partial\pi < 0$, we must have $\hat{x}_l < x_l^{SB}$ and, hence, $S'(\cdot) > 0$ and $\bar{u} + W - u(\hat{x}_l, \theta_l) > 0$ using (5) and $u'(\cdot) > 0$.

³²A function $f(x, t)$ is said to display increasing (decreasing) marginal returns, if $\frac{\partial f}{\partial x}$ is strictly increasing (decreasing) in t . This property is sometimes also referred to as (strict) supermodularity.

Therefore, $\frac{d\Phi}{d\pi} < 0$, i.e., the agent's rent is decreasing in π as one would expect. Since $\Phi(\hat{x}_l(\pi^{FB}), \pi^{FB}) < \Phi(x_l^{FB}, \pi^{FB}) = 0$, there exists a unique $\hat{\pi} \in (\frac{1}{2}, \pi^{FB})$ such that $\Phi(\hat{x}_l(\hat{\pi}), \hat{\pi}) = 0$. Now consider increasing π from $1/2$ to $\hat{\pi}$, i.e., over the range where $\Phi(x_l^S, \pi) > 0$ so that the (IR_h) constraint is not binding. Then, we are in case b) of Lemma 1 and $x_l^S = \hat{x}_l(\pi)$. Hence, x_l^S is decreasing in π for $\pi \in [1/2, \hat{\pi}]$. At $\hat{\pi}$, the (IR_h) -constraint becomes binding. For values $\hat{\pi} \leq \pi \leq \pi^{FB}$, we are in case a) of Lemma 1, where x_l^S is implicitly determined by $\Phi(x_l^S, \pi) = 0$. By the implicit function theorem,

$$\frac{\partial x_l^S}{\partial \pi} = -\frac{\partial \Phi(\cdot)/\partial \pi}{\partial \Phi(\cdot)/\partial x_l^S} = -\frac{-\frac{1}{\pi^2} [\bar{u} + W - u(x_l^S, \theta_l)]}{\frac{1-q}{\mu} S'(x_l^S, \theta_l)}, \quad (18)$$

which is strictly positive as $x_l^S \leq x_l^{FB}$ using a similar argument as above. Continuity follows from the Theorem of the Maximum. To summarize, x_l^S at first decreases in π below x_l^{SB} , obtains a minimum at $\hat{\pi}$ and then increases again until $x_l^S = x_l^{FB}$ at $\pi = \pi^{FB}$. The existence of the critical value $\underline{\pi}$ as stated in Proposition 1 now follows from an intermediate value argument. Finally, note that (5) and $x_l^S \leq x_l^{FB}$ together with $u' > 0$ ensure that $\bar{u} + W - u(x_l^S, \theta_l) \geq 0$ over the entire range. This implies $t_l^S \geq -W$, so that the contract characterized in Lemma 1 also satisfies the wealth constraint for the θ_l -type agent. \square

Proof of Proposition 2: Suppose $u'(x, \theta_i) < 0 \forall x \in [\underline{x}, \bar{x}]$. Again, we confine attention to the case where $\pi < \pi^{FB}$ and investigate how x_l^S varies with π , starting from $x_l^S = x_l^{SB} = \hat{x}_l(\frac{1}{2})$ at $\pi = \frac{1}{2}$. Due to $u' < 0$, the objective function in (9) has increasing marginal returns and $\hat{x}_l(\pi)$ is therefore a strictly increasing function of π . Since $\Phi(\hat{x}_l(\frac{1}{2}), \frac{1}{2}) > 0$, case b) of Lemma 1 again applies for π sufficiently close to $\frac{1}{2}$, implying $x_l^S = \hat{x}_l(\pi)$. Moreover, note that $\bar{\pi}$ as defined in the text is such that $\hat{x}_l(\bar{\pi}) \geq x_l^{FB} \Leftrightarrow \bar{\pi} \geq \pi^{FB}$. There are two possibilities to consider.

First, if $\pi^{FB} < \bar{\pi}$, then $\hat{x}_l(\pi) < x_l^{FB}$ for all $\pi \leq \pi^{FB}$. Since $\frac{\partial \Phi}{\partial \hat{x}_l} > 0$ for $\hat{x}_l < x_l^{FB}$, we have $\Phi(\hat{x}_l(\pi^{FB}), \pi^{FB}) < \Phi(x_l^{FB}, \pi^{FB}) = 0$, so that by continuity there must exist a $\hat{\pi} \in (1/2, \pi^{FB})$ with $\Phi(\hat{x}_l(\hat{\pi}), \hat{\pi}) = 0$ such that x_l^S is characterized as in Lemma 1 a) for $\pi \geq \hat{\pi}$. Over this range, $\Phi(x_l^S, \pi) = 0$, so that x_l^S continues to be increasing [see (18) and note that $x_l^{SB} < x_l^S < x_l^{FB}$]. Hence, x_l^S monotonically increases in this case from x_l^{SB} to x_l^{FB} as π varies from $\frac{1}{2}$ to π^{FB} . Second, we may have $\bar{\pi} < \pi^{FB}$, so that $\hat{x}_l(\pi) > x_l^{FB}$ for all $\pi \in (\bar{\pi}, \pi^{FB}]$. From $\frac{\partial \Phi}{\partial \hat{x}_l} < 0$ for

$\hat{x}_l > x_l^{FB}$, it follows again that $\Phi(\hat{x}_l(\pi^{FB}), \pi^{FB}) < \Phi(x_l^{FB}, \pi^{FB}) = 0$. Furthermore, $\Phi(\hat{x}_l(\bar{\pi}), \bar{\pi}) = \Phi(x_l^{FB}, \bar{\pi}) > \Phi(x_l^{FB}, \pi^{FB}) = 0$, because $\frac{\partial \Phi}{\partial \pi} < 0$. By continuity there thus exists a $\hat{\pi} \in (\bar{\pi}, \pi^{FB})$ with $\Phi(\hat{x}_l(\hat{\pi}), \hat{\pi}) = 0$ such that x_l^S is characterized as in Lemma 1 a) for $\pi \geq \hat{\pi}$. Over this range, x_l^S is decreasing in π as can be seen from (18) using $x_l^{FB} < x_l^S$. To summarize, for $\bar{\pi} < \pi^{FB}$, x_l^S at first increases in π above x_l^{FB} for $\pi > \bar{\pi}$, obtains a maximum at $\hat{\pi}$, and then decreases again until $x_l^S = x_l^{FB}$ at $\pi = \pi^{FB}$. Again, it is straightforward to show that $\bar{u} + W - u(x_l^S, \theta_l) \geq 0$ over the entire range, which implies $t_l^S \geq -W$, so that the contract characterized in Lemma 1 is indeed optimal. \square

Proof of the Corollary: The first part of the corollary follows immediately from the proof of Proposition 1. In order to prove the second part, suppose $u'(x, \theta_i) < 0$. The agent's rent is given by $\max\{\Phi(\hat{x}_l(\pi), \pi), 0\}$, where $\Phi(\hat{x}_l(\pi), \pi)$ varies with π according to (17). The first term in (17) is negative due to $\hat{x}_l \geq x_l^{SB}$ and (5). The second term is also negative if $\hat{x}_l(\pi) \geq x_l^{FB}$, which is equivalent to $\pi \geq \bar{\pi}$. For $\hat{x}_l(\pi) < x_l^{FB}$, however, the second term is positive. It is straightforward to construct examples where the second effect overcompensates the first effect for some $\pi < \bar{\pi}$. For instance, let $u(x, \theta) = -\frac{1}{2}(x - \theta)^2$, $v(x, \theta) = x$, $q = .5$, $\theta_l = .5$, $\theta_h = 1$, $\underline{x} = 1$, $\bar{x} = 3$, and $W = \bar{u} = 0$. It can easily be verified that $\bar{\pi} = 2/3$ and $\pi^{FB} = .8$, and that the agent's rent is increasing for $\pi \in [.5, .6]$. \square

References

- Arruñada, B., Garicano, L., and Vázquez, L. “Contractual Allocation of Decision Rights and Incentives: The Case of Automobile Distribution.” *Journal of Law, Economics, and Organization*, Vol. 17 (2001), pp. 257–284.
- Baron, D.P. and Besanko, D. “Regulation, Asymmetric Information, and Auditing.” *RAND Journal of Economics*, Vol. 15 (1984), pp. 447–470.
- Brander, J.A. and Spencer, B.J. “Moral Hazard and Limited Liability: Implications for the Theory of the Firm.” *International Economic Review*, Vol. 30 (1989), pp. 833–849.
- Border, K. and Sobel, J. “Samurai Accountant: A Theory of Auditing and Plunder.” *Review of Economic Studies*, Vol. 54 (1987), pp. 447–470.
- Crémer, J. “Arm’s Length Relationships.” *Quarterly Journal of Economics*, Vol. 110 (1995), pp. 275–295.
- Crémer, J. and McLean, R. P. “Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions.” *Econometrica*, Vol. 56 (1988), pp. 1247–57.
- Demougin, D. and Fluet, C. “Monitoring versus Incentives.” *European Economic Review*, Vol. 45 (2001), pp. 1741–1764.
- Demougin, D. and Garvie, D. A. “Contractual Design with Correlated Information under Limited Liability.” *Rand Journal of Economics*, Vol. 22 (1991), pp. 477–489.
- Dewatripont, M. and Maskin, E. “Contractual Contingencies and Renegotiation.” *RAND Journal of Economics*, Vol. 26 (1995), pp. 704–719.
- Dnes, A.W. “The Economic Analysis of Franchise Contracts.” *Journal of Institutional and Theoretical Economics*, Vol. 152 (1996), pp. 297–324.
- Edlin, A.S. and Shannon, C. “Strict Monotonicity in Comparative Statics.” *Journal of Economic Theory*, Vol. 81 (1998), pp. 201–219.
- Emons, W. “The Theory of Warranty Contracts.” *Journal of Economic Surveys*, Vol. 3 (1989), pp. 34–57.

- Gary-Bobo, R. and Spiegel, Y. “Optimal State-Contingent Regulation Under Limited Liability.” *mimeo* (2003), Tel Aviv University.
- Jullien, B. “Participation Constraints in Adverse Selection Models.” *Journal of Economic Theory*, Vol. 93 (2000), pp. 1–47.
- Khalil, F. “Auditing without Commitment.” *RAND Journal of Economics*, Vol. 28 (1997), pp. 629–640.
- Khalil, F. and Lawarrée, J.P. “Catching the Agent on the Wrong Foot: Ex Post Choice of Monitoring.” *Journal of Public Economics*, Vol. 82 (2001), pp. 327-347.
- Kofman, F. and Lawarrée, J. “Collusion in Hierarchical Agency.” *Econometrica*, Vol. 61 (1993), pp. 629–656.
- Laffont J.-J. and Martimort, D. *The Theory of Incentives: The Principal-Agent Model*, Princeton, N.J.: Princeton University Press, 2001.
- Laffont J.-J. and Tirole, J. *A Theory of Incentives in Procurement and Regulation*. Cambridge, Mass.: MIT Press, 1993.
- Lafontaine, F. and Slade, M.E. “Retail Contracting and Costly Monitoring: Theory and Evidence.” *European Economic Review*, Vol. 40 (1996), pp. 923-932.
- Lewis, T.R. and Sappington, D.E.M. “Countervailing Incentives in Agency Problems.” *Journal of Economic Theory*, Vol. 49 (1989), pp. 294–313.
- Lewis, T.R. and Sappington, D.E.M. “Contracting with Wealth-Constrained Agents.” *International Economic Review*, Vol. 41 (2000), pp. 743–767.
- Lewis, T.R. and Sappington, D.E.M. “Optimal Contracting with Private Knowledge of Wealth and Ability.” *Review of Economic Studies*, Vol. 68 (2001), pp. 21–44.
- Maggi, G. and Rodríguez-Clare, A. “Costly Distortion of Information in Agency Problems.” *RAND Journal of Economics*, Vol. 26 (1995), pp. 675–689.
- McAfee, R. P. and Reny, P.-J. “Correlated Information and Mechanism Design.” *Econometrica* 60 (1992), pp. 395-421.

- Melumad, N.D. and Mookherjee, D. "Delegation as Commitment: the Case of Income Tax Audits." *RAND Journal of Economics*, Vol. 20 (1989), pp. 139–163.
- Nalebuff, B. and Scharfstein, D. "Testing in Models of Asymmetric Information." *Review of Economic Studies*, Vol. 54 (1987), pp. 265–277.
- Riley, J.G. "Ex Post Information in Auctions." *Review of Economic Studies*, Vol. 55 (1988), pp. 409–430.
- Riordan, M.H. "What is Vertical Integration?" In M. Aoki, B. Gustafsson, and O.E. Williamson, eds., *The Firm as a Nexus of Treaties*. London: Sage Publications, 1990.
- Riordan, M.H. and Sappington, D.E.M. "Optimal Contracts with Public ex post Information." *Journal of Economic Theory*, Vol. 45 (1988), pp. 189–199.
- Sappington, D. "Limited Liability Contracts between Principal and Agent." *Journal of Economic Theory*, Vol. 29 (1983), pp. 1–21.