Games and Economic Behavior 91 (2015) 75–96

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Pessimistic information gathering ☆

Elisabetta Iossa [a,b,c,*,1], David Martimort [d,2]

[a] University of Rome Tor Vergata, Italy
[b] CEPR, United Kingdom
[c] IEFE-Bocconi and EIEF, Italy
[d] Paris School of Economics—EHESS, France

ARTICLE INFO

Article history:
Received 2 June 2013
Available online 2 April 2015

JEL classification:
H41
D82

Keywords:
Information gathering
Moral hazard
Asymmetric information
Pessimism

ABSTRACT

An agent gathers information on productivity shocks and accordingly produces on behalf of a principal. Information gathering is imperfect and whether it succeeds or not depends on the agent's effort. Contracting frictions come from the fact that the agent is pessimistic on the issue of information gathering, and there are both moral hazard in information gathering, private information on productivity shocks and moral hazard on operating effort. An optimal menu of linear contracts mixes high-powered, productivity-dependent screening options following "good news" with a fixed low-powered option otherwise.

1. Introduction

Since its early inception, agency theory has put much emphasis on how uncertainty impacts on the design of contracts and the overall performances of principal–agent arrangements.[3] Although implications for organizational design have been significant and far reaching, most of the extant agency literature has taken information structures as exogenous. Whether an agent takes actions before learning the state of nature (like in hidden action models) or after (like with hidden information) is thus a choice of the modeler to depict a given setting. However, in many circumstances agents do invest in information gathering to better tailor their actions to the realization of shocks that may impact on their performances. A more complete and probably best descriptive view of agency relationships should thus endogenize information structures. This paper belongs to a burgeoning literature that precisely contributes to such research program.

☆ For useful comments, we wish to thank Malin Arve, Antoine Bommier, Bernhard Ganglmair, Perrin Lefebvre, Manuel Marfan, David Miller, Wilfried Sand-Zantman and especially Dszeso Salay as well as participants to seminar at University of Pavia and to the 10th Annual International Industrial Organization Conference in Boston. An earlier version of this paper was circulated under the title *"Hidden Action or Hidden Information? How Information Gathering Shapes Contract Design"*. The comments of four referees and the associate editor were instrumental in helping us reshape our ideas and change accordingly this title. All remaining errors are ours.

* Corresponding author. The address of correspondence is: Dept. Economics and Finance, University of Rome Tor Vergata, Via Columbia 2, Rome 00133, Italy.

E-mail addresses: elisabetta.iossa@uniroma2.it (E. Iossa), david.martimort@parisschoolofeconomics.eu (D. Martimort).

1 This author wishes to thank financial support from the University of Padua, Grant No. CPDA121089/12.

2 This author wishes to thank financial support from the program Investissements d'Avenir of the French government (ANR-10-LABX-93-01).

3 Arrow (1986) and Hart and Holmström (1987) offer early important overviews while Laffont and Martimort (2002, Chapter 1) offer a historical perspective.

http://dx.doi.org/10.1016/j.geb.2015.03.014
0899-8256/

Contracting frictions. We consider a principal (referred to as *she* in the sequel) who hires an agent (*he*) to gather information which is relevant for a productive task he exerts on her behalf. Output depends not only on the agent's operating effort but also on a productivity shock. Learning the value of that shock is costly even though the value of information is positive; the agent's effort at the operating stage should indeed be tailored to the exact value of that shock.

This contractual relationship is *a priori* plagued with several frictions that are associated to various sorts of asymmetric information occurring during the course of contracting. First, *ex ante moral hazard* arises when the agent's effort in gathering information remains non-verifiable. Second, this information gathering technology is imperfect; the agent only gets *ex post private information* on productivity shocks with a probability which increases with his *ex ante* effort. Lastly, *ex post moral hazard* also arises when the agent's effort at the operating stage remains non-verifiable. In such context, neither can the principal infer whether a good performance comes from the agent having exerted a high operating effort or from a favorable productivity shock, nor can she ascertain whether this shock was known by the agent when he chose his operating effort.[4]

Finally, and this is a key ingredient of our modeling, the principal and the agent have different subjective beliefs on the outcome of information gathering. More precisely, the agent is more pessimistic than the principal on the likelihood that he will get informed.[5] There are several motivations behind such assumption.[6] First, the principal and the agent may just have different beliefs as a result of differences in learning on the likelihood of being informed. Overly optimistic agents might not have been tough enough in selecting projects in the past and may have been fired from the organization. Entrepreneurs and CEOs may be more optimistic than their subordinates on the likelihood of unveiling the profitability of new projects, or on the possibility to identify opportunities for new investments. Instead, agents may want to keep more pessimistic stances so as to preserve the quasi-rent they may hold under the status quo. Second, and even if our modeling preserves payoffs linearity, this assumption can be viewed as a reduced form for modeling the agent's risk aversion. An agent might be reluctant to operate under contracts that would tailor too precisely his performances to his claims on shocks that he might have learned. He might instead preferred contracts that make no use of such information and leave him with a less risky payoff.[7] Finally, the agent may need to raise funds to finance his investment in information gathering. *Ex post* private information on productivity shocks might create an agency problem with outside financiers who want to recoup their investments out of the proceeds kept by the agent. The impact of such agency cost (again in reduced form) is akin to reducing the value of getting informed.[8]

Contracting distortions. To see why a difference in beliefs generates agency costs, consider first the benchmark scenario where the agent and the principal actually have common beliefs. As contracting takes place *ex ante*, a simple fixed-price contract would make the agent residual claimant. Leaving all proceeds from production to the agent would be enough to induce efficient efforts, not only at the operating stage but also when gathering information. All the surplus could then be pocketed by the principal through a fixed fee. Agency costs are null under those circumstances.

When beliefs differ, the principal finds it costly to induce the right amount of information gathering. Part of the incentives is dissipated by the agent's pessimism. A pessimistic agent must thus be compensated with an extra *belief premium* that increases with his effort in information gathering. Agency costs now matter and reducing those costs calls for reducing that effort.

When gathering information is itself verifiable, the second-best effort that the principal would like to implement is even more distorted below what the agent would choose when being residual claimant (Proposition 1); a pessimistic agent would gather too much information. As a result, the second-best amount of information gathering is obtained by increasing the fixed repayment paid by the agent to the principal when he gets informed.

Even if it is attractive because it introduces a dichotomy between *ex ante* and *ex post* incentives, a menu of fixed-price contracts, with repayments that would be contingent on whether information has been gathered or not, is not always feasible. Indeed, under many circumstances where the agent's expertise is called for, his principal is neither able to ascertain whether information has ever been collected, nor what kind of information has been learned if any. *De facto*, tailoring repayments to whether information gathering has succeeded or not is impossible.

In the more likely, albeit more complex, scenario where neither information collection nor productivity shocks are common knowledge and verifiable, contracts must thus not only induce the right *ex ante* effort in information gathering but also force the agent to reveal whether he has observed the productivity shock and its possible value.[9] For sure, leaving all proceeds of production to the agent would mean paying too high a belief premium and this cannot be optimal. Of course,

[4] The agency model we consider thus belongs to the class of so-called "mixed models" which combine elements of moral hazard and asymmetric information. That an observable performance blends the impact of effort and an innate parameter is a familiar modeling trick of the scenting literature that goes back to Mirrlees (1971)'s seminal model of optimal income taxation and Laffont and Tirole (1986)'s well known model of incentive regulation. Laffont and Martimort (2002, Chapter 7) offers a typology of those and other related models.

[5] Section 5.2 briefly addresses the polar case of an agent who is more optimistic than the principal.

[6] Contracting problems between a principal and an agent with different risk perceptions have been studied in other contexts by De la Rosa (2011), Jeleva and Villeneuve (2004), Villeneuve (2005), Eliaz and Spiegler (2008), Spinnewijn (2013) and Grubb (2009).

[7] Section 5.3 briefly addresses an extension of our modeling where risk aversion is more explicit.

[8] See Guriev (2001) for an interesting model along these lines.

[9] Our analysis will illustrate that, in many respects, this multi-dimensional screening problem is much easier to handle than in the standard screening framework developed for multiproduct monopolists by Rochet and Choné (1998) among others.

the principal could counter the agent's excessive incentives for information gathering and reduce this premium by keeping a constant share of output for himself. Choosing such a fixed sharing rule would make the *ex post* screening problem trivial although it would also induce low-powered incentives at the operational stage. In fact, the principal can improve on this outcome by offering a more complete menu of screening contracts.

When neither information collection nor productivity shocks are common knowledge, incentives to gather information are directly linked to the distributions of type-dependent bonuses that the principal is offering to screen possible realizations of the shock (Lemma 1). The dichotomy between *ex ante* and *ex post* incentives no longer holds. In particular, the agent's incentives to gather information depend on the convexity of the payoffs profile he withdraws from possibly learning shock realizations. To reduce the excessive incentives of a pessimistic agent to gather information and implement a "flatter" payoffs profile, the principal would like to reduce bonuses on the upper tail of the shocks distribution and increase bonuses on the lower tail.

Of course, the optimal bonuses are not only driven by distortions targeted to reduce *ex ante* incentives for information gathering but also by distortions that are needed to solve the screening problem *ex post*. When informed, the agent has indeed incentives to pretend that the productivity shock is less favorable to save on his repayment to the principal. Preventing that behavior calls for making less attractive the contract for lower realizations of the shock which means reducing bonuses and moving towards low-powered incentives as shocks are less favorable. Following *"good news,"* i.e., for shocks which are sufficiently favorable, *ex post* and *ex ante* distortions go in the same direction. The optimal menu of contracts entails linear contracts stipulating bonuses that are positively correlated with productivity shocks. The agent communicates whatever information has been learned and incentives at the operating stage are rather high powered. Instead, *ex post* and *ex ante* distortions go in opposite directions following *"bad news"*. Screening distortions still call for lower bonuses while implementing a flatter payoffs profile also requires to increase bonuses on the lower tail. Bunching arises as a result of these conflicting forces.[10] The same contract with rather low-powered incentives is offered whether the agent has learned *"bad news"* or remains uninformed (Theorem 1). One important take-away of our analysis is thus that the agent's pessimism calls for relatively flat menus of options which reduce the value of getting information.

Literature review. There is by now a sizable literature which endogenizes information structures in principal–agent models and social choice environments.[11] Earlier contributions have analyzed the principal's preferences over the agent's amount of private information but neglected his incentives to acquire information (Lewis and Sappington, 1991, 1993; Sobel, 1993). Subsequent research has instead pushed incentives on the forefront of the analysis. How much information is gathered at equilibrium, whether information is valuable *per se* or just a pure rent-seeking activity, and the shape of the optimal contract all depend on fine details of the modeling. This sensitivity makes definitive lessons on whether endogenizing information structures in agency problems brings something new to our tools kit at best unsettled. The timing for information gathering and contracting (Cremer and Khalil, 1992; Cremer et al., 1998a; Compte and Jehiel, 2008 and Terstiege, 2012), the modeling of information as a continuous or a discrete variable (Cremer and Khalil, 1994; Kessler, 1998), the amount of competition between agents (Cremer and Khalil, 1992; Compte and Jehiel, 2008), and whether information gathering refers to the outcome of the agent's effort (Iossa and Legros, 2004) are all ingredients that matter a lot.

Two papers are closest to ours. The first one is Cremer et al. (1998b) who extend the seminal model of Baron and Myerson (1982) by allowing an agent to gather information on his cost parameter after the principal's offer but before contract signing. An informed agent then chooses within a whole menu of screening options. Instead, the contract stipulates just a single output–payment pair if the agent remains uninformed. There is no residual uncertainty in information gathering; either the agent learns perfectly his type or he does not, and each possibility may be optimal depending on parameters.

We depart from this model along several lines. First, we investigate a timing where contracting takes place *ex ante*, i.e., before any uncertainty gets resolved.[12] Cremer et al. (1998b) instead introduces an *ex post* participation constraint that applies to an informed agent while they do not impose such constraint when the agent remains uninformed even though shocks on costs could drive profit below zero. We keep a more consistent specification of the participation constraints in a framework with *ex ante* contracting and focus on differences in belief as a source of contracting frictions.

Second, information gathering remains uncertain in this paper; there is always a non-zero probability that the agent remains uninformed unless the agent exerts no effort at all in gathering information. From an economic viewpoint, this assumption ensures that information gathering leads to a nontrivial moral hazard problem. From a technical viewpoint, this assumption avoids the cumbersome discussion of the various scenarios where the principal finds it optimal or not to induce information gathering. We indeed characterize the agent's effort in information gathering by means of a more compact (necessary and sufficient) first-order approach.

---

[10] See Lewis and Sappington (1989) for a seminal paper on countervailing incentives.

[11] Bergemann and Välimäki (2006) provide an exhaustive survey. Beyond models involving a single agent which are close to ours, several authors have analyzed information gathering in multi-agent contexts (Bergemann and Välimäki, 2002; Cremer et al., 2009; Gerardi and Yariv, 2008; Gershkov and Szentes, 2009; Shi, 2012).

[12] In our view, such timing better captures the contracting environments that motivated our study, especially in procurement contexts. For instance, Public–Private Partnerships are actually long-term projects whose contours are designed much before any uncertainty on future demand and cost parameters is released and thus much before private contractors may learn relevant information on those parameters.

Third, Cremer et al. (1998b) cast their analysis in the framework of Baron and Myerson (1982)'s seminal model. There, costs are supposed to be non-observable and output is the sole contracting variable. It follows from incentive compatibility that the agent gets the same the payment and output requirements when uninformed and when he has learned the mean of the distribution of marginal costs. Yet, in the first case realized costs may fluctuate a lot depending on the realized shock while in the second scenario, those costs remain unchanged. Introducing a little bit of *ex post* costs observability and conditioning payments on realized costs could thus significantly ease incentive compatibility constraints. In this paper, we take another path. We assume that output although contractible blends together the impacts of the agent's effort at the operating stage and the productivity shock. When informed, the agent has perfect knowledge of the mapping between his operating effort, output and compensation. He chooses accordingly within a menu of screening options. When uninformed, the agent still has freedom in choosing his operating effort, but he ignores how that effort will be mapped into final outputs and rewards that may fluctuate with shocks realizations.

The second important paper we build on is Szalay (2009). This author provides a rich framework to model uncertain information gathering. Incentive constraints are captured by means of a first-order approach. Yet, the timing of the contractual game remains similar to Cremer et al. (1998b) and thus differs from ours. Szalay (2009) highlights that high-powered contracts make the agent's informational rents more risky which boosts information acquisition,[13] while non-responsive contracts are used to reduce incentives for information gathering. We differ from Szalay (2009) by restricting the class of information technologies (the agent gets either full information or not at all) but this restriction is already rich enough to allow us to analyze a new agency cost of information gathering due to different prior beliefs. This framework allows us to derive an optimal contract that features at the same time both high- and low-powered incentives depending on whether the agent has gathered *"good", "bad"* or *"no news."*

Finally, our concerns on how *ex ante* incentives in information gathering and *ex post* incentives in truthtelling revealing such information is also shared by a burgeoning literature that analyzes the optimal organization of experts in specific contexts (Lambert, 1986; Demski and Sappington, 1987; Gromb and Martimort, 2007; Malcomson, 2009; Szalay, 2005, and Dai et al., 2006, among others). Experts must be incentivized to gather information and to truthful reveal their findings whenever such information is manipulable. Those papers investigate how those two agency problems interact even if those experts might not necessarily operate themselves the projects they recommend.

Organization of the paper. Section 2 describes the model. Section 3 presents some useful benchmarks. Section 3.1 discusses the scenario where the principal gathers information by himself. Section 3.2 studies the case where productivity shocks are observable so that the sole agency problem is at the *ex ante* stage when information is gathered. Section 4 moves to the more complex case where neither information gathering nor its outcome can be observed by the principal. Section 4.1 describes the set of incentive feasible allocations under that scenario. Section 4.2 characterizes the optimal contract and provides a number of comparative statics results. Section 5 discusses some possible extensions, focusing on the value of communication, the case of optimistic agents and risk aversion. Section 6 briefly concludes. Proofs are relegated to Appendix A.

2. The model

Consider a principal (thereafter *she*) who hires an agent (or *he*) to gather information on a productivity shock and then decide on an operating effort according to information he may have learned. To illustrate, in a public procurement context, a public authority delegates to a firm the task of producing a service and gathering information on demand for that service. Such information is relevant to determine the level of service, its quality or its volume. Yet, our methodology and findings are sufficiently general to allow broader interpretations. They would as well apply to other principal–agent relationships like shareholders–CEOs, manufacturer–retailers, clients–lawyers, lender–borrowers, and so on.

Technology. The project yields an observable return y which depends both on the realization of a productivity shock θ and the agent's operating effort e. For simplicity, we postulate the following multiplicative specification of the production function:

$$y = (1+\theta)e.$$

A lower (resp. higher) shock θ reduces (resp. increases) the marginal return on effort.

Ex post asymmetric information. When successful in gathering information, the agent privately learns the true value of the productivity shock θ. This shock is drawn according to the common knowledge cumulative distribution $F(\theta)$ with an

[13] Also related is Baker and Jorgensen (2003). Those authors analyze a moral hazard model à la Holmström and Milgrom (1991) with an agent having a CARA utility function, but whose measured performances depend both on some additive noise and a multiplicative shock (coined as "volatility") that affects the returns on his effort and which is *a priori* known by the agent. Because of a lack of tractability of the CARA model to handle the *ex post* asymmetric information, the analysis is restricted to the case where a single linear contract is offered. Baker and Jorgensen (2003) demonstrate that high-powered incentives are preferred in more volatile settings.

atomless and everywhere positive density $f(\theta)$ on the support $\Theta = [-\delta, \delta]$ (with $\delta \leq 1$ to maintain positive outputs under all circumstances). The density function $f(\theta)$ is symmetric and centered around zero so that $E_\theta(\theta) = 0$ where $E_\theta(\cdot)$ denotes the expectation operator. Let $\sigma^2 = E_\theta(\theta^2)$ also denote the variance of this shock.

To ensure monotonicity properties which are familiar for screening models, we also assume that $S(\theta) = \frac{F(\theta)}{f(\theta)(1+\theta)}$ and $R(\theta) = \frac{1-F(\theta)}{f(\theta)(1+\theta)}$ satisfy the following conditions:

Assumption 1.

$$R'(\theta) \leq 0 \leq S'(\theta) \quad \forall \theta \in \Theta.$$

Ex post moral hazard. The operating effort is non-observable. Whether the agent gets informed or not, he can still freely choose effort at the operating stage. The contract is thus subject to *ex post moral hazard*. The cost of exerting an operating effort e is $\psi(e) = \frac{e^2}{2\lambda}$ where $\lambda > 0$.[14] When the technology is more flexible, i.e., for greater values of λ, the agent's effort becomes more sensitive to what he may have learned on productivity shocks.

Ex ante moral hazard. Effort in information gathering is also non-observable. At cost $\varphi(a)$, the agent gathers information on the productivity shock θ with probability a. With probability $1 - a$, the agent remains uninformed. $\varphi(a)$ is increasing, convex, has non-negative third-derivative and satisfies the usual Inada conditions ($\varphi'(a) \geq 0$, $\varphi''(a) > 0$, $\varphi'''(a) \geq 0$, $\varphi'(0) = 0$ and $\varphi'(1) = +\infty$). These assumptions ensure interior solutions for all problems below, so that both regimes with or without information gathering have always positive probabilities.[15]

Preferences. The principal pockets the project return $y = (1 + \theta)e$ and pays back a monetary transfer t to the agent for his services. The principal's payoff is quasi-linear in (y, t):

$$W(y,t) = y - t.$$

The agent's (*ex post*) payoff net of his costs at the operating and the information gathering stages simply writes as:

$$U(t,e,a) = t - \psi(e) - \varphi(a).$$

Beliefs. The agent is pessimistic over the outcome of information gathering although he shares with the principal the same beliefs $F(\cdot)$ over the realization of the productivity shock. Formally, while the principal believes that information gathering succeeds with probability a, the agent has subjective beliefs that this probability is only ρa where $\rho \leq 1$.[16] That beliefs on the success of information gathering are subjective is best justified by the huge uncertainty surrounding that activity, which leaves much freedom to both parties in assessing the likelihood of its success. Instead, parties agree on their beliefs over the realization of the productivity shock because there is a well-known technology that generates those shocks.

Contracts. Contracts are signed before any information gathering takes place. They must be designed to align the principal's and the agent's objectives even if their beliefs might differ.

Contracts are linear in the realized outcome y and thus write as $t(y) = \alpha y - \beta$. The fixed repayment β is paid by the agent to access the production process. A piece rate bonus αy determines the share of production that the agent can keep for himself.[17]

A menu of such linear contracts helps screening the agent according to his information. Indeed, the agent first knows whether information gathering has been successful or not. Second, if successful at this earlier stage, the agent also learns the value of the productivity shock. A direct revelation mechanism must thus induce revelation of both pieces of information.[18] In other words, the agent can lie on the productivity shock when knowing it but he may also pretend having observed a given productivity shock even if he has not (and *vice versa*). From the Revelation Principle,[19] there is thus no loss of generality in having the principal offer menu of linear schedules of the form $\{(t(y,\hat{\theta}))_{\hat{\theta}\in\Theta}, t_u(y)\}$ where $\hat{\theta}$ is the agent's

[14] This quadratic specification is used for tractability. It simplifies the expression of some of our results below without any loss of economic insights.

[15] This assumption distinguishes our analysis from other models in the literature as, for instance, Cremer et al. (1998a) and (1998b) who view information gathering as having a deterministic outcome.

[16] This difference in beliefs introduces a friction in contracting that, although it is akin to risk aversion, preserves the linearity of payoffs. Because of the various sorts of frictions (*ex ante* and *ex post* moral hazard, asymmetric information) considered, the model would lose tractability had we assumed a concave Bernoulli utility function. The impact of linearity will become more explicit below when computing expected payoffs.

[17] The incentive properties of linear contracts have been studied, among others, by Laffont and Tirole (1986), Rogerson (1987, 2003) in the field of regulation and Melumad et al. (1992) and Mookherjee and Reichelstein (2001) when applied to the internal organization of the firm. In pure moral hazard environments, Holmström and Milgrom (1987, 1991) and Baker (1992) unveil conditions under which such schemes are optimal.

[18] Although there are no true dynamics in our framework, one could envision the menu of contracts as first asking the agent whether he is informed or not, and second which productivity shock he has observed if he is informed. This dynamic interpretation of the model bears some resemblance with dynamic mechanism design problems investigated by Courty and Li (2000) and Krähmer and Strausz (2011a).

[19] Myerson (1982).

announcement on the realized shock that he may have learned. When informed on the productivity shock θ, incentive compatibility implies that the agent optimally picks the scheme corresponding to the observed productivity shock $t(y,\theta) = \alpha(\theta)y - \beta(\theta)$. The agent then delivers an output $y(\theta)$ by accordingly adjusting his effort at the operating stage and he receives the corresponding payment $t(y(\theta),\theta)$. When instead uninformed, the agent optimally chooses the scheme $t_u(y) = \alpha_u y - \beta_u$, exerts an operating effort without knowing the productivity shock and gets a payment $t_u(y)$ when output y realizes.

3. Benchmarks

In this section, we establish a couple of benchmarks that allow us to better understand contracting frictions in more complex scenarios. In passing, we also define a few useful notions.

3.1. Information gathering by the principal

Suppose that the principal first gathers information by himself and then chooses accordingly the operating effort. Once he has learned that the productivity shock is θ, the principal chooses an operating effort $e^{fb}(\theta)$ that trades off its marginal cost with its marginal return:

$$\psi'(e^{fb}(\theta)) = 1 + \theta \Leftrightarrow e^{fb}(\theta) = \lambda(1 + \theta).$$

A better productivity shock increases the marginal benefit of effort. The optimal levels of effort and output $y^{fb}(\theta) = \lambda(1+\theta)^2$ are thus increasing with θ. Also, a more flexible technology is associated with a lower marginal cost of effort and induces greater effort and output.

If the principal does not gather information, he chooses his effort at the operating stage without knowing the exact realization of the productivity shock and its consequences on output. The corresponding (uninformed) level of effort e_u^{fb} is then:

$$\psi'(e_u^{fb}) = E_\theta(1 + \theta) \Leftrightarrow e_u^{fb} = e^{fb}(0) = \lambda. \tag{1}$$

Because shocks have zero mean, this optimal effort is exactly the same as if the principal had learned that the realized productivity shock is the mean of the distribution.

We can now characterize the optimal first-stage effort in information gathering, say a^{fb}, as a solution to the following problem:

$$a^{fb} \equiv \arg\max_a a E_\theta\big((1+\theta)e^{fb}(\theta) - \psi(e^{fb}(\theta))\big) + (1-a)\big(e^{fb}(0) - \psi(e^{fb}(0))\big) - \varphi(a).$$

The following first-order condition is both necessary and sufficient for optimality:

$$\varphi'(a^{fb}) = \frac{\lambda\sigma^2}{2}. \tag{2}$$

To understand this expression, it is useful to define the (positive) value of information in the absence of any agency problem as the difference in the (expected) operating surpluses when information is learned and when it is not:

$$\mathcal{V}^{fb} \equiv E_\theta\big((1+\theta)e^{fb}(\theta) - \psi(e^{fb}(\theta))\big) - E_\theta\big((1+\theta)e^{fb}(0) - \psi(e^{fb}(0))\big) = \frac{\lambda\sigma^2}{2}.$$

The optimality condition (2) just means that the marginal cost of effort in information gathering is equal to the value of information for the principal. More uncertainty on productivity shocks (i.e., σ^2 larger) increases the value of information and boosts information gathering.

Observable ex ante effort. Before moving to scenarios where agency problems matter, observe that the same outcome than described above can be achieved if the agent is delegated the decision to gather information but his effort at this stage and the information collected are both verifiable. This outcome can be readily achieved by means of a forcing contract stipulating the requested first-best effort level.

3.2. Information gathering with verifiable shocks

Suppose now that the agent gathers information, but that the productivity shock θ can be verified so that there is no agency problem *ex post*. Intuition suggests that the principal can certainly align her objectives with those of the agent at the operating stage under those circumstances by means of fixed-price contracts. Yet, providing *ex ante* incentives in information gathering remains an issue when the agent's effort at that stage is non-verifiable. This benchmark thus allows us to focus on the sole consequences of *ex ante* moral hazard in information gathering.

Ex ante participation and incentive constraints. Because θ is verifiable, the agent operates under the scheme $t(y,\theta) = \alpha(\theta)y - \beta(\theta)$ when the productivity shock θ has been learned and $t_u(y) = \alpha_u y - \beta_u$ when it has not. Keeping our previous notation, we denote the agent's *ex post* payoffs under those scenarios respectively as:

$$U(\theta) \equiv \max_e \alpha(\theta)(1+\theta)e - \frac{e^2}{2\lambda} - \beta(\theta) = \frac{\lambda}{2}\alpha^2(\theta)(1+\theta)^2 - \beta(\theta), \tag{3}$$

$$U_u \equiv \max_e \alpha_u E_\theta((1+\theta)e) - \frac{e^2}{2\lambda} - \beta_u = \frac{\lambda}{2}\alpha_u^2 - \beta_u. \tag{4}$$

Remember that the agent is pessimistic and overweights the probability of not gathering information. The agent participates when his *ex ante* payoff remains non-negative:

$$\rho a E_\theta(U(\theta)) + (1-\rho a)U_u - \varphi(a) \geq 0. \tag{5}$$

The left-hand side above is strictly concave in a and the following incentive constraint characterizes the (interior) first-stage effort:

$$\frac{1}{\rho}\varphi'(a) = E_\theta(U(\theta)) - U_u. \tag{6}$$

The right-hand side of (6) is again the value of information for the agent. It is clearly non-negative because the principal could just offer the same scheme $t(y,\theta) \equiv t_u(y)$ under all circumstances. Inducing a positive *ex ante* effort from the agent requires to leave him a greater payoff when he acquires information: $E_\theta(U(\theta)) > U_u$.

The principal's problem. To understand the principal's problem let us decompose the agent's *ex ante* payoff as follows:

$$\rho a E_\theta(U(\theta)) + (1-\rho a)U_u - \varphi(a) = \mathcal{U} - \mathcal{R}$$

where

$$\mathcal{U} \equiv a E_\theta(U(\theta)) + (1-a)U_u - \varphi(a) \text{ and } \mathcal{R} \equiv a(1-\rho)(E_\theta(U(\theta)) - U_u). \tag{7}$$

The first term $\mathcal{U}$ is the agent's expected payoff had he shared with the principal the same beliefs over the success of information gathering. Otherwise, this payoff accounts for a *belief premium* $\mathcal{R}$ that measures how the principal and the agent's objectives are misaligned. This *belief premium* is indeed an extra payment that the principal must give to a pessimistic agent to induce his participation.[20] Inducing information gathering from a pessimistic agent thus entails an *agency cost*. Inserting the expression of effort coming from the incentive constraint (6) and making its dependence on a explicit, we may rewrite this agency cost $\mathcal{R}(a)$ as:

$$\mathcal{R}(a) = a\varphi'(a)\left(\frac{1}{\rho} - 1\right). \tag{8}$$

Observe that increasing the agent's information gathering effort raises this agency cost.

Importantly, $\mathcal{R}(a)$ is independent of effort at the operating stage. This highlights a dichotomy between the moral hazard incentive problems at the operating and the information gathering stages. Yet, with verifiable productivity shocks, the only source of agency cost comes from the *ex ante* stage, making the agent residual claimant for effort provision at the operating stage is thus clearly optimal:

$$\hat{\alpha}_u = \hat{\alpha}(\theta) = 1 \quad \forall \theta \in \Theta. \tag{9}$$

Such fixed-price contracts align the principal's and the agent's objectives and induce the efficient effort at the operating stage:

$$\hat{e}_u = e^{fb}(0) \text{ and } \hat{e}(\theta) = e^{fb}(\theta) \quad \forall \theta \in \Theta. \tag{10}$$

Incentives to gather information are solely provided by means of different reimbursements depending on whether information has been gathered or not. Next proposition summarizes the design of the optimal contract.

Proposition 1. *Assume that productivity shocks are verifiable but effort in information gathering is not. A menu of fixed-price contracts, $\hat{t}(y,\theta) \equiv \hat{t}_i(y) = y - \hat{\beta}$ and $\hat{t}_u(y,\theta) = y - \hat{\beta}_u$ with $\hat{\beta}_u < \hat{\beta}$, is optimal. Moreover, the effort in information gathering, although positive, is lower than its first-best level, $0 < \hat{a} < a^{fb}$ with:*

$$\varphi'(\hat{a}) + \mathcal{R}'(\hat{a}) = \frac{\lambda\sigma^2}{2}. \tag{11}$$

[20] From (6), $\mathcal{R}$ is non-negative when $\rho < 1$.

The agency cost $\mathcal{R}(a)$ is reduced by charging a lower fixed fee when the agent is uninformed ($\hat{\beta}_u < \hat{\beta}$), since it lowers the effort in information gathering below the first-best level. This distortion is more pronounced when the agent is more pessimistic (ρ smaller).

Had the principal not set a lower fee when the agent is uninformed, the agent would choose an effort below the first-best level a^{fb} but still too high. Indeed, with $\hat{\alpha}_u = \hat{\alpha}(\theta) = 1$ and $\hat{\beta}_u = \hat{\beta}(\theta)$ for all θ, the agent is made residual claimant both for gathering information gathering and operating assets. His payoffs whether informed or not would respectively be given by:

$$U(\theta) \equiv \max_e (1+\theta)e - \psi(e) - \beta_u = \frac{\lambda}{2}(1+\theta)^2 - \beta_u \text{ and } U_u \equiv \max_e E_\theta(1+\theta)e - \psi(e) - \beta_u = \frac{\lambda}{2} - \beta_u.$$

Therefore, the agent's *ex ante* payoff would be:

$$\rho a E_\theta(U(\theta)) + (1-\rho a)U_u - \varphi(a) = \frac{\lambda}{2}(1+\rho a\sigma^2) - \varphi(a) - \beta_u.$$

This would lead to a level of effort a_0 such that:

$$\frac{1}{\rho}\varphi'(a_0) = \frac{\lambda\sigma^2}{2} = \varphi'(a^{fb}) \Rightarrow a_0 < a^{fb}. \tag{12}$$

Because the agent is pessimistic (i.e., $\rho \in (0,1)$) his effort falls below the first best, $a_0 < a^{fb}$. More importantly, from (11) and (12), this effort remains greater than the optimal level that the principal wishes to induce, i.e., $\hat{a} < a_0$. Intuitively, although both the principal and the agent wants to reduce effort a to decrease the premium $\mathcal{R}(a)$, they want to do so with different intensities. The pessimistic agent wants to gather more information than what the principal would like in a second-best world. This agent has less incentives to reduce this premium by himself because the principal ends up paying for that premium anyway to ensure his participation. To induce such lower effort $\hat{a} < a_0$, the principal has then to ask for a higher fee when the agent is informed, making information gathering less attractive.

Verifiability of information gathering/non-observability of shocks. Let us now consider a slightly more complex information structure where the mere fact that information has been gathered can be verified, although the agent privately observes the true realization of the shock. The solution described in Proposition 1 is still valid under that scenario. Indeed, the optimal fixed-price contracts do not depend on the precise value of the shock. Those contracts could as well have been offered had this shock been privately known.

4. Non-observable information gathering

We now turn to the more complex and more interesting scenario where the principal ignores both whether information has been gathered or not, and the possible realization of that shock. Clearly, the solution found in Proposition 1 is no longer feasible. Because $\hat{\beta}_u < \hat{\beta}(\theta)$, the agent would always report being uninformed so as to minimize his fee, and this would destroy his incentives to gather information. Raising the fee when the agent is uninformed, so as to have $\hat{\beta}_u = \hat{\beta}(\theta)$, would induce truthful revelation but then, as we have seen, the level of effort a_0 chosen by the pessimistic agent would be excessive.

The principal can obtain a higher payoff by modifying not only fees but also bonuses in order to induce information gathering and revelation. This rough intuition already suggests that the dichotomy between *ex post* and *ex ante* incentives will now disappear.

4.1. Incentive feasible allocations

We first describe the set of incentive feasible allocations when all contracting frictions (*ex ante* moral hazard on a, *ex post* moral hazard on e and private information on θ) are at play.

Ex post private information. The agent must be induced to reveal what he has learned. Let again denote by $U(\theta)$ the information rent of an agent who now has private information on the shock θ. Incentive constraints for this aspect of the contracting problem amounts to:

$$U(\theta) \equiv \max_{(\hat{\theta},e)} \alpha(\hat{\theta})(1+\theta)e - \psi(e) - \beta(\hat{\theta}). \tag{13}$$

This expression already encompasses all truthtelling constraints preventing an informed agent to misrepresent the productivity shock. Following earlier models in the screening literature,[21] output aggregates both the agent's effort and his private

[21] See Mirrlees (1971), Laffont and Tirole (1986) and Laffont and Martimort (2002, Chapter 7) for an overview of models blending together adverse selection and moral hazard elements.

information. When truthtelling constraints hold, the agent picks within the schedule $t(y,\theta)$ his most preferred output $y(\theta) = (1+\theta)e(\theta)$ with an operating effort that reflects actual bonus:

$$e(\theta) = \lambda(1+\theta)\alpha(\theta) \quad \forall\theta \in \Theta. \tag{14}$$

We immediately get the standard characterization of the rent profile $U(\theta)$.

Lemma 1. *$U(\theta)$ is absolutely continuous, non-decreasing and convex with*

$$U(\theta) = U(0) + \lambda \int_0^\theta (1+x)\alpha^2(x)dx \quad \forall\theta \in \Theta. \tag{15}$$

The bonus $\alpha(\theta)$ is non-decreasing in θ.

The intuition behind this Lemma can be best seen by writing (15) in its differential form (which is valid almost everywhere) as:

$$U'(\theta) = \lambda(1+\theta)\alpha^2(\theta) \quad \forall\theta \in \Theta. \tag{16}$$

When he has observed a shock θ, the agent may be tempted to claim that the shock is only $\theta - d\theta$, save on the *ex post* operating effort and pay accordingly a lower fee to the principal. This gain is worth $\lambda(1+\theta)\alpha^2(\theta-d\theta)d\theta \approx \lambda(1+\theta)\alpha^2(\theta)d\theta$. To avoid such manipulation, type θ must receive an extra rent worth $U(\theta) - U(\theta - d\theta) \approx U'(\theta)d\theta$ precisely equal to this gain.

Beyond those incentives to manipulate information when learned, two other important incentive constraints must be added. First, an informed agent must also be prevented from pretending being uninformed and choosing an output y along the scheme $t_u(\cdot)$. The following incentive compatibility constraint must thus also hold:

$$U(\theta) \geq V(\theta) \equiv \max_e \alpha_u(1+\theta)e - \beta_u - \psi(e) \quad \forall\theta \in \Theta.^{22} \tag{17}$$

Second, a uninformed agent must prefer to pick the scheme $t_u(\cdot)$ and choose his operating effort not knowing the productivity shock θ rather than pretending knowing that a shock $\hat{\theta}$ has realized:

$$U_u = \max_e \alpha_u e E_\theta(1+\theta) - \beta_u - \psi(e) \equiv V(0) \geq \max_{(e,\hat{\theta})} \alpha(\hat{\theta}) e E_\theta(1+\theta) - \beta(\hat{\theta}) - \psi(e) \equiv U(0). \tag{18}$$

The right-hand side above simply comes from observing that the average shock θ is zero. The uninformed agent can always behave as being hit by such mean value shock while reciprocally, an agent having learned a shock with mean value can always pretend being uninformed (i.e., (17) holds). This remark immediately implies:

$$U(0) = U_u. \tag{19}$$

In other words, the agent gets the same payoff whether uninformed or informed on the mean value of the shock. This important property turns out to be useful to simplify the expression of the *ex ante* moral hazard incentive constraint. Turning now to the information gathering stage, using the characterization of the rent profile $U(\theta)$ obtained from Lemma 1, integrating by parts and taking into account (19), we indeed get:

Lemma 2. *The information gathering incentive constraint (6) can be rewritten as:*

$$\frac{1}{\rho}\varphi'(a) = \lambda E_\theta\left(\frac{1_{\theta>0} - F(\theta)}{f(\theta)}(1+\theta)\alpha^2(\theta)\right) \tag{20}$$

where $1_{\theta>0} = \begin{cases} 1 \text{ if } \theta>0 \\ 0 \text{ otherwise} \end{cases}$.

The incentive constraint (20) has important implications for the shape of the optimal contract. When information remains private to the agent, his incentives to gather information are driven by his payoffs at the operating stage. Eq. (20) can be interpreted as earlier. The left-hand side is the agent's marginal cost of effort conveniently modified to take into account his pessimism. The right-hand side is again the value of information from the agent's viewpoint. This quantity depends now on the whole profile of bonuses at the operating stage. It encapsulates the fact that the profile of *ex post* payoffs also satisfies truthtelling constraints.

[22] Note that the reservation payoff $V(\theta)$ is itself an implementable profile corresponding to a fixed bonus α_u. As such, it satisfies Lemma 1 and thus:

$$V(\theta) = V(0) + \frac{\lambda\alpha_u^2}{2}\theta(2+\theta) \quad \forall\theta \in \Theta.$$

4.2. Optimal contracts

The non-verifiability of productivity shocks puts lots of structure on the optimization problem. First, truthtelling constraints imply that $\alpha(\theta)$ must remain non-decreasing as reported in Lemma 1. Second, effort at the information gathering stage now also depends on bonuses at the operating stage through the *ex ante* incentive constraint (20). Indeed, first- and second-stage efforts are now linked. There is no longer any dichotomy between *ex post* and *ex ante* incentives. Bonuses must be modified to boost information gathering.

Theorem 1 characterizes optimal bonuses and first stage effort in information gathering at the optimal contract. This characterization shows that offering a menu of screening bonuses on the upper tail of the distribution is always optimal.

To this end, denote by $\Omega(a)$ the (positive)[23] implicit solution in $(0, \delta)$ to the following equation in θ^*[24]:

$$R(\theta^*) = \frac{\frac{a}{1-a}\int_{-\delta}^{\theta^*}(1_{\theta>0} - F(\theta))(1+\theta)d\theta}{1 + \frac{a}{1-a}\int_{-\delta}^{\theta^*}(1+\theta)^2 f(\theta)d\theta}. \tag{21}$$

Theorem 1. *The optimal contract with a pessimistic agent entails the following properties.*

1. *For $\theta \in [-\delta, \Omega(a^*)]$, the incentive bonus is independent of θ and identical whether the agent gets such "bad news" or remains uninformed:*

$$\alpha^*(\theta) = \alpha_u^* < 1 \quad \forall \theta \in \left[-\delta, \Omega(a^*)\right].$$

2. *For $\theta \in [\Omega(a^*), \delta]$, the incentive bonus $\alpha^*(\theta)$ is increasing with θ and greater than when the agent remains uninformed:*

$$\alpha_u^* \le \alpha^*(\theta) = \frac{\alpha_u^*}{\alpha_u^* + (1-\alpha_u^*)\frac{R(\theta)}{R(\Omega(a^*))}} \le 1 \quad \theta \in \left[\Omega(a^*), \delta\right] \tag{22}$$

with an equality on the r.h.s at $\bar{\theta}$ only and on the l.h.s at $\Omega(a^*)$ only.

3. *The optimal information gathering effort a^* is lower than what the agent would choose on his own and thus also lowers than the first best:*

$$a^* < a_0 < a^{fb}.$$

Screening for "good news." A menu of linear contracts allows to screen the agent according to the productivity shock he has learned. Of course, this does not come as a surprise, the principal can certainly not lose by offering different options. What is more interesting is precisely when and how those screening possibilities are used. Indeed, the optimal scheme entails bunching on a lower tail $[-\delta, \Omega(a^*)]$ of the distribution while screening only arises for sufficiently "good news," i.e., on an interval $[\Omega(a^*), \delta]$. On the bunching area, the agent gets the same bonus whether he learns *"bad news"* or remains uninformed.

To understand the shape of the optimal contract, it is important to come back on the particular form taken by the *ex ante* incentive constraint (20). To build some preliminary intuition, suppose first that the principal offers a single linear scheme which applies whether the agent gets informed or not, i.e., $t_i(y, \theta) = t_u(y) \equiv \alpha y - \beta$ for all $\theta \in \Theta$ for some (α, β). From Section 3.2, we already know that such scheme cannot be optimal when the productivity shock is verifiable; fixed-price contracts have to specify different fees depending on whether information has been learned or not as a means to better align the principal and the agent's incentives to collect information. Yet, a single linear scheme nevertheless solves the screening problem *ex post* in an obvious manner, a property which could be attractive in itself. Turning now to the agent's *ex ante* incentives, we may first integrate by parts on $[-\delta, 0]$ and $[0, \delta]$ respectively to get $E_\theta\left(\frac{1_{\theta>0} - F(\theta)}{f(\theta)}(1+\theta)\right) = \frac{\sigma^2}{2}$. Using this expression, we may thus rewrite the incentive constraint (20) in the case of a linear contract with slope α as:

$$\frac{1}{\rho}\varphi'(a) = \frac{\lambda}{2}\sigma^2\alpha^2. \tag{23}$$

Moving away from a fixed-price contract by reducing the bonus below 1 certainly reduces the agent's value of information and thus dampens his incentives to gather information. It thus better aligns the agent's and the principal's objectives.

Of course, the principal can still improve on this outcome by choosing a type-dependent bonus (accompanied with a type-dependent fee to respect truthtelling constraints). To see how, we may indeed rewrite the right-hand side of (20) as:

$$\lambda\left(\int_0^\delta (1-F(\theta))(1+\theta)\alpha^2(\theta)d\theta - \int_{-\delta}^0 F(\theta)(1+\theta)\alpha^2(\theta)d\theta\right).$$

[23] See Appendix A for a proof.

[24] We prove in Appendix A that this solution is indeed unique.

Starting from a fixed-price contract (that would be optimal had shocks been verifiable) and reducing the bonus below one following *"good news"* (i.e. $\theta \geq 0$) while increasing it following *"bad news"* (i.e. $\theta \leq 0$) would reduce the agent's effort in information gathering and move it closer to the principal's optimal choice.

On the upper tail, such downward distortions have also good properties from a screening point of view. Indeed, remember that when he has observed a shock θ, the agent may be tempted to claim that the shock is only $\theta - d\theta$, save on the *ex post* operating effort and pay accordingly a lower fee to the principal. Such incentives to misrepresent the state of nature downwards are better controlled by making the allocation less attractive in that state which requires reducing $\alpha(\theta)$. In other words, the *ex ante* and the *ex post* incentive problems both require distortions which go in the same direction. As a result, there is still the possibility to screen the agent according to the realization of the productivity shocks when those shocks are *"good news."*

Bunching for "bad news." On the lower tail of the shock distribution, increasing the bonus to boost information gathering is now in conflict with the downward distortion requested for screening purposes. The best way to solve this conflict is to have some bunching when shocks are sufficiently *"bad news"*.

Importantly, this bunching area never covers the whole interval. Indeed, there are always gains of screening shocks in the neighborhood of $\theta = \delta$. Indeed, screening distortions are not necessary at the upper bound of the shocks distribution, a familiar *"no distortion at the top"* result from the screening literature that results in a fixed-price contract being offered for $\theta = \delta$. The agent who announces having observed the most productive shock is thus made residual claimant for his performances. Less attractive reports are followed by lower bonuses. From there, the bonus profile defined in (22) actually continuously goes from 1 to reach α_u^* as shocks become less favorable.

The boundary of the bunching area, say $\Omega(a^*)$ is defined through condition (21). The economic intuition behind this definition is the following. The left-hand side is the familiar hazard rate that applies to characterize screening distortions on the upper tail of the shocks distribution. The right-hand side is also a measure of those distortions but it now applies over the whole bunching area and it is weighted by the likelihood ratio $\frac{a}{1-a}$ that information gathering was successful. The boundary of the bunching area is thus defined so that screening distortions are the same on the separating part of the optimal contract (the upper tail) and on the bunching part (the lower tail). The optimal bonus is continuous at that boundary.

When the ex ante effort is very costly, the optimal effort a^* goes to zero, and the bunching area covers almost the whole types set. Intuitively, the principal does not lose much from offering a single linear scheme targeted to an uninformed agent under those circumstances. *A contrario*, a less costly technology for information gathering increases a^* and reduces the bunching area since $\Omega(\cdot)$ is decreasing.

Fixed fees. The first-order necessary (and here sufficient) conditions for optimality at the revelation stage imply that the fixed fee $\beta^*(\theta)$ solves the following differential equation:

$$\beta^{*\prime}(\theta) = \lambda(1+\theta)^2\alpha^{*\prime}(\theta).$$

From Proposition 1, $\beta^*(\theta)$ is thus increasing on the upper tail $[\Omega(a^*), \delta]$ and constant on the lower tail $[-\delta, \Omega(a^*)]$, $\beta^*(\theta) = \beta_u^*$.

On the other hand, the agent's participation constraint being binding at the optimum, we get:

$$\rho a^* E_\theta(U^*(\theta)) + (1-\rho a^*)U^*(0) = \varphi(a^*).$$

Rearranging this expression, using (6) and (19), we obtain the expression of β_u^*:

$$U^*(0) = \varphi(a^*) - a^*\varphi'(a^*) = U_u^* = \frac{\lambda}{2}\alpha_u^{*2} - \beta_u^* < 0$$

where the right-hand side equality follows from direct computations of the agent's payoff when he remains uninformed and the inequality from the convexity of $\varphi(\cdot)$.

This condition together with the continuity and the monotonicity of the optimal profile $U^*(\cdot)$ implies that $U^*(\theta) \leq 0$ on a whole interval $[-\delta, \theta_1]$ for some $\theta_1 > 0$. The agent gets a positive rent only if he reports sufficiently *"good news"*. The optimal contract is *most of the time* biased towards negative payoffs and the agent's *ex ante* participation constraint only holds in expectation because the rent $U^*(\theta)$ increases sufficiently quickly with θ over the upper tail of the distribution.[25]

Effort in gathering information. The principal modifies the optimal menus by implementing low-powered incentives for *"bad news"* but, overall, menus are somewhat "flatter" to make the agent be less willing to invest in information gathering in the first place.

Indeed, the sign of the Lagrange multiplier for the incentive constraint (20) captures whether the agent exerts too much effort or too little on his own. It turns out that this multiplier is negative when the agent is pessimistic as we show in

[25] Adding ex post participation constraints for all possible values of θ would bring our analysis closer to the work of Cremer et al. (1998b).

Appendix A. This sign has important consequences on the shape of the optimal contract. It implies that the optimal contract remains continuous in our set-up with an ex ante participation constraint when the agent is pessimistic while it exhibits a strong discontinuity in the framework of Cremer et al. (1998b) where the agent is protected by interim participation constraints. Theorem 2 below shows that a similar discontinuity arises when the agent is instead optimistic and the Lagrange multiplier is positive.[26]

5. Discussion

5.1. The value of communication

In an important paper, Laffont and Tirole (1986) have shown how a menu of linear schemes can sometimes be used to implement the optimal nonlinear contract when the information structure is exogenously given. Taking a reverse perspective, we may ask here whether, on the other hand, the menu of linear contracts that we have used so far could be replaced by its upper envelope in our context where the information structure is endogenized. We prove below that the principal would achieve a very different payoff by offering the nonlinear payment schedule obtained as the upper envelope of the linear payments found above. In other words, there is a positive value of communication between the principal and the agent.[27]

To see why, let us define the upper envelope of the optimal menu of linear schemes found above as a maximum of (increasing) linear functions:

$$T^*(y) = \max_{\theta \geq \Omega(a^*)} \alpha^*(\theta) y - \beta^*(\theta).$$

By construction, $T^*(y)$ is itself increasing, convex with a slope in between α^*_u and $\alpha^*(\bar{\theta}) = 1$.

When he is informed on the realization of the shock θ, the agent exerts the same effort with $T^*(y)$ as when he chooses his most preferred option $\alpha^*(\theta)y - \beta^*(\theta)$ (for $\theta \geq \Omega(a^*)$) within with the menu of linear contracts.

Consider instead an uninformed agent facing $T^*(y)$ instead of the linear payment rule $\alpha^*_u y - \beta^*_u$ that he should be taking within the menu of linear options. Incentives at the operating stage are modified accordingly. Using a (necessary and sufficient first-order) condition for optimality and taking into account that $T^{*\prime}(y) \geq \alpha^*_u$, the uninformed agent now chooses an effort e_u at the operating stage that satisfies:

$$e_u = \lambda E_\theta((1+\theta)T^{*\prime}((1+\theta)e_u)) \geq \lambda \alpha^*_u E_\theta(1+\theta) = \lambda \alpha^*_u \Rightarrow e_u > e^*_u$$

where e^*_u is the operating effort under the optimal linear scheme $\alpha^*_u y - \beta^*_u$. Hence, the convexity of $T^*(y)$ implies that an uninformed agent would now have too much incentives at the operating stage.

There is thus a positive value of communicating through a menu of contracts and it comes precisely from the fact that the agent has to (truthfully) claim that he is uninformed so as to operate on a less powered incentive scheme.

5.2. Optimistic agent

Suppose now that the agent is more optimistic than the principal on the possibility of gathering information, i.e., $\rho > 1$. For tractability, we also posit a quadratic specification $\varphi(a) = \frac{\varphi a^2}{2}$, and that φ is large enough to ensure that the optimal first-stage effort always remains within $[0, 1]$ under all circumstances below.

Verifiable shocks. The analysis under that scenario follows closely that of Section 3.2 with the sole exception that $\mathcal{R}(a) = a\varphi'(a)\left(\frac{1}{\rho} - 1\right)$ is now negative; it is now less costly for the principal to induce an optimistic agent to gather information. Assuming nevertheless that the overall cost of effort, namely $\varphi(a) + \mathcal{R}(a) = \varphi\left(\frac{1}{\rho} - \frac{1}{2}\right)a^2$, remains strictly convex in a, which requires $\rho < 2$, Proposition 1 carries over *mutatis mutandis*. The only difference comes from the fact that now:

$$\hat{a} = \frac{\rho\lambda\sigma^2}{2(2-\rho)\varphi} > \frac{\rho\lambda\sigma^2}{2\varphi} = a_0.$$

In other words, the optimistic agent does not invest enough in information gathering compared to what the principal would like him to do. To induce such effort, the principal now requests a lower fee when information has been gathered and thus $\hat{\beta}_u > \hat{\beta}$.

Non-observable information gathering. The profile of optimal bonuses must now be modified to boost information gathering. From (20), this can be done by implementing rent profiles $U(\theta)$ which are more *"convex"* than those obtained with

[26] Szalay (2009) also criticizes those discontinuities as coming from the specification of the (deterministic) information gathering technology made by Cremer et al. (1998b).

[27] This finding is of course reminiscent of the work of Melumad and Reichelstein (1989), although these authors address this issue in the context of a risk-neutral agent who has exogenous private information on a technological parameter before exerting a non-verifiable effort whose return is random.

fixed-price contracts; a scenario very close to the analysis in Szalay (2009) (who, remember, considers interim participation constraints as in Cremer et al. (1998b)). In particular, offering extra incentives following *"good news"* (i.e., $\alpha(\theta) > 1$ for $\theta > 0$) and lower incentives following *"bad news"* (i.e., $\alpha(\theta) < 1$ for $\theta < 0$) would certainly ease the incentive constraint (20).

Such distortions nevertheless conflict with the monotonicity requirement on $\alpha(\cdot)$ that follows from Lemma 1. To see why, observe that the *"no distortion at the top"* condition familiar from the screening literature would call for setting $\alpha(\delta) = 1$ and thus monotonicity is violated for $\theta > 0$ since we should also have $\alpha(0^+) > 0$. A symmetric argument shows that this monotonicity also fails for $\theta < 0$. To reconcile truth-telling constraints and *ex ante* incentives, the optimal contract becomes a rather simple menu. Only three options are offered. The first one is a fixed-price contract taken when the agent remains uninformed. The second one entails very high powered incentives and is taken by an agent who has learned *"good news"*. The last option has instead low powered incentives and is targeted to an agent having learned *"bad news"*. Those features are summarized in the next proposition.

Theorem 2. *The optimal contract with an optimistic agent entails the following discontinuous bonuses:*

$$\alpha^*(\theta) = \begin{cases} \alpha_1 < 1 & \textit{if}\ \theta < 0 \\ \alpha_2 > 1 & \textit{if}\ \theta \geq 0 \end{cases} \ \textit{and}\ \alpha_u^* = 1.$$[28]

The discontinuity exhibited by the optimal contract above bears some resemblance with the result found in Cremer et al. (1998b). Those authors analyze a standard Baron–Myerson environment with *ex post* participation constraint and show that the optimal contract entails lower (resp. greater) distortions than in Baron and Myerson (1982) depending on whether the agent learns that his cost parameter is above its mean value or below. We borrow from this important paper the trick that an uninformed agent behaves as having a type which is at the mean value of the distribution. Beyond, the settings are significantly different. In particular, and although the principal and his agent have different priors, the agent's participation constraint is written at the *ex ante* stage in our setting. This difference in timing implies that, had the information structure being exogenously given, the optimal allocation would already entail a downward distortion in Cremer et al. (1998b)'s framework for the usual rent extraction reasons while it would not in ours. A fixed-price contract making the agent residual claimant for the output would implement the first best. When the profile of information rent corresponding to the Baron–Myerson allocation does not suffice to induce information gathering, the contract in Cremer et al. (1998b) must thus be further distorted to make this rent profile more convex so as to induce this effort, just as what arises in our setting when the agent is too optimistic. Yet, these extra distortions can be imposed without violating monotonicity conditions because this reference point is already sufficiently distorted downward. This is not the case in our setting since the reference point is a flat fixed-price contract. Making the allocation more convex may then conflict with the monotonicity requirement on bonuses.

5.3. (First-order) risk aversion and recursive utility

With minor modifications, our framework can easily be adapted to handle differences in the degree of risk aversion between the principal and his agent instead of differences in beliefs. This section sketches how it can be so. Suppose that the agent now exhibits (first-order) risk aversion in a framework with recursive utility à la Epstein and Zin (1990). More precisely, the agent has first-period preferences over effort a and the certainty equivalent of his future payoffs x given by:

$$v(x) - \varphi(a) = \begin{cases} \alpha x - \varphi(a) & \text{if}\ x \geq 0, \\ x - \varphi(a) & \text{otherwise.} \end{cases}$$

The parameter $\alpha \in [0, 1]$ captures the agent's risk aversion. The agent is more reluctant to get a random albeit positive payoff when α decreases.

With those notations, we can thus write the agent's *ex ante* payoff as:

$$\alpha a E_\theta(U(\theta)) + (1 - a)U_u - \varphi(a)$$

From this, we may derive the expression of the agent's *ex ante* incentive constraint as:

$$\alpha E_\theta(U(\theta)) - U_u = \varphi'(a). \tag{24}$$

For a menu of contracts such that the agent's *ex ante* payoff is zero, we thus have:

$$E_\theta(U(\theta)) = \frac{1}{\alpha}\left(\varphi(a) + (1 - a)\varphi'(a)\right) > 0 \text{ and } U_u = \varphi(a) - a\varphi'(a) < 0.$$

Following the same steps as before and using the same notations, we may rewrite the principal's agency cost of providing incentives as:

[28] We could as well have chosen another tie-breaking rule at $\theta = 0$ and set $\alpha^*(0) = \alpha_1$.

$$aE_\theta(U(\theta))+(1-a)U_u-\varphi(a)=\mathcal{R}_r(a)=\left(\frac{1}{\alpha}-1\right)a(\varphi(a)+(1-a)\varphi'(a)). \tag{25}$$

We can now transform Lemma 2 so as to rewrite (24) as[29]:

$$\varphi'(a)+\frac{1-\alpha}{\alpha}\left(\varphi(a)+(1-a)\varphi'(a)\right)=\lambda E_\theta\left(\frac{1_{\theta>0}-F(\theta)}{f(\theta)}(1+\theta)\alpha^2(\theta)\right). \tag{26}$$

Two features make this model very close to our main setting assuming different priors. First, the agency cost $\mathcal{R}_r(a)$ is positive and increasing in a. As in our main framework, this calls for modifying the menu of contracts so as to reduce effort in information gathering. Second, had the agent been offered fixed-price contracts (i.e., $\alpha(\theta)\equiv 1$ for all θ), he would again exert an effort a_0 such that $\varphi'(a_0)+\frac{1-\alpha}{\alpha}\left(\varphi(a_0)+(1-a_0)\varphi'(a_0)\right)=\frac{\lambda\sigma^2}{2}$ which is lower than the first best level a^{fb}. Although details differ due to changes in functional forms, these two features make the analysis of this case of recursive utility with first-order risk aversion very close to that investigated in our main scenario.

6. Conclusion

We have analyzed a contracting environment where an agent may invest to gather private information on a productivity shock that will affect the marginal benefit of his operating effort. When choosing his effort at the operating stage, the agent may or may not be informed on the state of nature and both scenarios arise with positive probabilities as a result of the agent's earlier investment in (imperfect) information gathering. The agent is more pessimistic than the principal on the success of information gathering. The optimal menu of contract entails screening on the upper tail of the distribution while, as a result of countervailing incentives in information gathering and screening, incentives to operate do not depend on realized shock on the lower tail.

Several extensions of our analysis might be worth pursuing. First, it would be interesting to analyze whether information gathering and operating assets are tasks that should be bundled or instead split between two different agents. Two conflicting effects are at play in this regard. On the one hand, our analysis demonstrates that bundling may be beneficial on the upper tail of the distribution since screening and information gathering distortions go in the same direction. On the other hand, those distortions go in opposite directions on the lower tail which may call for splitting those tasks.[30]

Second, we could follow Szalay (2009) and investigate the nature of information gathering with more general information structures. Instead of modeling information gathering as an *"all-or-nothing"* phenomenon as we do here following Cremer et al. (1998b), Szalay (2009) introduces a more flexible technology where the agent gets a more precise signal on the realized productivity shocks as he exerts more effort in information gathering. It could be worth to extend his analysis to the case of asymmetric prior beliefs investigated in this paper.

We hope to investigate those issues in future research.

Appendix A

Proof of Proposition 1. We suppose that the agent's effort at the information gathering stage cannot be verified. Whether information has been gathered or not and the realized productivity shocks (in case information gathering has been successful) can be verified. Optimal efforts are given by:

$$e(\theta)=\lambda(1+\theta)\alpha(\theta) \text{ and } e_u=\lambda\alpha_u. \tag{A1}$$

The principal's problem can thus be written as:

$$(\mathcal{P}^c):\max_{(\alpha(\theta),\alpha_u,U(\theta),U_u,a)} aE_\theta\left(\lambda(1+\theta)^2\left(\alpha(\theta)-\frac{1}{2}\alpha^2(\theta)\right)-U(\theta)\right)+(1-a)\left(\lambda\left(\alpha_u-\frac{1}{2}\alpha_u^2\right)-U_u\right)$$

subject to (5) and (6).

Inserting (6) into (5), we may rewrite the agent's participation constraint as:

$$aE_\theta\left(U(\theta)\right)+(1-a)U_u-\varphi(a)\geq\mathcal{R}(a). \tag{A2}$$

Observing that (A2) is necessarily binding, we may rewrite the optimization problem as:

$$(\mathcal{P}^c):\max_{(\alpha(\theta),\alpha_u,a)} a\lambda E_\theta\left((1+\theta)^2\left(\alpha(\theta)-\frac{1}{2}\alpha^2(\theta)\right)\right)+(1-a)\lambda\left(\alpha_u-\frac{1}{2}\alpha_u^2\right)-\varphi(a)-\mathcal{R}(a).$$

[29] We are using here that $U(0)=U_u=\varphi(a)-a\varphi'(a)$ which is again implied by incentive compatibility just as in Lemma A.1 in Appendix A.

[30] Specific organizational issues raised when planning and implementation are either merged or split between two separate agents have been addressed in Lewis and Sappington (1997), Khalil et al. (2006), Krähmer and Strausz (2011b) and Hoppe and Schmitz (2013).

This objective is exactly the same as if the effort at the information gathering stage were observable except for the extra agency cost of information gathering $\mathcal{R}(a)$ (from (6) and (8)).

Optimizing immediately yields the expressions of the optimal bonuses given by (9). The corresponding efforts at the operating stage are then given by (10). The optimality condition with respect to a yields (11).

Taking into account (9) and inserting into (3) and (4), we obtain the following expressions of the rents depending on whether the agent gets informed or not:

$$\hat{U}(\theta) = \frac{\lambda}{2}(1+\theta)^2 - \beta(\theta) \text{ and } \hat{U}_u = \frac{\lambda}{2} - \beta_u. \tag{A3}$$

With those expressions, the incentive constraint (6) becomes:

$$\frac{1}{\rho}\varphi'(\hat{a}) = E_\theta\left(\frac{\lambda}{2}(1+\theta)^2 - \beta(\theta)\right) - \left(\frac{\lambda}{2} - \beta_u\right). \tag{A4}$$

Since this incentive constraint depends only on the expected fee $E_\theta(\beta(\theta))$, we can choose to set a constant fee:

$$\beta(\theta) \equiv \hat{\beta} \quad \forall\theta.$$

From (A4) and (11), it then follows that:

$$\hat{\beta}_u - \hat{\beta} = \frac{1}{\rho}\varphi'(\hat{a}) - \frac{\lambda\sigma^2}{2} = -\left(\frac{1}{\rho} - 1\right)\hat{a}\varphi''(\hat{a}) < 0$$

where the negative sign comes from the fact that (11) also implies $\hat{a} > 0$. □

Proof of Lemma 1. When the agent knows the realization of the shock θ and chooses the linear contract $t(y,\theta) = \alpha(\theta)y - \beta(\theta)$, he exerts an effort level such that:

$$e(\theta) = \arg\max_e \alpha(\hat{\theta})(1+\theta)e - \beta(\hat{\theta}) - \psi(e). \tag{A5}$$

The first-order condition for this strictly concave problem yields (14). Taking into account the expression of the effort level, we can rewrite

$$U(\theta) = \max_{\hat{\theta}\in\Theta} \frac{\lambda}{2}\alpha^2(\hat{\theta})(1+\theta)^2 - \beta(\hat{\theta}). \tag{A6}$$

From this, we immediately obtain that $U(\theta)$ is the maximum of convex functions and as such it is itself convex, absolutely continuous, and thus it admits everywhere a sub-differential which is almost everywhere single valued. At any such point of differentiability, (16) holds. Absolute continuity yields (15).

Using simple revealed arguments and (A6), we get for $\theta \geq \hat{\theta}$

$$\frac{\lambda}{2}\alpha^2(\theta)(1+\theta)^2 - \beta(\theta) \geq \frac{\lambda}{2}\alpha^2(\hat{\theta})(1+\theta)^2 - \beta(\hat{\theta}).$$

Writing the same incentive constraint but now between $\hat{\theta}$ and θ, summing on both sides and simplifying yields immediately that $\alpha(\theta) \geq \alpha(\hat{\theta})$ for $\theta \geq \hat{\theta}$. $\alpha(\cdot)$ is monotonically increasing and thus a.e. differentiable with at any point of differentiability:

$$\alpha'(\theta) \geq 0. \qquad \square$$

Simplifying the incentive feasible set. The set of incentive feasible allocations is characterized by constraints (5)–(6)–(17) and (18) plus the monotonicity condition on $\alpha(\cdot)$. The first step of our analysis is to simplify this characterization to facilitate optimization. This step is accomplished by studying each of those constraints in turn.

First observe that the agent who has learned that the shock is just the mean of the distribution could behave as being uninformed and *vice versa*. When uninformed, the agent thus gets the same payoff as when he knows that the productivity shock is the mean. Next Lemma immediately follows.

Lemma A.1. *$U(\theta)$ is more convex than $V(\theta)$*[31] *with:*

$$U(0) = V(0) = U_u \tag{A7}$$

and

$$\alpha(0^-) \leq \alpha_u \leq \alpha(0^+). \tag{A8}$$

[31] Formally, our notion of being "more convex" means that the convex epigraph of U is a subset of the convex epigraph of V: $epi\quad U = \{y| \quad y \geq U(\theta), \theta \in \Theta\} \subset epi\quad V = \{y| \quad y \geq U(\theta), \theta \in \Theta\}$ with (A7) ensuring that those epigraphs have a common point. This definition is consistent with other notions of relative convexity like those found in Palmer (2003), for instance.

Proof of Lemma A.1. From (17), we get $U(0) \geq V(0)$. But taking expectations on the left-hand side of (18), we get $U_u = V(0) \geq U(0)$. Since $U(\theta)$ is absolutely continuous and convex, it admits right- and left-derivatives at 0 where its subgradient is $\partial_\theta U(0) = [\lambda\alpha(0^-), \lambda\alpha(0^+)]$. Because $V(\theta)$ is also convex, minorizes $U(\theta)$ at $\theta = 0$ and has derivative α_u at that point, (A8) must hold. □

If $\alpha(\cdot)$ is continuous at 0, condition (A8) is akin to a "smooth-pasting" requirement, i.e., the rent profile $U(\theta)$ tangentially touches $V(\theta)$ at 0. Smooth-pasting holds at the optimal contract, as we will see below.

From (A7) and (17), we get for $\theta \geq 0$

$$\frac{1}{\theta}(U(\theta) - U(0)) = \frac{\lambda}{\theta}\int_0^\theta (1+x)\alpha^2(x)dx \geq \frac{\lambda\alpha_u^2}{2}(2+\theta)$$

Taking limits on both sides when $\theta \mapsto 0^+$ yields $\alpha^2(0^+) \geq \alpha_u^2$. Proceeding similarly for $\theta \mapsto 0^-$, we get $\alpha_u^2 \geq \alpha^2(0^-)$. Finally, we obtain (A8).

Lemma 1 and condition (A8) altogether imply that $\alpha(\theta) \geq \alpha(0^+) \geq \alpha_u$ for $\theta \geq 0$. From this, it follows that

$$\lambda\int_0^\theta (1+x)\alpha^2(x)dx \geq \frac{\lambda\alpha_u^2}{2}\theta(2+\theta) \quad \forall\theta \geq 0.$$

Thus, $U(\theta) \geq V(\theta)$ for all $\theta \geq 0$. The proof is similar for $\theta \leq 0$. Condition (17) holds everywhere. □

Proof of Lemma 2. Using (15), we write:

$$E_\theta(U(\theta)) = U(0) + \int_0^\delta f(\theta)\left(\int_0^\theta (1+x)\alpha^2(x)dx\right)d\theta + \int_{-\delta}^0 f(\theta)\left(\int_0^\theta (1+x)\alpha(x)dx\right)d\theta.$$

Integrating by parts each of those integrals on the right-hand side leads to

$$E_\theta(U(\theta)) - U(0) = E_\theta\left(\lambda\frac{1_{\theta>0} - F(\theta)}{f(\theta)}(1+\theta)\alpha^2(\theta)\right).$$

Taking into account (A7) gives us (20). □

Proof of Theorem 1. We first set up the principal's problem before finding the optimal contract.

The Principal's problem. Taking into account the expression of operating efforts in terms of bonuses (see (A1)), this optimization problem writes as:

$$(\mathcal{P}): \max_{(\alpha(\theta),\alpha_u,a)} a\lambda E_\theta\left((1+\theta)^2\left(\alpha(\theta) - \frac{1}{2}\alpha^2(\theta)\right)\right) + (1-a)\lambda\left(\alpha_u - \frac{1}{2}\alpha_u^2\right) - \varphi(a) - \mathcal{R}(a)$$

subject to $\alpha(\theta)$ non-decreasing and satisfying (A8); and (20).

From Lemmas A.1 and 2, these constraints are sufficient to fully describe the constrained set. We now proceed in two steps. First, we replace (A8) and the requirement of monotonicity of $\alpha(\cdot)$ by the weaker constraints:

$$\alpha(\theta) \leq \alpha_u \quad \forall\theta \leq 0; \text{ and } \alpha_u \leq \alpha(\theta) \quad \forall\theta \geq 0. \tag{A9}$$

Second, we optimize the new problem $(\mathcal{P}^*)$ so obtained with (20) and (A9) as constraints. Lastly, we will show that the solution to this relaxed problem is such that $\alpha(\cdot)$ is indeed everywhere non-decreasing. □

Optimization. We first observe that the fact that $\alpha(\theta)$ is non-decreasing and satisfy (A8) altogether with Lemma A.1 imply that the incentive constraint (17) binds only on a single connected interval $[\theta_*, \theta^*]$ (with $\theta^* \geq \theta_*$) including $\theta = 0$ (but possibly reduced to a single point at $\theta = 0$). To show this, suppose indeed it is binding on two such disconnected intervals $[\theta_1, \theta_2]$ (with possibly $\theta_1 = \theta_2$) and $[\theta_*, \theta^*]$ with $-\delta \leq \theta_1 \leq \theta_2 < \theta_*$. On (θ_2, θ_*) we have $U(\theta) > V(\theta)$ which implies that necessarily there exists a subinterval with non-empty interior $(\theta_2 + \epsilon, \theta_2 + 2\epsilon)$ (with ϵ small enough) such that $U'(\theta) > V'(\theta)$, i.e., $\alpha(\theta) > \alpha_u$, on that subinterval. This would contradict the monotonicity of $\alpha(\cdot)$ and in particular the fact that $\alpha(0^-) \leq \alpha_u$.

Consider now the interval $[\theta_*, \theta^*]$ and suppose it has a non-empty interior. Since $U(\theta) = V(\theta)$ on that interval, differentiating with respect to θ yields $V'(\theta) = U'(\theta)$ and thus $\alpha(\theta) = \alpha_u$. Of course, on $[\underline{\theta}, \theta_*]$, the monotonicity of $\alpha(\cdot)$ implies that we must have $\alpha(\theta) \le \alpha_u$. This property is used below to prove that indeed, $\theta_* = \underline{\theta}$.

With those earlier findings in mind, we can rewrite the information gathering incentive constraint (A7) as:

$$\lambda\left(\int_{\theta^*}^{\delta}(1-F(\theta))(1+\theta)\alpha^2(\theta)d\theta - \int_{-\delta}^{\theta_*}F(\theta)(1+\theta)\alpha^2(\theta)d\theta + \int_{\theta_*}^{\theta^*}(1_{\theta>0}-F(\theta))(1+\theta)\alpha_u^2 d\theta\right)$$
$$= \frac{1}{\rho}\varphi'(a). \tag{A10}$$

First, we omit constraint (A9) for a while. Second, we write the Lagrangean of the so-relaxed problem $(\mathcal{P}^*)$ as:

$$\begin{aligned} L(\alpha(\cdot), \alpha_u, a, \mu) = {} & a\lambda\left(\int_{\theta^*}^{\delta}(1+\theta)^2\left(\alpha(\theta)-\frac{1}{2}\alpha^2(\theta)\right)f(\theta)d\theta + \int_{-\delta}^{\theta_*}(1+\theta)^2\left(\alpha(\theta)-\frac{1}{2}\alpha^2(\theta)\right)f(\theta)d\theta\right) \\ & + \lambda\left(1-a+a\int_{\theta_*}^{\theta^*}(1+\theta)^2 f(\theta)d\theta\right)\left(\alpha_u-\frac{1}{2}\alpha_u^2\right) - \mathcal{R}(a) - \varphi(a) \\ & + \mu\lambda\left(\int_{\theta^*}^{\delta}(1-F(\theta))(1+\theta)\alpha^2(\theta)d\theta - \int_{-\delta}^{\theta_*}F(\theta)(1+\theta)\alpha^2(\theta)d\theta\right. \\ & \left. + \alpha_u^2\int_{\theta_*}^{\theta^*}(1_{\theta>0}-F(\theta))(1+\theta)d\theta\right) - \frac{\mu}{\rho}\varphi'(a) \end{aligned}$$

where μ is the multiplier of (A10), shown below to be negative. Following Ruszczynski (2006, Theorem 3.25) and observing that the constraint qualification of Robinson's kind trivially holds,[32] necessary conditions for optimality are as follows.

- We optimize w.r.t. α_u and pointwise w.r.t. $\alpha(\theta)$.
  1. On $[\theta_*, \theta^*]$, we get:

$$\left(1-a+a\int_{\theta_*}^{\theta^*}(1+\theta)^2 f(\theta)d\theta\right)(1-\alpha_u) + 2\alpha_u\mu\int_{\theta_*}^{\theta^*}(1_{\theta>0}-F(\theta))(1+\theta)d\theta = 0. \tag{A11}$$

  2. On $[\theta^*, \delta]$, we get:

$$\begin{aligned} & a(1+\theta)^2(1-\alpha(\theta))f(\theta) + 2\mu(1-F(\theta))(1+\theta)\alpha(\theta) = 0 \\ & \Leftrightarrow \alpha(\theta) = \frac{1}{1-\frac{2\mu R(\theta)}{a}}. \end{aligned} \tag{A12}$$

  Note that, under Assumption 1, $\alpha(\theta)$ is increasing on $[\theta^*, \delta]$ when $\mu < 0$ and is worth 1 at $\theta = \delta$ if $\theta^* < \delta$.

  3. On $[-\delta, \theta_*]$, and if $\alpha(\theta)$ is not constrained by (A9), we get:

$$\begin{aligned} & a(1+\theta)^2(1-\alpha(\theta))f(\theta) - 2\mu F(\theta)(1+\theta)\alpha(\theta) = 0 \\ & \Leftrightarrow \alpha(\theta) = \frac{1}{1+\frac{2\mu S(\theta)}{a}}. \end{aligned} \tag{A13}$$

  Observe that, under Assumption 1, $\alpha(\theta)$ so defined is monotonically increasing on $[-\delta, \theta_*]$ when $\mu < 0$ and $\alpha(-\delta) = 1$ if $\theta_* > -\delta$, so that $\alpha(\theta) \ge 1$ for all $\theta \in [-\delta, \theta_*]$. In particular, this unconstrained solution always violates (A9) since the monotonicity condition on $\alpha(\cdot)$ implies that $\alpha_u \le \alpha(\theta^*) \le \alpha(\bar{\theta}) = 1$. Necessarily, the constraint (A9) thus binds on the whole interval $[-\delta, \theta_*] \bigcup [\theta_*, \theta^*]$. We conclude that

$$\alpha(\theta) = \alpha_u \quad \forall \theta \in [-\delta, \theta^*]. \tag{A14}$$

[32] See Ruszczynski (2006, Lemma 3.16).

We may then also rewrite (A10) as:

$$\lambda\left(\int_{\theta^*}^{\delta}(1-F(\theta))(1+\theta)\alpha^2(\theta)d\theta+\int_{\underline{\theta}}^{\theta^*}(1_{\theta>0}-F(\theta))(1+\theta)\alpha_u^2 d\theta\right)=\frac{1}{\rho}\varphi'(a). \tag{A15}$$

- Inserting the condition (A14) into (A11) and taking into account that $\theta_*=-\delta$ yields:

$$\alpha_u=\frac{1}{1-\frac{2\mu\int_{-\delta}^{\theta^*}(1_{\theta>0}-F(\theta))(1+\theta)d\theta}{1-a+a\int_{-\delta}^{\theta^*}(1+\theta)^2 f(\theta)d\theta}}. \tag{A16}$$

  Observe that $\alpha_u<1$ when $\mu<0$.
- Optimizing the Lagrangean w.r.t. θ^* yields:

$$a(1+\theta^*)^2 f(\theta^*)\left(\alpha(\theta^{*+})-\frac{1}{2}\alpha^2(\theta^{*+})-\alpha_u+\frac{1}{2}\alpha_u^2\right)+\mu(1-F(\theta^*))(1+\theta^*)\left(\alpha_u^2-\alpha^2(\theta^{*+})\right)=0. \tag{A17}$$

  An obvious solution to (A17) consists in having $\alpha(\theta)$ continuous at θ^*:

$$\alpha(\theta^{*+})=\frac{1}{1-\frac{2\mu R(\theta^*)}{a}}=\alpha_u \tag{A18}$$

  From there and assuming that $\mu\neq 0$ (an assertion to be checked below), we immediately deduce that θ^* is such that is solves (21) and thus $\theta^*=\Omega(a)$.
- Finally, optimizing the Lagrangean w.r.t. a yields:

$$\varphi'(a)+((1-\rho)a+\mu)\varphi''(a)$$
$$=\rho\lambda\left(\left(\int_{\theta^*}^{\delta}(1+\theta)^2\left(\alpha(\theta)-\frac{1}{2}\alpha^2(\theta)\right)f(\theta)d\theta\right)-\left(\alpha_u-\frac{1}{2}\alpha_u^2\right)\left(1-\int_{-\delta}^{\theta^*}(1+\theta)^2 f(\theta)d\theta\right)\right). \tag{A19}$$

Proof of $\theta^*>0$. From Assumption 1, the left-hand side of (21) is a decreasing function of θ^* which is equal to $\frac{1}{2f(0)}$ for $\theta^*=0$ and zero for $\theta^*=\delta$. The right-hand side, that will be denoted $\chi(\theta)$, is negative at $\theta=0$ and positive for $\theta=\delta$ (since after integrating by parts, the numerator becomes $\int_{-\delta}^{\delta}(1_{\theta>0}-F(\theta))(1+\theta)d\theta=\frac{1}{2}E_\theta((1+\theta)^2-1)=\frac{\sigma^2}{2}>0$). It is zero for θ_0 such that $\frac{\sigma^2}{2}=\int_{\theta_0}^{\delta}(1-F(\theta))(1+\theta)d\theta$. Moreover, it can be checked that $\dot{\chi}(\theta^*)=0$ and $\dot{\chi}(\theta)<0$ in the neighborhood of $\theta=\delta$. Hence, there exists a unique $\theta^*>0$ that solves (21) and $\chi(\theta)$ has a maximum there. Let denote this solution as $\theta^*=\Omega(a)$. It can be readily seen that $\Omega(a)$ decreases with a.

Proof of $\alpha_u<1$ and $\mu<0$. Eliminating μ/a from (A12) and (A16) and simplifying by using (21), we get:

$$\alpha(\theta)=\Lambda(\alpha_u,a,\theta)\equiv\frac{\alpha_u}{\alpha_u+(1-\alpha_u)\frac{R(\theta)}{R(\Omega(a))}}\quad\forall\theta\geq\Omega(a). \tag{A20}$$

This, together with the fact that $\theta^*=\Omega(a)>0$, allows us to rewrite (A15) as:

$$\lambda\left(\int_{\Omega(a)}^{\delta}(1-F(\theta))(1+\theta)\left[\alpha^2\right]_{\alpha_u}^{\Lambda(\alpha_u,a,\theta)}d\theta+\frac{\sigma^2}{2}\alpha_u^2\right)=\frac{1}{\rho}\varphi'(a).^{33} \tag{A21}$$

The principal's problem can be rewritten as a simple optimization over (α_u,a) as:

$$(\mathcal{P}):\quad\max_{(\alpha_u,a)}\lambda\left(\alpha_u-\frac{1}{2}\alpha_u^2\right)(1+a\sigma^2)+\lambda a\int_{\Omega(a)}^{\delta}(1+\theta)^2 f(\theta)\left[\alpha-\frac{1}{2}\alpha^2\right]_{\alpha_u}^{\Lambda(\alpha_u,a,\theta)}d\theta-\varphi(a)-\mathcal{R}(a)$$

subject to (A21).

[33] We use the compact notation $[f]_y^x=f(x)-f(y)$.

From (A21), we may define a as a function of α_u, say $\Xi(\alpha_u)$. Observe that:

$$\frac{\varphi''(\Xi(\alpha_u))}{\lambda\rho}\Xi'(\alpha_u) = \sigma^2\Xi(\alpha_u) + 2\int_{\Omega(\Xi(\alpha_u))}^{\delta} (1-F(\theta))(1+\theta)\left(\Lambda(\alpha_u,\Xi(\alpha_u),\theta)\left(\frac{\partial\Lambda}{\partial\alpha_u}(\alpha_u,\Xi(\alpha_u),\theta)\right.\right.$$
$$\left.\left.+\frac{\partial\Lambda}{\partial a}(\alpha_u,\Xi(\alpha_u),\theta)\Xi'(\alpha_u)\right)-\alpha_u\right)d\theta. \quad \text{(A22)}$$

We compute $\Lambda(1,a,\theta)=1$. Inserting into (A21) yields $\Xi(1)=a_0$. Observe also that $\frac{\partial\Lambda}{\partial\alpha_u}(\alpha_u,\Xi(\alpha_u),\theta) = \frac{\frac{R(\theta)}{R(\Omega(\Xi(\alpha_u)))}}{\left(\alpha_u+(1-\alpha_u)\frac{R(\theta)}{R(\Omega(\Xi(\alpha_u)))}\right)^2}$ and $\frac{\partial\Lambda}{\partial a}(\alpha_u,a,\theta) = \frac{\alpha_u(1-\alpha_u)\frac{R(\theta)R'(\Omega(a))}{R^2(\Omega(a))}}{\left(\alpha_u+(1-\alpha_u)\frac{R(\theta)}{R(\Omega(a))}\right)^2}\Omega'(a)$. Thus, we also get $\frac{\partial\Lambda}{\partial\alpha_u}(1,a_0,\theta)=\frac{R(\theta)}{R(\Omega(\Xi(\alpha_u)))}$ and $\frac{\partial\Lambda}{\partial a}(1,a_0,\theta)=0$. Inserting those findings into (A22) yields:

$$\frac{\varphi''(a_0)}{\lambda\rho}\Xi'(1)=\sigma^2 a_0+2\int_{\Omega(a_0)}^{\delta}(1-F(\theta))(1+\theta)\left(\frac{R(\theta)}{R(\Omega(a_0))}-1\right)d\theta>0.$$

Inserting $a=\Xi(\alpha_u)$ into the maximand of $(\mathcal{P})$, this maximand becomes a function of α_u only, say $V(\alpha_u)$:

$$V(\alpha_u)=\lambda\left(\alpha_u-\frac{1}{2}\alpha_u^2\right)(1+\Xi(\alpha_u)\sigma^2)$$
$$+\lambda\Xi(\alpha_u)\int_{\Omega(\Xi(\alpha_u))}^{\delta}(1+\theta)^2 f(\theta)\left[\alpha-\frac{1}{2}\alpha^2\right]_{\alpha_u}^{\Lambda(\alpha_u,\Xi(\alpha_u),\theta)}d\theta-\varphi(\Xi(\alpha_u))-\mathcal{R}(\Xi(\alpha_u)).$$

We compute the derivative of $V(\alpha_u)$ as:

$$V'(\alpha_u)=\left(\lambda\sigma^2\left(\alpha_u-\frac{1}{2}\alpha_u^2\right)-\frac{\varphi'(\Xi(\alpha_u))}{\rho}-\Xi(\alpha_u)\varphi''(\Xi(\alpha_u))\left(\frac{1}{\rho}-1\right)\right)\Xi'(\alpha_u)+\lambda(1-\alpha_u)(1+\sigma^2\Xi(\alpha_u))$$
$$+\lambda\left(\int_{\Omega(\Xi(\alpha_u))}^{\delta}(1+\theta)^2 f(\theta)\left[\alpha-\frac{1}{2}\alpha^2\right]_{\alpha_u}^{\Lambda(\alpha_u,\Xi(\alpha_u),\theta)}d\theta\right)\Xi'(\alpha_u)+\lambda\Xi(\alpha_u)\int_{\Omega(\Xi(\alpha_u))}^{\delta}(1+\theta)^2 f(\theta)$$
$$\times\left[(1-\Lambda(\alpha_u,\Xi(\alpha_u),\theta))\left(\frac{\partial\Lambda}{\partial\alpha_u}(\alpha_u,\Xi(\alpha_u),\theta)+\frac{\partial\Lambda}{\partial a}(\alpha_u,\Xi(\alpha_u),\theta)\Xi'(\alpha_u)\right)-(1-\alpha_u)\right]d\theta. \quad \text{(A23)}$$

That $\Lambda(1,a,\theta)=1$ also implies that the last three terms on the right-hand side of (A23) are zero for $\alpha_i=1$. Finally, we get:

$$V'(1)=\left(\frac{\lambda\sigma^2}{2}-\frac{\varphi'(a_0)}{\rho}-a_0\varphi''(a_0)\left(\frac{1}{\rho}-1\right)\right)\Xi'(1)=-a_0\varphi''(a_0)\left(\frac{1}{\rho}-1\right)\Xi'(1)<0.$$

From this, it necessarily follows that the optimal value of α_u satisfies:

$$\alpha_u^*<1. \quad \text{(A24)}$$

Inserting into (A16), it follows that $\mu<0$ as supposed.

Proof that $a^*<a_0$. From (A24) and the fact that $\int_{-\delta}^{\Omega(a^*)}(1-F(\theta))(1+\theta)d\theta\geq 0$ (which follows from the fact that $\Omega(a^*)$ is defined through (21) whose left-hand side is non-negative) and using that $\alpha_u^*\leq\alpha^*(\theta)\leq 1$ for all $\theta\in[\Omega(a^*),\delta]$, we get the following majorization

$$\frac{1}{\rho}\varphi'(a^*)\leq\lambda\left(\int_{\Omega(a^*)}^{\delta}(1-F(\theta))(1+\theta)d\theta+\int_{-\delta}^{\Omega(a^*)}(1_{\theta>0}-F(\theta))(1+\theta)d\theta\right)=\frac{\lambda\sigma^2}{2}=\frac{1}{\rho}\varphi'(a_0)$$

which ends the proof. □

Proof of Theorem 2. The principal's problem $(\mathcal{P})$ again writes as in the Proof of Theorem 1.

Relaxed problem $(\mathcal{P}^{**})$. We first neglect the monotonicity constraint on $\alpha(\cdot)$ and thus look for a solution to the relaxed problem $(\mathcal{P}^{**})$ so obtained. For this relaxed problem, we also conjecture that the incentive constraint (17) only binds at $\theta = 0$ (and check this conjecture *ex post*). We can now write the Lagrangean of the so-relaxed problem $(\mathcal{P}^{**})$ as:

$$L(\alpha(\cdot), \alpha_u, a, \mu) = a\lambda \int_{-\delta}^{\delta} (1+\theta)^2 \left(\alpha(\theta) - \frac{1}{2}\alpha^2(\theta)\right) f(\theta)d\theta + (1-a)\lambda\left(\alpha_u - \frac{1}{2}\alpha_u^2\right)$$
$$- \mathcal{R}(a) - \varphi(a) + \mu\left(\lambda\left(\int_0^{\delta}(1-F(\theta))(1+\theta)\alpha^2(\theta)d\theta - \int_{-\delta}^{0} F(\theta)(1+\theta)\alpha^2(\theta)d\theta\right) - \frac{\varphi'(a)}{\rho}\right)$$

where μ is the multiplier of (20), shown below to be positive.

Optimality conditions. Following Ruszczynski (2006, Theorem 3.25) and observing that the constraint qualification of the Robinson's kind trivially holds, the necessary conditions for optimality write as follows.

1. Optimizing w.r.t. α_u yields:

$$\alpha_u = 1. \tag{A25}$$

2. Optimizing pointwise w.r.t. $\alpha(\theta)$, we get:

$$\alpha(\theta) = \begin{cases} \frac{1}{1-\frac{2\mu R(\theta)}{a}} & \text{if } \theta \in (0, \delta], \\ \frac{1}{1+\frac{2\mu S(\theta)}{a}} & \text{if } \theta \in [-\delta, 0).^{34} \end{cases} \tag{A26}$$

3. Optimizing w.r.t. a and taking into account that $\varphi(a) = \frac{\varphi a^2}{2}$ yields:

$$(2-\rho)a + \mu = \frac{\rho\lambda}{\varphi}\left(\int_{-\delta}^{\delta}(1+\theta)^2\left(\alpha(\theta) - \frac{1}{2}\alpha^2(\theta)\right) f(\theta)d\theta - \frac{1}{2}\right). \tag{A27}$$

Inserting the values of $\alpha(\theta)$ from (A26) into (A27) gives us an implicit definition of a as a function of μ, say $a(\mu)$. Observe that $a(\mu)$ is decreasing when $\mu > 0$, so is $\frac{a(\mu)}{\mu}$. Inserting this latter expression again into (A26) gives us an expression of the bonus that only depends on μ as well, say $\alpha(\theta, \mu)$. From Assumption 1 and when $\mu > 0$, $\alpha(\theta, \mu)$ is decreasing with θ both for $\theta < 0$ and for $\theta > 0$, which would violate the monotonicity condition on those two intervals.

Finally, μ itself is defined implicitly as the solution to:

$$a(\mu) = \frac{\rho\lambda}{\varphi}\left(\int_0^{\delta}(1-F(\theta))(1+\theta)\alpha^2(\theta,\mu)d\theta - \int_{-\delta}^{0} F(\theta)(1+\theta)\alpha^2(\theta,\mu)d\theta\right). \tag{A28}$$

Observe that for $\mu = 0$, the left-hand side above is worth $\frac{\rho\lambda\sigma^2}{2(2-\rho)\varphi}$ while the right-hand side is worth $\frac{\rho\lambda\sigma^2}{2\varphi}$ which is lower when $\rho > 1$. Moreover, the right-hand side of (A28) is increasing in μ. Hence, the solution to (A28) has necessarily $\mu > 0$.

Solution to $(\mathcal{P})$. Because $\mu > 0$, the solution to the relaxed problem does not satisfy the monotonicity requirement on $\alpha(\cdot)$. There is thus bunching on both $[-\delta, 0]$ and $[0, \delta]$ and given that the solution to the relaxed problem is nowhere increasing (except with an upward jump at 0), we must look for a solution to $(\mathcal{P})$ which features bunching on each of those intervals. Let call α_1 and α_2 the respective values of those bonuses. With those notations, the Lagrangean of problem $(\mathcal{P})$ becomes:

[34] Observe that $\alpha(\theta)$ is discontinuous at $\theta = 0$ but that the choice of $\alpha(0)$ is arbitrary as long as the monotonicity condition would be satisfied, a point to which we come back below.

$$L(\alpha(\cdot),\alpha_u,a,\mu)=a\lambda\left(\left(\alpha_1-\frac{1}{2}\alpha_1^2\right)\int_{-\delta}^{0}(1+\theta)^2 f(\theta)d\theta+\left(\alpha_2-\frac{1}{2}\alpha_2^2\right)\int_{0}^{\delta}(1+\theta)^2 f(\theta)d\theta\right)$$
$$+(1-a)\lambda\left(\alpha_u-\frac{1}{2}\alpha_u^2\right)-\mathcal{R}(a)-\varphi(a)$$
$$+\mu\left(\lambda\left(\alpha_2^2\int_0^{\delta}(1-F(\theta))(1+\theta)d\theta-\alpha_1^2\int_{-\delta}^{0}F(\theta)(1+\theta)d\theta\right)-\frac{\varphi'(a)}{\rho}\right).$$

The necessary conditions for optimality write as follows.

1. Again, α_u satisfies (A25).
2. α_1 and α_2 are respectively given by:

$$\alpha_2=\frac{1}{1-\frac{2\mu\int_0^{\delta}(1-F(\theta))(1+\theta)d\theta}{a\int_0^{\delta}(1+\theta)^2 f(\theta)d\theta}},\tag{A29}$$

$$\alpha_1=\frac{1}{1+\frac{2\mu\int_{-\delta}^{0}F(\theta)(1+\theta)d\theta}{a\int_{-\delta}^{0}(1+\theta)^2 f(\theta)d\theta}}.\tag{A30}$$

   Observe that $\alpha_2>1>\alpha_1$ when $\mu>0$.
3. Optimizing w.r.t. a and taking into account that $\varphi(a)=\frac{\varphi a^2}{2}$ yields:

$$(2-\rho)a+\mu=\frac{\rho\lambda}{\varphi}\left(\left(\alpha_1-\frac{1}{2}\alpha_1^2\right)\int_{-\delta}^{0}(1+\theta)^2 f(\theta)d\theta+\left(\alpha_2-\frac{1}{2}\alpha_2^2\right)\int_{0}^{\delta}(1+\theta)^2 f(\theta)d\theta-\frac{1}{2}\right).\tag{A31}$$

Proof that $\mu > 0$. Inserting the values of α_1 and α_2 from (A29) and (A30) into (A31) gives us an implicit definition of a as a function of μ, that is again denoted as $a(\mu)$. Observe that $a(\mu)$ is decreasing when $\mu>0$, so is $\frac{a(\mu)}{\mu}$. Inserting this latter expression again into (A29) and (A30) gives us expressions of the bonuses α_1 and α_2 that depend on μ as well, say $\alpha_1(\mu)$ and $\alpha_2(\mu)$. Observe that, when $\mu>0$, $\alpha_2(\mu)$ is increasing with μ, while $\alpha_1(\mu)$ is decreasing. Finally, μ itself is defined implicitly as the solution to:

$$a(\mu)=\frac{\rho\lambda}{\varphi}\left(\alpha_2^2(\mu)\int_0^{\delta}(1-F(\theta))(1+\theta)d\theta-\alpha_1^2(\mu)\int_{-\delta}^{0}F(\theta)(1+\theta)d\theta\right).\tag{A32}$$

Observe that for $\mu=0$, (A28) implies that the left-hand side above is again worth $\frac{\rho\lambda\sigma^2}{2(2-\rho)\varphi}$ while $\alpha_1(\mu)=\alpha_2(\mu)=1$ then implies that the right-hand side is worth $\frac{\rho\lambda\sigma^2}{2\varphi}$ which is lower when $\rho>1$. Moreover, the right-hand side of (A32) is increasing in μ while the left-hand side is decreasing. Hence, the solution to (A28) has necessarily $\mu>0$. □

References

Arrow, K., 1986. Agency and markets. In: Arrow, K., Intriligator, M. (Eds.), Handbook of Mathematical Economics, vol. 3. North-Holland, Amsterdam.
Baker, G., 1992. Incentive contracts and performance measurement. J. Polit. Economy 100, 598–614.
Baker, G., Jorgensen, B., 2003. Volatility, noise and incentives. Mimeo, Harvard University.
Baron, D., Myerson, R., 1982. Regulating a monopolist with unknown costs. Econometrica 50, 911–930.
Bergemann, D., Välimäki, J., 2002. Information acquisition and efficient mechanism design. Econometrica 70, 1007–1033.
Bergemann, D., Välimäki, J., 2006. Information in mechanism design. In: Blundell, R., Newey, W., Persson, T. (Eds.), Proceedings of the 9th World Congress of the Econometric Society. Cambridge University Press, New York, pp. 186–221.
Compte, O., Jehiel, P., 2008. Gathering information before signing a contract: a screening perspective. Int. J. Ind. Organ. 26, 206–212.
Courty, P., Li, H., 2000. Sequential screening. Rev. Econ. Stud. 67, 697–717.
Cremer, J., Khalil, F., 1992. Gathering information before signing a contract. Amer. Econ. Rev. 82, 566–578.
Cremer, J., Khalil, F., 1994. Gathering information before a contract is offered. Europ. Econ. Rev. 38, 675–682.
Cremer, J., Khalil, F., Rochet, J.C., 1998a. Strategic information gathering before a contract is offered. J. Econ. Theory 81, 163–200.
Cremer, J., Khalil, F., Rochet, J.C., 1998b. Contracts and productive information gathering. Games Econ. Behav. 25, 174–193.
Cremer, J., Spiegel, Y., Zheng, C., 2009. Auctions with costly information acquisition. Econ. Theory 38, 41–72.
Dai, C., Lewis, T., Loppomo, G., 2006. Delegating management to experts. RAND J. Econ. 37, 503–520.
De La Rosa, L., 2011. Overconfidence and moral hazard. Games Econ. Behav. 73, 429–451.
Demski, J., Sappington, D., 1987. Delegated expertise. J. Acc. Res. 25, 68–89.

Eliaz, K., Spiegler, R., 2008. Consumer optimism and price discrimination. Theoretical Econ. 3, 459–497.
Epstein, L., Zin, S., 1990. First-order risk aversion and the equity premium puzzle. J. Monet. Econ. 26, 387–407.
Gerardi, D., Yariv, L., 2008. Information acquisition in committees. Games Econ. Behav. 62, 436–459.
Gershkov, A., Szentes, B., 2009. Optimal voting schemes with costly information acquisition. J. Econ. Theory 144, 36–68.
Gromb, D., Martimort, D., 2007. Collusion and the organization of delegated expertise. J. Econ. Theory 137, 271–299.
Grubb, M., 2009. Selling to overconfident consumers. Amer. Econ. Rev. 99, 1770–1807.
Guriev, S., 2001. On microfoundations of the dual theory of choice. Geneva Pap. Risk Ins. Theory 26, 117–137.
Hart, O., Holmström, B., 1987. The theory of contracts. In: Bewley, T. (Ed.), Advances in Economic Theory, Proceedings of the 1985 World Congress of the Econometric Society. Cambridge University Press, Cambridge.
Holmström, B., Milgrom, P., 1987. Aggregation and linearity in the provision of intertemporal incentives. Econometrica 55, 303–328.
Holmström, B., Milgrom, P., 1991. Multitask principal–agent analyses: incentive contracts, asset ownership, and job design. J. Law, Econ., Organ. 7, 24–52.
Hoppe, E., Schmitz, P., 2013. Public–private partnerships versus traditional procurement: innovation incentives and information gathering. RAND J. Econ. 44, 56–74.
Iossa, E., Legros, P., 2004. Auditing and property rights. RAND J. Econ. 35, 356–372.
Jeleva, M., Villeneuve, B., 2004. Insurance contracts with imprecise probabilities and adverse selection. Econ. Theory 23, 777–794.
Kessler, A., 1998. The value of ignorance. RAND J. Econ. 29, 339–354.
Khalil, F., Kim, D., Shin, D., 2006. Optimal task design: to integrate or separate planning and implementation? J. Econ. Manage. Strategy 15, 457–478.
Krähmer, D., Strausz, R., 2011a. The benefits of sequential screening. http://epub.ub.uni-muenchen.de/13191/1/363.pdf.
Krähmer, D., Strausz, R., 2011b. Optimal procurement contracts with pre-project planning. Rev. Econ. Stud. 78, 1015–1041.
Laffont, J.J., Martimort, D., 2002. The Theory of Incentives: The Principal–Agent Model. Princeton University Press.
Laffont, J.J., Tirole, J., 1986. Using costs to regulate firms. J. Polit. Economy 94, 614–641.
Lambert, R., 1986. Executive effort and selection of risky projects. RAND J. Econ. 17, 77–88.
Lewis, T., Sappington, D., 1989. Countervailing incentives in agency problems. J. Econ. Theory 49, 294–313.
Lewis, T., Sappington, D., 1991. All or nothing information control. Econ. Letters 37, 111–113.
Lewis, T., Sappington, D., 1993. Ignorance in agency problems. J. Econ. Theory 61, 169–183.
Lewis, T., Sappington, D., 1997. Information management in incentive problems. J. Polit. Economy 105, 796–821.
Malcomson, J., 2009. Principal and expert agent. B. E. J. Theor. Econ. 9 (Online). http://dx.doi.org/10.2202/1935-1704.1528. ISSN 1935-1704.
Melumad, N., Mookherjee, D., Reichelstein, S., 1992. A theory of responsibility centers. J. Acc. Econ. 15, 445–484.
Melumad, N., Reichelstein, S., 1989. Value of communication in agency. J. Econ. Theory 47, 334–368.
Mirrlees, J., 1971. An exploration in the theory of optimum income taxation. Rev. Econ. Stud. 38, 175–208.
Mookherjee, D., Reichelstein, S., 2001. Incentives and coordination in hierarchies. B. E. J. Theor. Econ. 1 (Online), 1534–5963. http://dx.doi.org/10.2202/1534-5963.1009.
Myerson, R., 1982. Optimal coordination mechanisms in generalized principal–agent problems. J. Math. Econ. 10, 67–81.
Palmer, J., 2003. Relative convexity. Mimeo, ECE Dept., UCSD, Tech. rep.
Rochet, J.C., Choné, P., 1998. Ironing, sweeping, and multidimensional screening. Econometrica 66, 783–826.
Rogerson, W., 1987. On the optimality of linear contracts. Mimeo, Northwestern University.
Rogerson, W., 2003. Simple menus of contracts in cost-based procurement and regulation. Amer. Econ. Rev. 93, 919–926.
Ruszczynski, A., 2006. Nonlinear Optimization. Princeton University Press.
Shi, X., 2012. Optimal auctions with information acquisition. Games Econ. Behav. 74, 666–686.
Sobel, J., 1993. Information control in the principal–agent problem. Int. Econ. Rev. 34, 259–269.
Spinnewijn, J., 2013. Insurance and perceptions: how to screen optimists and pessimists. Econ. J. 123, 606–633.
Szalay, D., 2005. The economics of clear advice and extreme options. Rev. Econ. Stud. 72, 1173–1198.
Szalay, D., 2009. Contracts with endogenous information. Games Econ. Behav. 65, 586–625.
Terstiege, S., 2012. Endogenous information and stochastic contracts. Games Econ. Behav. 76, 535–567.
Villeneuve, B., 2005. Competition between insurers with superior information. Europ. Econ. Rev. 49, 321–340.