WORKING PAPERS

N° 1739

APRIL 2026

"On The (Relative) Merits of Money Burning For Optimal Contracting"

Javier Gonzalez-Morin and David Martimort

Toulouse School of Economics

On The (Relative) Merits of Money Burning For Optimal Contracting*

Javier Gonzalez-Morin† and David Martimort‡

April 21, 2026

Abstract

We consider the following screening model of procurement. An agent (the seller) has private information on his cost parameter. A principal (the buyer) learns an ex post signal on this parameter. The signal is private information to the principal and proper incentives to reveal this signal must be designed. In related contexts, money burning, i.e., the *ex post* destruction of some of the gains from trade, has shown to be useful to provide such incentives. We demonstrate that money burning allows the principal to implement the first-best output with zero information rent for the agent; although it is never optimal to do so since output distortions are less costly. More generally, money burning is rarely optimal, and only used as a tool of last resort if output distortions are no longer feasible. In particular, when output must be chosen before the non-verifiable signal realizes, money burning becomes more attractive.

Keywords. Optimal contracting, asymmetric information, ex post signal, money burning.

JEL codes. D82.

1. Introduction

We consider the following screening model of procurement. The principal (the buyer, *she*) contracts for the delivery of good or service. The agent in charge (the seller, *he*) has

*We thank participants to the TSE Microeconomic Theory Workshop, SAEe 2023 and EEA Congress 2025 for useful comments. This paper has also been funded by the Agence Nationale de la Recherche under grant ANR-17-EURE-0010 (Investissements d'Avenir program). The usual disclaimer applies.

†Toulouse School of Economics, Toulouse, javier.gonzalez-morin@tse-fr.eu

‡Toulouse School of Economics, Toulouse, david.martimort@tse-fr.eu

private information on his cost parameter. In this adverse selection context, the principal should offer a menu of contracts to maximise her surplus and let the agent select the right option within the menu according to his type. However, achieving successful self-selection requires that the principal leaves some information rent to the agent. To reduce this rent that she views as costly, the principal must distort quantities downwards. The so called *rent/efficiency* trade-off as exemplified in Baron and Myerson (1982) is at play.

Of course, a more direct access to information would facilitate the principal's task in extracting more information rent from her agent and could thus be useful to better regulate the agent. As a result, lesser output distortions might be implemented; and the aforementioned rent/efficiency trade-off would be tilted towards efficiency. Benchmarking of the seller's performances with those of related sellers on the market place, or even more extreme forms of yardstick information where individual compensations are linked to peers' performances are all information management techniques that are often used in practice to this end and that fall under this umbrella.

On the theory side, the benefit of an additional ex post public signal, which is correlated with the agent's type and can be used by a principal to fully extract the agent's information rent is well documented, at least since the work of Riordan and Sappington (1988).[1] When the principal has access to an ex post public signal on the agent's cost parameter, the optimal contract can be designed so as to fully extract the agent's information rent and implement the first-best.[2] Contingent payments that depend on how the agent's report is confirmed or not by the principal's signal suffice to do the trick.

Much less is known when the principal ex post signal is non-verifiable, or private information. Indeed, there are many instances where the principal may obtain information that cannot be verified by a Court of Law. For instance, there could be a leakage of information, such as insider knowledge being revealed by an employee of the seller's unit, that makes the principal learn about the agent. In such scenario, the principal cannot write down an explicit contract contingent on the realized signal. The principal may behave opportunistically and report whatever signal would minimize her payment to the agent. The principal thus faces the challenge of designing ex ante a contract that prevents her own opportunism ex post. In addition to the more standard incentive compatibility constraints on the agent, incentive constraints on the principal's side must also be imposed to induce her to truthful report whatever signal she has learned. In our context, simple sell-out contracts that leave the principal's payoff unchanged across signal realizations and make the agent residual claimant for the gains from trade solve the credibility

[1] More recent works on this front are Kessler et al. (2005), Krähmer (2020) and Kattwinkel and Knoepfle (2023).

[2] The logic is also familiar from the work of Crémer and McLean (1985), Crémer and McLean (1988), McAfee and Reny (1992), Lopomo et al. (2022) and Rahman (2012a) in multi-agent environments.

problem.

Two scenarios could then envisioned. In the first scenario, there is no waste of money and whatever is paid by the principal for the agent's services ends up being pocketed by the agent. This requirement is akin to assuming that trade is balanced, or that any putative waste of resource would be renegotiated away on efficiency grounds. Solving simultaneously with the sole use of a single payment scheme the principal's and the agent's incentive problems may be difficult in those circumstances.[3] Breaking the budget could thus be attractive; the second scenario that is the focus of this paper. Money burning, since it creates a wedge between what the principal pays for the agent's services and what the latter receives, precisely performs this task. In practice, money burning can be viewed as a metaphor for the costs of conflicts arising from contradictory information, costly and useless actions, or artificial costs implemented to prevent voluntary resignations or contractual terminations.

A priori, one could think that this extra degree of freedom in contracting might allow to implement cheaper incentives and thereby improve the terms of the rent/efficiency trade-off. Money burning gives an extra possibility for keeping the principal's payoff constant across realizations of her signal, satisfying thereby her incentive constraints, without changing the agent's payments and thus perturbating his own incentives. As a result, money burning could allow to reduce output distortions and lower informational rents. To illustrate, we demonstrate that money burning certainly allows to implement first-best output while it permits full rent extraction.

Although attractive and intuitive, this line of reasoning stressing the beneficial role of money burning is somewhat incomplete. Money burning may just not be optimal, even when allowed. The waste of surplus might be too large. The intuition is the following. Remember that the principal's incentives hold provided that her net payoff is constant over possible realizations of her signal. Without money burning, this constancy can be achieved by varying at the same time payments and outputs. A higher payment to the agent in case the observed signal confirms the agent's report should then be compensated by a greater output so as to leave the principal's net surplus unchanged. Under broad circumstances, the value of contracting is positive at a given pair of signal-cost realizations and it is always better to use such variations in output and payment rather than just burn money. Under weak conditions (technically, the principal's problem should be regular, i.e., local incentive compatibility constraints are sufficient for global incentive compatibility and the agent's participation constraint binds for the worst realization of his cost parameter), it is never optimal to burn money. Adjusting output requirements

[3] This tension is reminiscent of the moral hazard in teams problem à la Holmström (1982) or to the difficulty of satisfying requirement of budget balance in mechanism design environment as in Laffont and Maskin (1980).

and payments to the principal's signal suffice. Fine tuning production and payment is always preferable as long as surplus has a positive social value *even when taking incentive constraints* into account.

Of course, adjusting output itself might not be always feasible. It is especially so when production is a commitment that requires investments or dedicated assets that cannot be acquired or sold at wish along the production process. In those circumstances, output can certainly be contingent on the agent's report early on but might not be on the principal's own claim ex post, once she learns her signal. Together with contingent payments, money burning becomes a complementary tool to write a contract contingent on the signal. We show in fact that money burning puts the agent under countervailing incentives (Lewis and Sappington, 1989a,b). Exaggerating his cost allows the agent to produce a lower quantity targeted to a less efficient type at a lower cost; saving thereby an information rent. Yet, exaggerating cost makes it more likely that the principal's signal strongly conflicts with this claim; which in turn increases the probability of money burning. Since money burning destroys surplus and the agent ends up being residual claimant for that surplus, increasing that probability reduces the benefits of exaggerating costs. Money burning then ends up being useful on two grounds. First, it might allow to extract more rent from the agent while keeping the same output distortions as absent any signal. This scenario arises when the incentive problem remains regular. The sole role of money burning is that it reduces the payment for the agent's services but the standard Baron-Myerson output distortion prevails. Second, money burning cannot only extract more rent but also reduces output distortions. This case arises when countervailing incentives are strong enough to keep a whole subset of the agent's types indifferent between participating or not. To prove the existence of such milder distortions, we use techniques from the principal-agent literature with type-dependent participation constraints, especially the work of Jullien (2000) and Martimort and Stole (2022).

Literature Review. Our paper borrows its framework from Riordan and Sappington (1988). Those authors have also developed a simple procurement model with the additional ingredient that the principal receives an ex post signal correlated with the agent's private information. There, the signal is verifiable and contractible, allowing for first-best implementation and full rent extraction under fairly weak conditions on the conditional distribution of the signal and the agent's cost function. The intuition is that correlation between the ex post signal and the agent's type can be used to cross-check his report with the observed signal and punish (at least stochastically) any discrepancy.[4] This paper extends this basic environment on two fronts. First, the ex post signal is non-verifiable;

[4]This implementation of the first-best bears some ressemblance with the results of Crémer and McLean (1985), Crémer and McLean (1988) and McAfee and Reny (1992) in multi-agent environments. The difference is that in those environments, there is no ex post signal and cross-checking arises by comparing agents' reports.

which *a priori* prevents first-best implementation. Second, we elaborate on the potential benefits of money burning under those circumstances.

On this front, Strausz (2006) also considered a pure adverse selection procurement model with binary signals and types. This author shows how the non-verifiability of the signal prevents the principal from implementing the first-best output with full rent extraction but can still be useful to improve the rent-efficiency trade-off even when budget balance is maintained. Dequiedt and Martimort (2015) also considered an adverse selection setup in which two agents have correlated types which are drawn from a continuous distribution.[5] The principal must run bilateral contracts with each of them, so that the report made by one of the agents cannot be publicly used for the allocation regarding the other agent. However, because the principal may still use the information that she gathers, the contract allows for principal's discretion in the last step. In both Strausz (2006) and Dequiedt and Martimort (2015), the benefits of using the signal, though non-verifiable, is that the agent's virtual cost (in the Myersonian parlance) can now be tailored to its precise realization; which lead to implement output contingent on this signal. Neither Strausz (2006) nor Dequiedt and Martimort (2015) consider the possibility of burning money. Because the present paper provides conditions under which money burning is actually useless, we also offer in passing a rationale for the approach taken by this earlier literature.

Subjective evaluations have also been an important and lively topic for managerial economics over the recent years. More specifically, departing from the benchmark à la Holmström (1979) where principals may rely on objective measures of performances to solve a moral hazard problem, several authors have investigated the design of compensation schemes when only non-verifiable measures of performances are available. Here, the work of MacLeod (2003) stands as a leading reference.[6] This author studies a moral hazard environment in which agent's performance is private information to the principal. Preventing the principal's opportunism might thus call for a fixed payment that dampens any incentives on the agent's side. Money burning is not only sufficient but also necessary to provide incentives to both parties. Since money burning is costly, it is not used unless in the worst state where the agent's performance is the lowest. There is a form of wage compression but it comes also with reduced effort in comparison with the scenario of an objective performance measure. In contrast, a screening environment has (generally) more contracting variables available. The principal's incentive constraints can also be adjusted by modifying trading volume; which might make money burning just suboptimal. This result bears thus some resemblance with Fuchs (2007) who studied a repeated moral hazard environment. His main finding is that the principal can rely

[5] See also Celik et al. (2021).

[6] Relevant references on this front also include Baker et al. (1994), Rajan and Reichelstein (2006), Rahman (2012b), Zábojník (2014) and Cheng (2021) among others.

on the threat of terminating the contract with the agent instead of wasting payments. Termination destroys the remaining surplus form the relationship, even if such surplus is positive. However, the principal must rely on such threat to provide incentives to the agent.[7] Along the same vein, Lang (2023) demonstrates that stochastic compensations may be more attractive than money burning because a risk-averse agent may incur some disutility from that randomness which has a similar impact to money burning.

Finally, Khalil et al. (2015) study a mixed model with both adverse selection and moral hazard elements. The principal can there also learn a non-verifiable signal and money burning is allowed. Those authors demonstrate that the date at which the non-verifiable signal is learned matters a great deal. If the effort decision is taken after the signal is realised, money burning is *a priori* optimal but rescaling the project in response to the signal may be a more attractive strategy. This result is reminiscent of the sub-optimality found for money burning in our pure adverse selection contexts. Yet, we also demonstrate the attractiveness of money burning when production is committed earlier on.

Organization of the paper. Section 2 presents our model. Section 3 presents several benchmarks: the complete information scenario, the case where no signal is available (which boils down to Baron and Myerson (1982)'s well known model), the case where an ex post signal is available but it can be contracted upon. Section 4 describes the set of incentive feasible allocations when the ex post signal is non-verifiable. Section 5 characterizes the optimal contract when output can be made contingent on the principal's report on her signal. Section 5.1 shows that money burning can be used to implement the first best outcome but Section 5.2 argues that it is not optimal when the principal's problem is regular. Section 6 characterizes the optimal contract when output cannot be made contingent on the principal's report on her signal. Money burning may now be attractive in case of strong conflicts between the principal's and the agent's assessments of costs. We again characterize the optimal contract under those circumstances and show the role of money burning as vehicle for countervailing incentives that might facilitate rent extraction but also ease production. Proofs are relegated to Appendix A.

2. The Model

Technology and Preferences. We consider the following simple model of procurement. A principal (thereafter *she*) wants to purchase a good or service from an agent (*he*). The principal derives utility $v(q,t)$ from purchasing a non-negative quantity q at price y:

$$v(q,y) = S(q) - y$$

[7] A similar result can also be found in Levin (2003).

where the gross surplus function S is twice continuously differentiable, increasing and strictly concave ($S' > 0 > S''$), and satisfies the Inada conditions $\lim_{q \to 0} S'(q) = +\infty$, $\lim_{q \to \infty} S'(q) = 0$ and $S(0) = 0$. Let $\mathcal{Q} \subseteq \mathbb{R}_+$ be the set of possible quantities and $\mathcal{T} \subseteq \mathbb{R}$ is the set of possible payments.

The agent has a quasi-linear utility function:

$$u(q, t; \theta) = t - \theta q$$

where θ is his marginal cost of production and t stands for the agent's reward, which *a priori* should be lower than the principal's payment when money burning is allowed.

Information Structure. The agent has private information on the cost parameter θ. This parameter is drawn from a distribution according to a cumulative distribution function F, atomless and continuously differentiable, admitting a density function f and with support $\Theta = [\underline{\theta}, \overline{\theta}]$ with $0 < \underline{\theta} < \overline{\theta}$. As it is standard in most of the screening literature,[8] we shall impose the following familiar assumption:

Assumption 1. Monotone Hazard Rate Property:

$$\frac{F(\theta)}{f(\theta)} \quad \textit{non-decreasing.}$$

This standard procurement model is augmented by adding the possibility for the principal to learn an informative signal σ on the agent's type. The set Σ of possible signal realizations of this signal is a compact interval of the real line, with its lowest and its maximal elements respectively denoted by $\underline{\sigma}$ and $\overline{\sigma}$. Let $\Delta(\Sigma)$ be the set of probability distributions on Σ and let $g(\sigma|\theta) \in \Delta(\Sigma)$ be the distribution of signals conditional on θ. This conditional distribution is assumed to have full support so that no signal may reveal a given type with certainty.

Following an approach already taken in MacLeod (2003), this signal σ stands for a subjective evaluation of the agent's performance. It is private information to the principal and thus non-verifiable by any third party. In a moral hazard context with subjective evaluation, MacLeod (2003) argues that a money burning is a useful tool to provide incentives to both the principal and the agent. Accordingly, we follow this author and allow the principal's payment, y, and the agent's rewards t to differ, i.e., $y = t + b$, with $b \in \mathcal{B} = [0, B]$ with B being an upper bound on the amount of money that can be burned.

Assumptions. Let denote by $g_\theta(\sigma|\theta)$, the derivative with respect to the cost parameter

[8] Bagnoli and Bergstrom (2005).

θ of the the conditional density of signal. For future reference, the conditional cumulative distribution will be denoted as $G(\sigma|\theta) = \int_{\underline{\sigma}}^{\sigma} g(\tilde{\sigma}|\theta)d\tilde{\sigma}$. For technical reasons, we shall assume that g is twice continuously differentiable in θ and that the score $\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}$ is absolutely continuous in θ. Furthermore, we shall also impose the following conditions.

Assumption 2. Monotone Likelihood Ratio Property:

$$\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)} \quad \textit{non-decreasing in } \sigma, \quad \forall \theta \in \Theta.$$

Assumption 2 is a familiar requirement that ensures that higher values of the signal σ are more likely to arise when the agent's true type increases. It is well known that this condition also implies first-order stochastic dominance, $G_\theta(\sigma|\theta) \leq 0$, i.e., as the agent's type increases it becomes less likely that signals become good news. Next condition focuses on scenarios where there is some form of decreasing returns in this respect.

Assumption 3. Convexity of the Conditional Cumulative Distribution Function:

$$G_{\theta\theta}(\sigma|\theta) \geq 0, \quad \forall \sigma \in \Sigma, \, \forall \theta \in \Theta$$

Convexity of the conditional distribution of signal is a more technical assumption needed at several places below to ensure that global incentive compatibility constraints always hold.

Example 1. *As a running example, we take θ to be uniform on $[\underline{\theta}, \overline{\theta}]$:*

$$F(\theta) = \frac{\theta - \underline{\theta}}{\overline{\theta} - \underline{\theta}} \quad \textit{and} \quad f(\theta) = \frac{1}{\overline{\theta} - \underline{\theta}} \qquad \textit{for } \theta \in [\underline{\theta}, \overline{\theta}],$$

and the signal distribution conditional on θ is:

$$g(\sigma|\theta) = (1 - \alpha) + \alpha\left(2\sigma F(\theta) + 2(1 - \sigma)[1 - F(\theta)]\right) \qquad \textit{for } \sigma \in [0, 1],$$

where $\alpha \in [0, 1]$ is a measure of the informativeness of the signal: if $\alpha = 0$, the signal is uniformly distributed on $[0, 1]$; if $\alpha > 0$, the signal σ is positively correlated with θ. In particular, Assumptions 1, 2 and 3 are satisfied.

Moreover, we specify $S(q) = 2q^{1/2}$, so that Inada conditions are satisfied.

Terminology. Since the relevant incentive problem is to deter over-reporting of the cost parameter θ, we define news relative to the agent's report. A high realization of σ will be called *good news*, as it is more consistent with a high reported cost and therefore makes over-reporting less likely. A low realization of σ will be called *bad news*, as it is more indicative of a low cost parameter and therefore more suggestive of over-reporting. Thus, "good" and "bad" refer to consistency with the report, not to whether the signal is favorable or unfavorable about the true type.

Mechanisms. A direct revelation mechanism (*DRM*) in our environment is an allocation rule $\{(q(\hat{\theta},\hat{\sigma}), t(\hat{\theta},\hat{\sigma}), b(\hat{\theta},\hat{\sigma}))\}_{\hat{\theta}\in\Theta,\hat{\sigma}\in\Sigma}$, that stipulates a quantity $q(\hat{\theta},\hat{\sigma})$, a payment $t(\hat{\theta},\hat{\sigma})$ to the agent and an amount of money to be burnt $b(\hat{\theta},\hat{\sigma})$ as a function of the agent's report $\hat{\theta}\in\Theta$ on her type and the principal's report on his signal $\hat{\sigma}\in\Sigma$.

Timing. The game unfolds as follows.

1. The principal proposes a *DRM*, namely $\{(q(\hat{\theta},\hat{\sigma}), t(\hat{\theta},\hat{\sigma}), b(\hat{\theta},\hat{\sigma}))\}_{\hat{\theta}\in\Theta,\hat{\sigma}\in\Sigma}$.

2. The agent learns θ.

3. The agent accepts or rejects the mechanism. Following acceptance, the agent sends a report $\hat{\theta}\in\Theta$ to the mechanism. Otherwise, the principal and the agent both get their reservation values that are normalized at zero.

4. The principal observes the report $\hat{\theta}$, learns the signal σ, and then reports $\hat{\sigma}\in\Sigma$ to the mechanism.[9,10]

5. The allocation $(q(\hat{\theta},\hat{\sigma}), t(\hat{\theta},\hat{\sigma}), b(\hat{\theta},\hat{\sigma}))$ is implemented. Payoffs are realized.

A *DRM* as described above, with the sequential timing of communication stands as a metaphor for other indirect mechanisms that would appear as more natural in practice but may be more complex to analyze without the apparatus of direct communication. To illustrate, the production process may be a dynamic one, with earlier production stages producing a signal on the agent's underlying cost that is privately observed by the principal. The agent's report in the direct communication game corresponds to the choice of a whole menu of production plans before the informative signal realizes and is learned by the principal in the indirect version. Such a plan stipulates a final output, a payment received the principal and the amount of money burnt as a function of this output. A menu is a family of such plans indexed by the signal realization. Finally, the principal picks an item within a menu of such contingent plans according to the signal she has learned. To reflect the sequential turn of actions in the indirect version of the game,

[9]The exact timing has no impact. It could be that the principal learns her signal before observing the agent's report with no change in the optimal contract.

[10]Our timing is such that the principal learns her private information after having designed the mechanism; in contrast with Cella (2008) who analyzed an informed principal's problem with correlated information.

communication in the direct version is also sequential. As proved in a related context by Dequiedt and Martimort (2015), the Revelation Principle still applies in our procurement context provided incentive constraints are properly defined for both the principal and the agent.[11]

In the first part of the paper, we will consider an environment where such output plans, contingent on the principal's report on his signal are feasible. In practice, it corresponds to scenarios where production is an ongoing process that can easily be modified. Output adjustments may be useful to better screen the agent's cost parameter. In Section 6, we consider an alternative timing where the signal is learned after output has been committed. This scenario is meant to capture environments where the agent must incur some (unmodeled) investment (productive assets) that determines the scale of production and, in the short run, no additional investment or no sale of existing assets is feasible.

3. BENCHMARKS

COMPLETE INFORMATION. When θ is common knowledge, there is no point using a subjective evaluation and money burning is useless. A simple forcing contract can be used to induce the first-best output and extracts all surplus from the agent. The first-best outcome entails:

$$S'(q^{fb}(\theta)) = \theta, \quad U^{fb}(\theta) = 0, \quad b^{fb}(\theta) = 0, \quad \forall \theta \in \Theta. \tag{3.1}$$

For future reference, let $W^{fb}(\theta) = S(q^{fb}(\theta)) - \theta q^{fb}(\theta)$ be the first-best surplus and let $\mathcal{W}^{fb} = \mathbb{E}_\theta(W^{fb}(\theta))$ its expected value.

ASYMMETRIC INFORMATION WITH NO SIGNAL. This scenario boils down to the standard procurement model à la Baron and Myerson (1982). In particular, there is no need for money burning while the rest of the solution is well known and highlights the main features of the rent-efficiency trade-off of screening models:

$$S'(q^{bm}(\theta)) = \theta + \frac{F(\theta)}{f(\theta)}, \quad U^{bm}(\theta) = \int_\theta^{\bar{\theta}} q^{bm}(x)dx \geq 0, \quad b^{bm}(\theta) = 0, \quad \forall \theta \in \Theta. \tag{3.2}$$

There is a downward distortion of output for all types except at the top and all types except the least-efficient one receives a positive information rent. Assumption 1 ensures that q^{bm} so defined is non-increasing as requested by incentive compatibility while the assumptions made on the surplus function ensure that q^{bm} is always positive so that all types always produce.

[11] See Appendix B for details.

Asymmetric Information with a Verifiable Signal. Suppose now that the signal σ is verifiable and thus can be contracted upon. Clearly, money burning is useless in this context. Money burning wastes resource and it has no benefits whatsoever in relaxing the agent's incentive constraints; payments contingent on σ suffices.

From Riordan and Sappington (1988), and following the logic of Crémer and McLean (1988) and McAfee and Reny (1992)'s more general results, it is well-known that the first-best outcome can be be implemented with no surplus left to the agent under weak assumptions on the distribution of the verifiable signal. To see how, consider the following payment schedule

$$t(\hat{\theta},\sigma) = S(q^{fb}(\hat{\theta})) - W^{fb}(\hat{\theta}) + z(\hat{\theta},\sigma) \quad \forall \hat{\theta} \in \Theta, \forall \sigma \in \Sigma. \tag{3.3}$$

Those payments make the agent residual claimant for the choice of the output $q^{fb}(\hat{\theta})$ but request him to pay a fee $W^{fb}(\hat{\theta})$ independent of σ that is designed to extract the agent's surplus. This fee is augmented by an extra payment $z(\hat{\theta},\sigma)$ contingent on the realized signal. This payment corrects any stochastic discrepancy that may arise between the agent's report $\hat{\theta}$ and the realized ex post signal.

Proposition 1. *Suppose that σ is verifiable and that Assumption 3 holds. There exists an optimal contract that implements the first-best outcome (3.1). In such optimal contract, money burning is useless and the transfers described in (3.3) are such that:*

$$z(\hat{\theta},\sigma) = q^{fb}(\hat{\theta}) \frac{g_\theta(\sigma|\hat{\theta})}{g(\sigma|\hat{\theta})} \frac{1}{\mathbb{E}_\sigma\left(\left(\frac{g_\theta(\sigma|\hat{\theta})}{g(\sigma|\hat{\theta})}\right)^2 \middle| \hat{\theta}\right)} \tag{3.4}$$

The first-best implementation when σ is verifiable relies on the possibility for the principal to offer lotteries $z(\hat{\theta},\sigma)$ that tie the agent's ex post payment to the realization of the signal; punishing the agent when the ex post signal conflicts too much with his report and rewarding otherwise. In particular, such lotteries can be reinterpreted as the principal estimating the type of the agent by maximum likelihood methods, and comparing this estimate with the agent's report. The score $s(\sigma|\theta) \equiv \frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}$ indicates the sign of the gradient of the log-likelihood with respect to θ. If the agent reports $\hat{\theta}$ and the realized signal is such that $s(\sigma|\hat{\theta}) < 0$, then the principal would find that the likelihood would be higher with a lower report of the type. Since the main incentive problem is to prevent over-reporting the type, the principal must punish such combination of signal and report, and thus $z(\hat{\theta},\sigma) < 0$. On the contrary, the principal rewards the agent with $z(\hat{\theta},\sigma) > 0$ when $s(\sigma|\hat{\theta}) > 0$, and there is neither a punishment nor a reward when the reported type of the agent matches the maximum likelihood estimate based on σ. The size of rewards and punishments is scaled by the inverse of the Fisher information $\mathcal{I}(\theta) \equiv \mathbb{E}_\sigma\left(\left(\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\right)^2 \middle| \theta\right)$, which is the variance of the score. If the signal contains little information about θ, the principal must use extreme lotteries to prevent misreporting. Thus, the higher this information, the smaller the scale of the transfers for any realization of the signal.

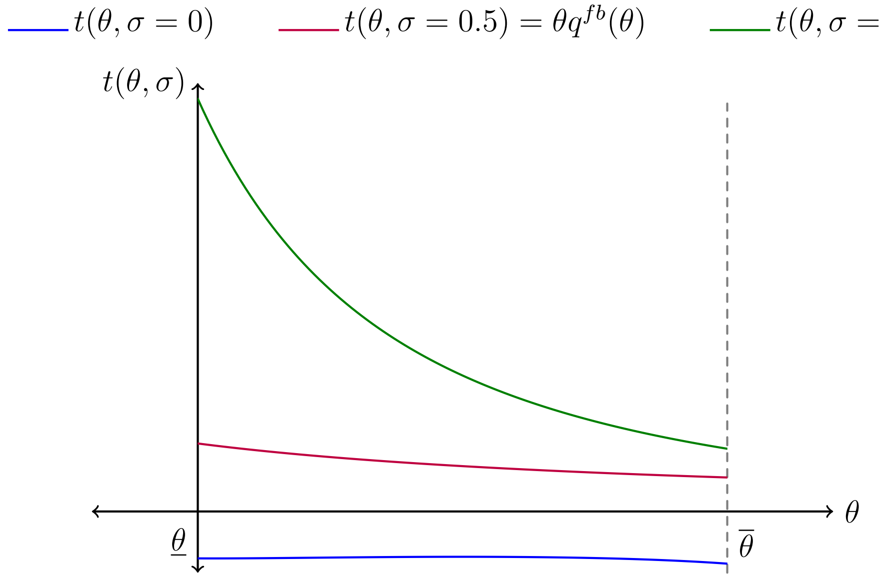


Figure 1: Transfers that implement the first-best allocation of output and zero rent when σ is verifiable. For $\sigma = 0.5$, the score of the signal is zero, which corresponds to a case with neither good nor bad news: the agent is compensated for the total cost of production $\theta q^{fb}(\theta)$. For $\sigma = 0$, the principal punishes the agent with a negative transfer, in order to disincentivize over-reporting. For $\sigma = 1$, the principal rewards the agent with higher transfers.

4. Incentive-Feasible Allocations

When deciding whether to accept the mechanism or not, the agent must form an expectation about the possible reports of the principal. Expecting a truthful report of the principal's subjective signal, the agent's payoff must be computed taking expectation over σ conditional on the realised type θ. Accordingly, we define the agent's *interim rent* when his type is θ as

$$U(\theta) = \mathbb{E}_\sigma \left(t(\theta, \sigma) - \theta q(\theta, \sigma) \mid \theta \right).$$

The agent's participation constraint can thus be written as

$$U(\theta) \geq 0, \quad \forall \theta \in \Theta. \tag{4.1}$$

Being given that the principal will report truthfully her subjective signal, the agent's incentive-compatibility constraints write as

$$U(\theta) = \max_{\hat{\theta} \in \Theta} \mathbb{E}_\sigma \left(t(\hat{\theta}, \sigma) - \theta q(\hat{\theta}, \sigma) | \theta \right), \quad \forall \theta \in \Theta. \tag{4.2}$$

Additional incentive compatibility constraints must be incorporated to indeed induce the principal to truthfully report her signal. Importantly, note that these constraints are written ex post, once the agent has reported (truthfully in equilibrium) her type θ. For any such report, the principal must prefer not to manipulate her subjective evaluation of the agent; a condition written as

$$S(q(\theta,\sigma)) - t(\theta,\sigma) - b(\theta,\sigma) \geq S(q(\theta,\hat{\sigma})) - t(\theta,\hat{\sigma}) - b(\theta,\hat{\sigma}), \quad \forall\theta \in \Theta, \forall(\sigma,\hat{\sigma}) \in \Sigma^2. \tag{4.3}$$

Because σ is not payoff-relevant, the role of σ and $\hat{\sigma}$ can be reversed. In fact, the principal should be made indifferent between any of her available reports and (4.3) amounts to the existence of a function $H(\theta)$ that correspond to the constant value of the principal's net payoff over all realizations of her signal:

$$S(q(\theta,\sigma)) - t(\theta,\sigma) - b(\theta,\sigma) = H(\theta) \quad \forall\theta \in \Theta, \forall\sigma \in \Sigma. \tag{4.4}$$

The principal is thus willing to report truthfully as she is indifferent between any of her available reports. The contract takes the form of a sell-out contract. Anticipating such truthful reporting, the agent is right to evaluate his own incentives to report the truth and participate with the true expectation on signal conditional on his type.

The principal's incentive constraint (4.4) deserves further comments. In a standard moral hazard à la MacLeod (2003), the principal can only use transfers to incentivize her agent. She can reduce the wage being paid by reporting a particular realization of the subjective signal on the agent's performance and she would thus always choose that very report. But then the payment schedule can no longer depend on performance and incentives to exert effort can no longer be provided. Money burning is thus necessary to simultaneously induce truthful reporting by the principal and incentivize effort by the agent. In our adverse selection setup, money burning is no longer necessary. It could still be possible to satisfy (4.4) without burning money (i.e., setting $b(\theta,\sigma) \equiv 0$). In particular, different combinations of outputs and transfers $(q(\theta,\sigma), t(\theta,\sigma))$ could be implemented for different realizations of σ the principal's payoff is kept constant. Hence, money burning is not necessary in a screening environment.

Using the expression of the transfers $t(\theta,\sigma)$ coming from (4.4), the agent's incentive compatibility constraints can be written in terms of interim payoffs as

$$U(\theta) = \max_{\hat{\theta}\in\Theta} -H(\hat{\theta}) + \mathbb{E}_\sigma\left(S(q(\hat{\theta},\sigma)) - \theta q(\hat{\theta},\sigma) - b(\hat{\theta},\sigma)|\theta\right), \quad \forall\theta \in \Theta. \tag{4.5}$$

This expression makes it clear that the agent's is made residual claimant for the maximization of the overall surplus $S(q) - \theta q - b$, choosing among production plans $q(\hat{\theta},\sigma)$

contingent on the ex post signal σ that each corresponds to a payment back to the principal worth $H(\hat{\theta})$.

The Principal's Problem. Expressing also the maximand in terms of the agent's interim rent, the principal's problem is to choose an allocation $(q(\theta,\sigma), b(\theta,\sigma), U(\theta))_{(\theta,\sigma)\in\Theta\times\Sigma}$ that solves

$$(\mathcal{P}) : \max_{(q,b,U)} \quad \mathbb{E}_{\theta,\sigma}\left(S(q(\theta,\sigma)) - \theta q(\theta,\sigma) - b(\theta,\sigma) - U(\theta)\right)$$

subject to (4.1), (4.5) and

$$b(\theta,\sigma) \in [0, B], \quad \forall(\theta,\sigma) \in \Theta \times \Sigma. \tag{4.6}$$

This expression of the principal's maximization problem showcases an extension of the classic rent-efficiency trade-off. The principal would like to maximise the surplus that is created and reduce the rent that is left to the agent. In sharp contrast with the standard procurement model, part of the surplus can here be burnt. Yet, money burning can also be used to relax the agent's incentive compatibility constraint (4.5).

Necessary Conditions for Incentive Compatibility. A standard approach in mechanism design consists in studying a relaxed maximization problem where the whole set of incentive compatibility constraints (4.5) (whose description already encompasses the characterization of the principal's incentive compatibility constraints (4.4) is replaced by a necessary Envelope Condition. Next Lemma expresses this condition.

Lemma 1. *An allocation $(q(\theta,\sigma), b(\theta,\sigma), U(\theta))$ satisfying (4.5) is necessarily such that U is absolutely continuous and thus a.e. differentiable with a derivative when it exists given by*

$$\dot{U}(\theta) = \mathbb{E}_{\sigma}\left(-q(\theta,\sigma) + (S(q(\theta,\sigma)) - \theta q(\theta,\sigma) - b(\theta,\sigma))\frac{g_{\theta}(\sigma|\theta)}{g(\sigma|\theta)}\,\middle|\,\theta\right). \tag{4.7}$$

5. When the Signal Is Learned before Production

5.1. On the (Possible) Value of Money Burning

The full surplus-extraction result of Proposition 1 that holds even in the absence of any money burning crucially depends on the fact that the signal σ was supposed to be verifiable. Next Proposition shows that with non-verifiability of the signal the first-best can no longer be implemented unless money is burned.

Proposition 2. *Suppose that σ is non-verifiable. The first-best allocation $(q^{fb}(\theta), U^{fb}(\theta))$*

1. *cannot be implemented without money burning;*
2. *can be implemented with money burning when Assumption 3 holds and B is large enough. It is so by using money burning following bad news, i.e.,*

$$b(\theta,\sigma) = \begin{cases} B & \text{if } \sigma \in [\underline{\sigma}, \sigma^{fb}(\theta)), \\ 0 & \text{if } \sigma \in (\sigma^{fb}(\theta), \overline{\sigma}], \end{cases} \tag{5.1}$$

where $\sigma^{fb}(\theta)$ solves

$$-q^{fb}(\theta) = BG_\theta(\sigma^{fb}(\theta)|\theta). \tag{5.2}$$

The intuition for Item 1. is straightforward and can easily be grasped from the expression of the transfers in (3.3). Lotteries are no longer possible when the principal can manipulate her report on the realized signal and money burning is not feasible. The principal would just report the signal with the lowest such transfer $\min_{\sigma\in\Sigma} z(\hat{\theta},\sigma)$; which voids lotteries of their value.

Item 2. shows that, if enough money can be burned, full extraction of the agent's surplus is possible and the first-best output can be implemented provided that the principal burns the largest possible amount B whenever the signal she observes is sufficiently bad news relative to the agent's report. To understand this result, observe that the agent wants to overstate his cost parameter θ and claim he has a cost $\theta + d\theta$ to produce a lower quantity $q^{fb}(\theta + d\theta)$ at a lower marginal cost and thereby save $q^{fb}(\theta + d\theta)d\theta \approx q^{fb}(\theta)d\theta$. Of course, this effect would still be present in the absence of any ex post signal. With an ex post signal that can be manipulated by the principal, (4.4) constraint requires that the principal gets a payoff independent of the signal. In other words, the agent ends up being residual claimant not only for the whole gains from trade but also for the money burned. To create countervailing incentives[12] and thus better extract the agent's surplus, more money should thus be burned as the agent overstates his cost parameter.

5.2. On the Irrelevance of Money Burning

When The Incentive Problem Is Regular. In a first pass and to simplify the analysis, we first focus on a *relaxed* incentive problem $(\mathcal{P}^r)$ where the requirement of incentive compatibility (4.5) reduces to the necessary condition (4.7) while the participation

[12] Lewis and Sappington (1989a,b).

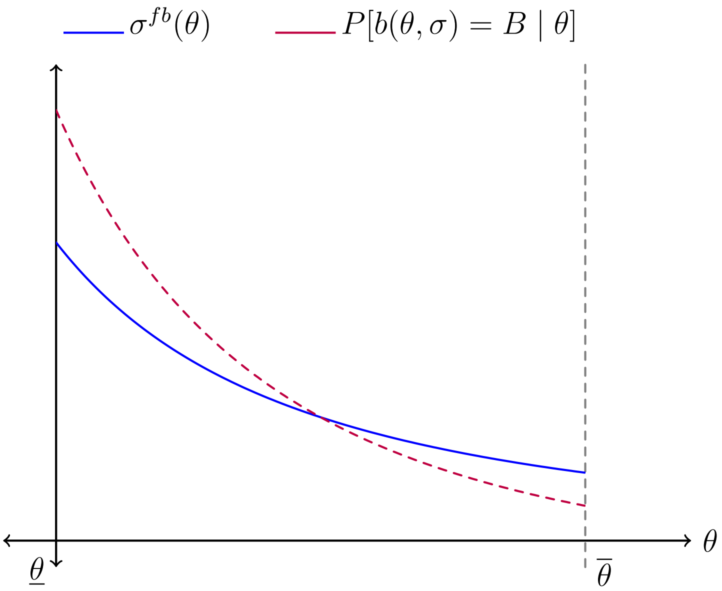


Figure 2: In blue, the threshold for the signal σ conditional on θ that determines money burning. In purple, dashed, the probability of money burning conditional on the reported type θ.

constraint (4.1) is now only imposed for the least efficient type:

$$U(\overline{\theta}) \geq 0. \tag{5.3}$$

DEFINITION 1. $(\mathcal{P})$ *is said to be regular when its solutions also solve* $(\mathcal{P}^r)$.

A key assumption to ensure regularity is the following generalization of Assumption 1.

ASSUMPTION 4. GENERALIZED MONOTONE HAZARD RATE PROPERTY:

$$\frac{\frac{F(\theta)}{f(\theta)}}{1+\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\frac{F(\theta)}{f(\theta)}} \quad \begin{cases} \textit{non-decreasing in } \theta, & \forall \sigma \in \Sigma, \\ \textit{non-increasing in } \sigma, & \forall \theta \in \Theta, \end{cases} \tag{5.4}$$

with

$$1+\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\frac{F(\theta)}{f(\theta)} > 0, \quad \forall (\theta,\sigma) \in \Theta \times \Sigma. \tag{5.5}$$

The first of those monotonicity conditions ensures that the second-best solution $q^{sb}(\theta,\sigma)$ remains non-increasing with type while the second monotonicity condition that follows from Assumption 2 shall ensure that the agent will produce more when signals are good news. Example 1 satisfies this property if $\alpha \leq \frac{1}{3}$, i.e., the signal is not too informative about the cost θ. The full characterization of the second-best solution is provided below.

Proposition 3. *Suppose that Assumption 4 holds. The optimal contract solution to* $(\mathcal{P}^r)$ *entails the following features.*

1. *The second-best output* $q^{sb}(\theta,\sigma)$ *entails a generalized Baron-Myerson distortion:*

$$S'(q^{sb}(\theta,\sigma)) = \theta + \frac{\frac{F(\theta)}{f(\theta)}}{1+\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\frac{F(\theta)}{f(\theta)}}, \quad \forall(\theta,\sigma)\in\Theta\times\Sigma. \tag{5.6}$$

   $q^{sb}(\theta,\sigma)$ *is positive, non-increasing (resp. non-decreasing) in* θ *(resp. in* σ*).*

2. *Money burning is useless:*

$$b^{sb}(\theta,\sigma) = 0, \quad \forall(\theta,\sigma)\in\Theta\times\Sigma. \tag{5.7}$$

3. *The agent's information rent is given by:*

$$U^{sb}(\theta) = \int_{\theta}^{\overline{\theta}} \mathbb{E}_\sigma\left(q^{sb}(\tilde{\theta},\sigma) - (S(q^{sb}(\tilde{\theta},\sigma)) - \tilde{\theta}q^{sb}(\tilde{\theta},\sigma))\frac{g_\theta(\sigma|\tilde{\theta})}{g(\sigma|\tilde{\theta})}\middle|\tilde{\theta}\right) d\tilde{\theta}. \tag{5.8}$$

4. *The incentive problem is regular provided that* $\Delta\theta = \overline{\theta} - \underline{\theta}$ *is small enough and*

$$1 + \frac{\frac{G_\theta(\sigma|\theta)}{g(\sigma|\theta)}\left(\frac{F(\theta)}{f(\theta)}\right)^2}{1+\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\frac{F(\theta)}{f(\theta)}}\frac{d}{d\sigma}\left(\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\right) \geq 0, \quad \forall(\theta,\sigma)\in\Theta\times\Sigma. \tag{5.9}$$

Output Distortions. To understand (5.6), it is first useful to think about the case where the signal σ would be uninformative (i.e., $g_\theta \equiv 0$). In that polar scenario, the signal has no value and the optimal output exhibits the standard Baron-Myerson distortion expressed in (3.2). With an informative signal, the principal instead would like to link the agent's payment to what she learns from the signal if the latter was verifiable and the first-best output could be implemented as shown in Proposition 1. When this signal is non-verifiable, the principal's incentive compatibility constraints (4.4) limit how flexible payments can be made to fulfil this task. Output distortions, contingent on the signal's realization, are always needed for any realization of the signal while full rent extraction is no longer possible as shown in (5.8). Because the agent's rent is costly for the principal, downward distortions of output below the first-best level are needed to reduce this rent. The logic is familiar from most screening models. Yet, details differ. The comparison of (5.6) with the standard Baron-Myerson formula (3.2) showcases the nature of those output distortions. At a rough level, the logic of screening models à la Baron and Myerson (1982) still holds; the optimal output is such that the marginal benefit of production is equal to what can be viewed as (marginal) *generalized virtual cost* of the agent.[13] Indeed, Assumption 2 implies that a signal σ such that $\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)} > 0$ (resp. < 0) is *good news* (resp. *bad news*) and it makes the principal think that the agent has not exaggerated his cost parameter. Output distortions are less (resp. more) attractive under those conditions.

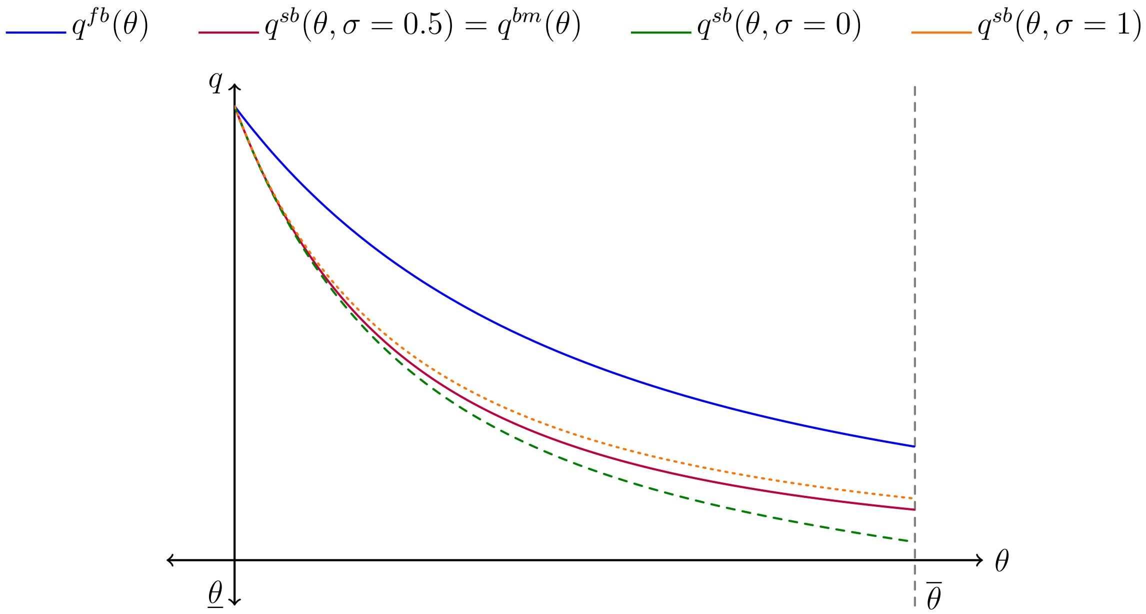


Figure 3: Output schemes. In blue, the first-best level of output. For $\sigma = 0.5$ (purple), the score $\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}$ of the signal is zero, which corresponds to a case with neither good nor bad news: output hence equals the Baron-Myerson amount. For $\sigma = 0$ (green, dashed), the principal exacerbates the downwards distortions, in order to disincentivize over-reporting. For $\sigma = 1$ (orange, dotted), the principal softens the distortions in output.

No Money Burning. Remarkably, output distortions are not only necessary but are also sufficient to solve the bilateral incentive problem between the principal and her agent. When (5.5) holds, the principal's optimization problem remains strictly concave in output. Production remains valuable while money burning is necessarily costly. Only output is used to take into account signal realizations and no money needs to be burned. For a regular problem, imposing budget balance between the principal and the agent is enough to reach the informationally constrained optimum.

The Value of the Signal. Using the fact that no money is burned, we may rewrite (4.7) as

$$\dot{U}^{sb}(\theta) = \mathbb{E}_\sigma \left(-q^{sb}(\theta,\sigma) + (S(q^{sb}(\theta,\sigma)) - \theta q^{sb}(\theta,\sigma)) \frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)} |\theta \right).$$

[13]The expression was coined by Dequiedt and Martimort (2015).

Integrating by parts the second term in this expectation yields

$$\dot{U}^{sb}(\theta) = \mathbb{E}_\sigma\left(-q^{sb}(\theta,\sigma)|\theta\right) - \mathbb{E}_\sigma\left((S'(q^{sb}(\theta,\sigma)) - \theta)q^{sb}_\sigma(\theta,\sigma)\frac{G_\theta(\sigma|\theta)}{g(\sigma|\theta)}|\theta\right). \tag{5.10}$$

For a non-informative signal, $g_\theta = G_\theta \equiv 0$, and the second term above disappears. Otherwise, the fact that q^{sb} is distorted downwards (i.e., $S'(q^{sb}(\theta,\sigma)) \geq \theta$) together with $q^{sb}_\sigma \geq 0$ and $G_\theta \leq 0$ (which are both implied by Assumption 2) implies that this second term on the right-hand side of (5.10) is always non-negative. This sign captures the fact that, overall, the principal can design countervailing incentives for the agent once the signal is available, even if it remains non-verifiable.

Average Distortions. Different realizations of the signal σ induce different distortions on output, as shown in (5.6). When the realized signal is good news, the optimal contract implements smaller distortions; while larger distortions are introduced with bad news. However, thanks to strict convexity of the distortions in the score $\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}$, Jensen's inequality allows us to derive:

$$\mathbb{E}_\sigma\left(\frac{\frac{F(\theta)}{f(\theta)}}{1+\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\frac{F(\theta)}{f(\theta)}}\middle|\theta\right) \leq \frac{\frac{F(\theta)}{f(\theta)}}{1+\mathbb{E}_\sigma\left(\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\middle|\theta\right)\frac{F(\theta)}{f(\theta)}} = \frac{F(\theta)}{f(\theta)} \tag{5.11}$$

Thus, $\mathbb{E}_\sigma\left(q^{sb}(\theta,\sigma)\right) \geq q^{bm}(\theta)$, with strict inequality provided that $g_\theta \not\equiv 0$. In other words, for each type θ, the expected distortion induced in the optimal contract is smaller when the principal has a private signal available.

"Very Bad News". As a corollary of those findings, money might only be burned when $(\mathcal{P})$ is regular if (5.5) fails for some combinations of signal and type. This possibility arises when σ is *very bad news* on the agent's cost, i.e., $g_\theta(\sigma|\theta)$ is sufficiently negative at such a combination. It characterizes a scenario of strong conflict between the principal's assessment of the agent's performance and the latter's report on his type. In Example 1, this occurs when $\alpha > \frac{1}{3}$.

To explore the consequences of this possibility for contract design, we first consider the case of a regular problem. Next Assumption, which is a strengthening of Assumption 1 gives us a clear configuration of the optimal contract under those circumstances.

Assumption 5. Mixed Monotonicity.

$$\frac{F(\theta)}{f(\theta)} + \frac{g(\sigma|\theta)}{g_\theta(\sigma|\theta)} \quad \textit{non-decreasing in } \theta, \quad \forall \sigma \in \Sigma.$$

Proposition 4. *Suppose that Assumption 5 holds and that there exists $\theta^* \in (\underline{\theta}, \overline{\theta}]$ such* *that*

$$1+\frac{g_\theta(\underline{\sigma}|\theta^*)}{g(\underline{\sigma}|\theta^*)}\frac{F(\theta^*)}{f(\theta^*)}=0. \tag{5.12}$$

Then, there exists a threshold value $\sigma^{sb}(\theta)\in(\underline{\sigma},\overline{\sigma}]$[14] *defined as*

$$1+\frac{g_\theta(\sigma^{sb}(\theta)|\theta)}{g(\sigma^{sb}(\theta)|\theta)}\frac{F(\theta)}{f(\theta)}=0,\quad \forall\theta\in\left[\theta^*,\overline{\theta}\right]. \tag{5.13}$$

such that solution to the relaxed problem $(\mathcal{P}^r)$ *is characterized as follows.*

1. *The optimal output entails:*

$$\begin{cases} q^{sb}(\theta,\sigma)=0 & \text{if } \theta\in\left[\theta^*,\overline{\theta}\right] \text{ and } \sigma\in\left[\underline{\sigma},\sigma^{sb}(\theta)\right],\\ S'(q^{sb}(\theta,\sigma))=\theta+\frac{\frac{F(\theta)}{f(\theta)}}{1+\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\frac{F(\theta)}{f(\theta)}}, & \text{otherwise.}\end{cases} \tag{5.14}$$

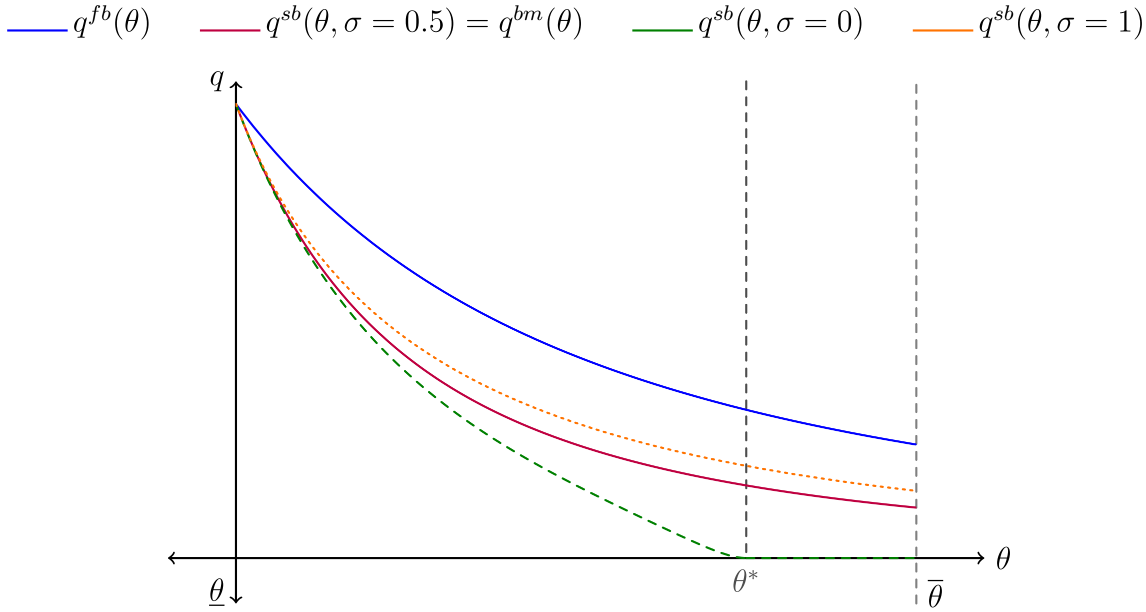
2. *Money is burned following very bad news only:*

$$b^{sb}(\theta,\sigma)=\begin{cases} B & \text{if } \theta\in\left[\theta^*,\overline{\theta}\right] \text{ and } \sigma\in\left[\underline{\sigma},\sigma^{sb}(\theta)\right],\\ 0 & \text{otherwise.}\end{cases} \tag{5.15}$$

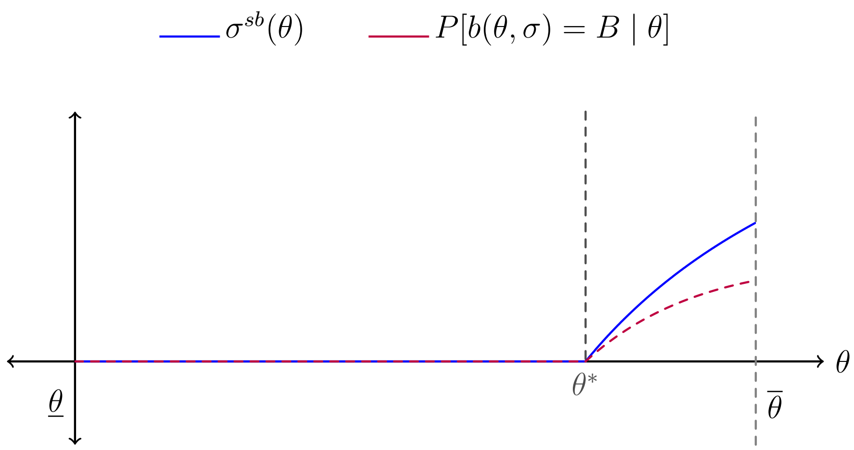
Proposition 4 demonstrates that burning money is now useful in the case of a strong conflict between the principal's assessment and the agent's report; the agent claiming a cost parameter which is high enough while the principal's evaluation is much more favorable. We already know from Proposition 3 that, following *bad news*, the principal wants to reduce the agent's output. The novelty is that, when Assumption 5 and Condition (5.12) hold altogether, the principal prefers not to ask for any production so as there are no gains from trade. At the same time, even more surplus can be destroyed when money is also burned. Money burning acts as a credible threat to extract more of the agent's surplus when reports are strongly conflicting. Instead, when the principal's signal does not conflict much with the agent's report, production remains attractive. The principal finds profitable to reduce output but retains from burning money.

Nonregularity. The availability of the signal σ may cause the contracting problem to be non-regular. Indeed Strausz (2006) shows in his binary model without money burning that the participation constraint may be binding for both the inefficient and the efficient type. In our continuous type space scenario, dealing with this problem requires the use of techniques borrowed from Jullien (2000) and Martimort and Stole (2022). In Appendix C we extend Propositions 3 and 4 to consider such a possibility.

[14]With the convention that $\sigma^{sb}(\theta)=\underline{\sigma}$ for $\theta\in[\underline{\theta},\theta^*)$.



(a) Output schemes. In blue, the first-best level of output. For $\sigma = 0.5$ (purple), the score $\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}$ of the signal is zero, which corresponds to a case with neither good nor bad news: output hence equals the Baron-Myerson amount. For $\sigma = 0$ (green, dashed), the principal exacerbates the downwards distortions, in order to disincentivize over-reporting. Distortions can be large enough to bring output to zero for the upper tail of types $\theta > \theta^*$. For $\sigma = 1$ (orange, dotted), the principal softens the distortions in output.



(b) In blue, the threshold for the signal σ conditional on θ that determines money burning. In purple, dashed, the probability of money burning conditional on the reported type θ.

Figure 4: Signal-dependent distortions and money burning thresholds.

6. When the Signal Is Learned after Production

In order to further illustrate the use of money burning, we consider an alternative scenario where the signal σ is realised after output q has already been chosen by the agent. Under those circumstances, production can no longer be made contingent on the principal's report about the signal. Yet, payments can be delayed to account for this possibility. We will thus simplify notations accordingly by suppressing the dependence of output on signal and we shall write $q(\hat{\theta})$. The principal's incentive compatibility constraints (4.4) show that there must exist a function H' such that the principal overall payment, including money burned, must be kept constant across all realizations of the signal as:

$$t(\theta,\sigma)+b(\theta,\sigma)=H'(\theta)\quad \forall\theta\in\Theta,\forall\sigma\in\Sigma. \tag{6.1}$$

If the principal were to use different payments for the agent's services as the signal he receives varies, those payments should be compensated by equal variations in the amount of money burned. To illustrate, suppose that the principal wants to punish the agent in case she gets a favorable signal by lowering the payment $t(\theta,\sigma_1)$ when σ_1 is small enough while, at the same time, she would like to reward the agent when the signal is less favorable, i.e., setting $t(\theta,\sigma_2)>t(\theta,\sigma_1)$ when $\sigma_2>\sigma_1$. Since output can no longer vary with the signal, the only way of making such signal-dependent rewards incentive compatible for the principal is to also increase money burning following σ_1 precisely by an amount $t(\theta,\sigma_2)-t(\theta,\sigma_1)$. In other words, money burning is given its best chance when the principal learns signal after production because it is only with money burning that the principal can credibly commit to incorporate her subjective evaluation in the agent's compensation.

Adapting our findings in Lemma 1, the agent's rent profile must now satisfy a simpler Envelope Condition

$$\dot{U}(\theta)=-q(\theta)-\mathbb{E}_\sigma\left(b(\theta,\sigma)\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}|\theta\right). \tag{6.2}$$

This condition is already illuminative of the key forces at play. By eventually choosing to burn money for all signal realizations σ such that $g_\theta(\sigma|\theta)<0$, the principal creates countervailing incentives for his agent. On the one hand, the latter wants again to overstate his cost parameter θ and claim he has a cost $\theta+d\theta$ to save $q(\theta+d\theta)d\theta\approx q(\theta)d\theta$ in doing so. On the other hand, the likelihood of a favorable signal (i.e., σ small enough) is greater at type θ than at type $\theta+d\theta$. Burning money for sufficiently favorable signals thus induces a significant cost on the exaggerating agent which limits those incentives to exaggerate costs.

We are now equipped for a characterization of the optimal contract. We shall distinguish two scenarios depending on how strong those countervailing incentives are.

6.1. Wage Compression

In the first scenario, the maximal amount of money that can be burned is not too large. The countervailing effect discussed above is of limited magnitude even if the principal uses money burning with some probability, doing so does not perturb too much the standard lessons of the Baron-Myerson scenario. The incentive problem $(\mathcal{P})$ remains regular, with the agent's participation constraint being again reduced to (5.3) and the Envelope Condition (6.2) being sufficient for (4.5).

PROPOSITION 5. *Suppose that Assumption 5 holds and again define θ^* as in (5.12). Then, the solution to the relaxed problem $(\mathcal{P}^r)$ entails the following properties.*

1. *The optimal output is always the Baron-Myerson outcome:*

$$q^{sb}(\theta) = q^{bm}(\theta), \quad \forall \theta \in \Theta. \tag{6.3}$$

2. *Money is burned following very bad news only:*

$$b^{sb}(\theta, \sigma) = \begin{cases} B & \text{if } \theta \in [\theta^*, \overline{\theta}] \text{ and } \sigma \in [\underline{\sigma}, \sigma^{sb}(\theta)], \\ 0 & \text{otherwise.} \end{cases} \tag{6.4}$$

$(\mathcal{P})$ is regular provided that

$$B \in \left[0, \min_{\theta \in [\theta^*, \overline{\theta}]} \left\{ \frac{-q^{bm}(\theta)}{G_\theta(\sigma^{sb}(\theta)|\theta)} \right\} \right]. \tag{6.5}$$

Proposition 5 delineates circumstances where the sole role of money burning is to reduce payments; a form of wage compression. It has no consequences whatsoever for the volume of trade and output remains at its Baron-Myerson level.

This result is reminiscent of MacLeod (2003). This author demonstrates that, under pure moral hazard, the optimal wage exhibits some form of compression. The principal commits to a fixed payment, independent of her own subjective evaluation of the agent, for all outcomes except the worst one and burn some money in that state. The incentive power of such a scheme is reduced in comparison with the ideal scenario where an objective performance evaluation would be available. As a result, the agent's effort and its liability rent are both reduced. Proposition 5 shows also that the agent's information rent is somewhat reduced in comparison in our pure screening scenario. Indeed, using the fact that (5.3) is binding at the optimum for a regular problem and the Envelope Condition (6.2), a simple integration gives us the expression of the agent's information rent as

$$U^{sb}(\theta) = \int_{\theta}^{\bar{\theta}} \left(q^{bm}(\tilde{\theta}) + BG_{\theta}(\sigma^{sb}(\tilde{\theta})|\tilde{\theta}) \right) d\tilde{\theta} \leq U^{bm}(\theta), \quad \forall \theta \in \Theta; \tag{6.6}$$

where the inequality is strict for all $\theta < \bar{\theta}$ since money burning occurs with positive probability for θ close enough of $\bar{\theta}$. From there, we may rewrite the agent's overall payment as

$$S(q^{bm}(\theta)) - BG(\sigma^{sb}(\theta)|\theta) - H^{sb}(\theta) = U^{sb}(\theta) + \theta q^{bm}(\theta) \tag{6.7}$$

$$\leq U^{bm}(\theta) + \theta q^{bm}(\theta) = S(q^{bm}(\theta)) - H^{bm}(\theta), \quad \forall \theta \in \Theta$$

where the right-hand side stands for the payment received in the absence of any signal and has been rewritten to stress, that even in that scenario, one possible implementation of the Baron-Myerson outcome is to let the agent enjoy all gross surplus of the transaction $S(q^{bm}(\theta))$ against a fee $H^{bm}(\theta)$. Condition (6.7) then shows that the agent is asked to disgorge more of the overall surplus of the transaction when a signal is available.

A remarkable finding of Proposition 5 is that the volume of trade remains unchanged, in contrast with MacLeod (2003)'s findings under pure moral hazard. The reason is simple. In our hidden information environment, the principal can use both quantity and payments for screening purposes. When production is chosen before any signal can be learned, money burning can help to save on payments but, as long as countervailing incentives are not too strong, the reduction of payments available when the signal can be used does not interact with output requirements. Instead, in a pure moral hazard environment, payments have always a direct impact on effort.

6.2. Output Expansion

Consider now a scenario where countervailing incentives are more pronounced. That is, condition (6.5) fails to hold: $B > \min_{\theta \in [\theta^*, \bar{\theta}]} \left\{ \frac{-q^{bm}(\theta)}{G_{\theta}(\sigma^{sb}(\theta)|\theta)} \right\}$. We consider a partially relaxed problem $(\mathcal{P}^{r\prime})$ where the participation constraint writes as (4.1), a pure state constraint, while incentive compatibility constraints (4.5) are still reduced to the Envelope Condition (6.2). The solution to this problem is now characterized.

PROPOSITION 6. *Suppose that Assumption 5 holds and again define θ^* as in (5.12). Then, the solution to the relaxed problem $(\mathcal{P}^{r\prime})$ entails the following properties.*

1. *There exists a probability measure μ whose support is $\Omega = \{\theta : U^{r\prime}(\theta) = 0\}$ such*

that

$$S'(q^{r\prime}(\theta)) = \theta + \frac{F(\theta) - M(\theta)}{f(\theta)}, \quad \forall \theta \in \Theta. \tag{6.8}$$

where $M(\theta) = \int_{[\underline{\theta},\theta)} \mu(d\tilde{\theta})$. *In particular, we have*

$$q^{r\prime}(\theta) \geq q^{bm}(\theta), \quad \forall \theta \in \Theta. \tag{6.9}$$

2. *Money is burned following very bad news only*

$$b^{r\prime}(\theta, \sigma) = \begin{cases} B & \text{if } \theta \in \left[\theta^*, \overline{\theta}\right] \text{ and } \sigma \in [\underline{\sigma}, \sigma^{r\prime}(\theta)], \\ 0 & \text{otherwise.} \end{cases} \tag{6.10}$$

where $\sigma^{r\prime}(\theta)$ *is now defined as*

$$1 + \frac{g_\theta(\sigma^{r\prime}(\theta)|\theta)}{g(\sigma^{r\prime}(\theta)|\theta)} \frac{F(\theta) - M(\theta)}{f(\theta)} = 0, \quad \forall \theta \in \left[\theta^*, \overline{\theta}\right]. \tag{6.11}$$

In particular, we have

$$\sigma^{r\prime}(\theta) \leq \sigma^{sb}(\theta) \text{ if } q^{r\prime}(\theta) \leq q^{fb}(\theta) \quad \forall \theta \in \Theta. \tag{6.12}$$

Moreover, it must be that

$$\min\{\Omega\} > \theta^* \tag{6.13}$$

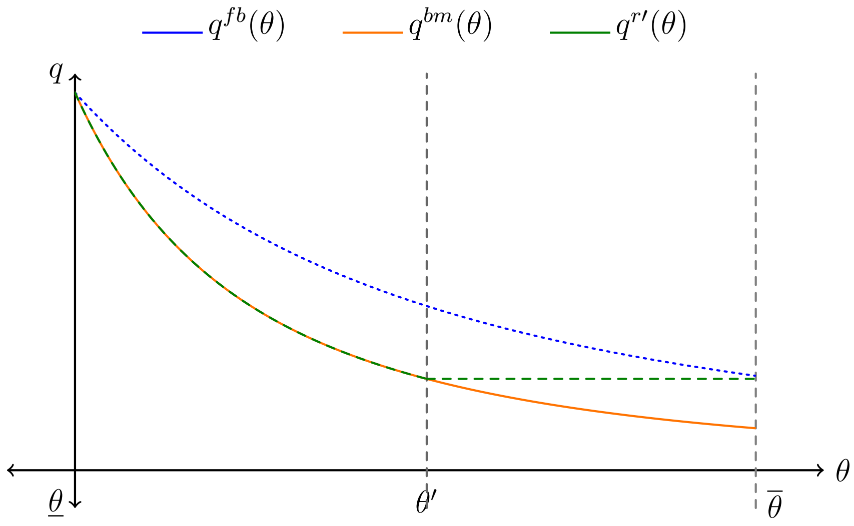
and

$$q^{r\prime}(\theta) = -BG_\theta(\sigma^{r\prime}(\theta)|\theta), \quad \forall \theta \in \mathring{\Omega}. \tag{6.14}$$

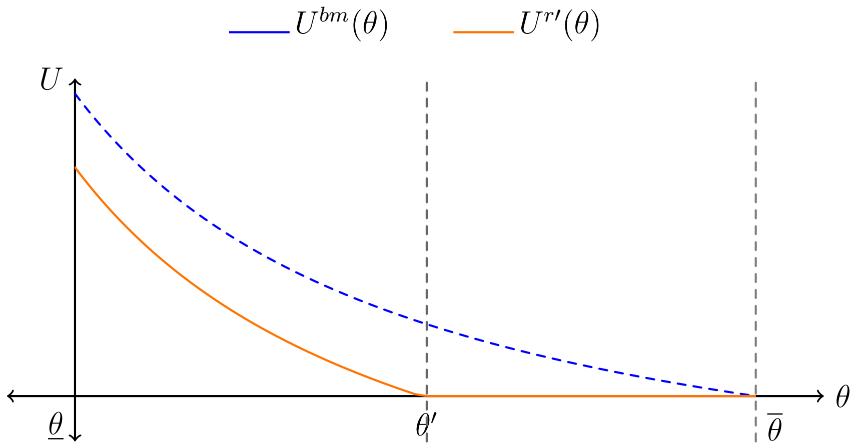
The incentive constraints (4.5) hold provided that $q^{r\prime}$ *is non-increasing and* $\sigma^{r\prime}$ *non-decreasing.*

Proposition 6 accounts for the possibility that countervailing incentives are so strong that the agent's participation constraint binds on an interval with non-empty interior.[15] The solution is then characterized by a probability measure over this set. On this set, money burning is optimally used to extract all the agent's information rent. Some distortions in output are still present, but these are lower compared to the Baron-Myerson solution. At the same time, money burning is less useful and limited in comparison with the scenario described in Proposition 5, since the rents of some mass of agent's types are already zero.

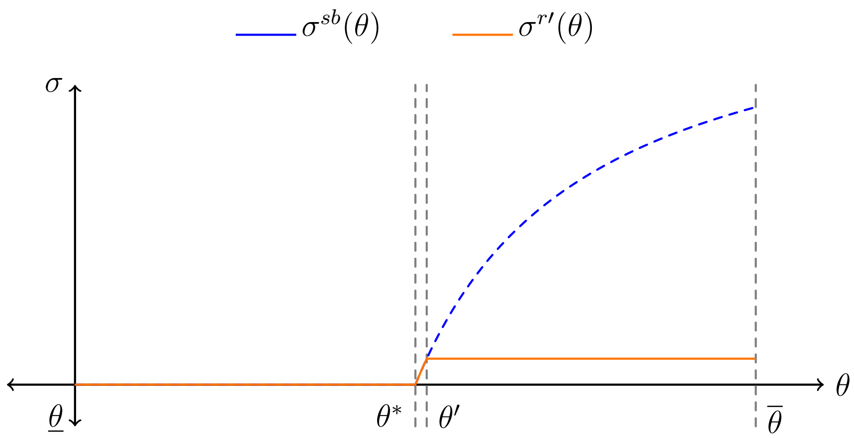
[15] If condition (6.5) holds, then the measure is such that $M(\theta) = 0$ for all $\theta \in \Theta$, meaning that the measure puts all the mass on the atom $\overline{\theta}$, which is the regular case.



(a) Output schemes. In blue (dotted), first-best output. In orange, Baron-Myerson scheme. In green (dashed), scheme with money burning.



(b) Rents. In blue (dashed), rents obtained by the agent in Baron-Myerson. In orange, rents with signals and money burning as defined in Proposition 6. There is a positive measure of types for which $U(\theta) = 0$.



(c) Money burning thresholds. In blue (dashed), threshold $\sigma^{sb}(\theta)$ as defined in 5.13. In orange, threshold $\sigma^{r\prime}(\theta)$ as defined in Proposition 6.

Figure 5: Optimal contract of Proposition 6.

Example 1. (continued) *When condition (6.5) fails to hold, the type space* $[\underline{\theta}, \overline{\theta}]$ *is partitioned into three subsets under the optimal contract. On* $[\underline{\theta}, \theta^*]$*, output equals the Baron-Myerson level, no money is burned, and rents are positive. On* (θ^*, θ')*, output equals the Baron-Myerson level, some money is burned with* $\sigma^{r\prime}(\theta) = \sigma^{sb}(\theta)$ *but rents are still strictly positive. Finally, on* $\Omega = [\theta', \overline{\theta}]$*, output is constant at* $q^{bm}(\theta')$*, money is burnt with* $\sigma^{r\prime}(\theta) \leq \sigma^{sb}(\theta)$*, and rents are zero. The type* $\theta' \in (\theta^*, \overline{\theta})$ *is the smallest* θ *such that*

$$-q^{bm}(\theta) - BG_\theta(\sigma^{sb}(\theta)|\theta) \geq 0,$$

which represents the value at which the rent would hit zero with the allocation described in Proposition 5 when (6.5) fails. As a consequence, there is output expansion on $\mathring{\Omega}$*.*

By combining (6.8) (6.11) and (6.14), we can find the cutoff $\sigma^{r\prime}(\theta)$ *on* $\mathring{\Omega}$*:*

$$S'(-BG_\theta(\sigma^{r\prime}(\theta)|\theta)) = \theta - \frac{g(\sigma^{r\prime}(\theta)|\theta)}{g_\theta(\sigma^{r\prime}(\theta)|\theta)}, \quad \forall \theta \in \mathring{\Omega}.$$

Since $g(\sigma|\theta)$ *is affine in* θ *in this example,* g_θ *and* G_θ *are constant in* θ*. It follows that the cutoff* $\sigma^{r\prime}(\theta)$ *is constant in* θ *on* $\mathring{\Omega}$*. From (6.14), it follows that* $q^{r\prime}(\theta)$ *is also constant on* $\mathring{\Omega}$*.*

The measure is such that:

$$M(\theta) = \begin{cases} 0 & \text{if } \theta \leq \theta', \\ F(\theta) + f(\theta)\frac{g(\sigma^{r\prime}(\theta)|\theta)}{g_\theta(\sigma^{r\prime}(\theta)|\theta)} \in (0,1) & \text{if } \theta' \leq \theta \leq \overline{\theta}, \end{cases}$$

In particular, the measure puts positive mass on the atom $\overline{\theta}$*. A representation of the optimal contract can be found in Figure 5.*

References

Bagnoli, M. and Bergstrom, T. (2005). Log-Concave Probability and Its Applications. *Economic Theory*, 26, 445–469.

Baker, G., Gibbons, R., and Murphy, K. J. (1994). Subjective Performance Measures in Optimal Incentive Contracts. *The Quarterly Journal of Economics*, 109(4), 1125–1156.

Baron, D. P. and Myerson, R. B. (1982). Regulating a Monopoly with Unknown Costs. *Econometrica*, 50(4), 911–930.

Celik, G., Shin, D., and Strausz, R. (2021). Public Good Overprovision by a Manipulative Provider. *The RAND Journal of Economics*, 52(2), 314–333.

Cella, M. (2008). Informed Principal with Correlation. *Games and Economic Behavior*, 64(2), 433–456.

Cheng, C. (2021). Moral Hazard in Teams with Subjective Evaluations. *The RAND Journal of Economics*, 52(1), 22–48.

Crémer, J. and McLean, R. P. (1985). Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent. *Econometrica*, 53(2), 345–361.

Crémer, R. and McLean, R. P. (1988). Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions. *Econometrica*, 56(6), 1247–1257.

Demougin, D. M. and Garvie, D. A. (1991). Contractual Design with Correlated Information under Limited Liability. *The RAND Journal of Economics*, 22(3), 477–489.

Dequiedt, V. and Martimort, D. (2015). Vertical Contracting with Informational Opportunism. *American Economic Review*, 105, 2141–2182.

Fuchs, W. (2007). Contracted with Repeated Moral Hazard and Private Evaluations. *American Economic Review*, 97, 1432–1448.

Gary-Bobo, R. and Spiegel, Y. (2006). Optimal State-Contingent Regulation under Limited Liability. *The RAND Journal of Economics*, 37(2), 431–448.

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

Holmström, B. (1982). Moral Hazard in Teams. *The Bell Journal of Economics*, 13(2), 324–340.

Jullien, B. (2000). Participation Constraints in Adverse Selection Models. *Journal of Economic Theory*, 93, 1–47.

Kattwinkel, D. and Knoepfle, J. (2023). Costless Information and Costly Verification: A Case for Transparency. *Journal of Political Economy*, 131(2), 504–548.

Kessler, A. S., Lülfesmann, C., and Schmitz, P. W. (2005). Endogenous Punishments in Agency with Verifiable Ex Post Information. *International Economic Review*, 46(4), 1207–1231.

Khalil, F., Lawarrée, J., and Scott, T. J. (2015). Private Monitoring, Collusion and the Timing of Information. *The RAND Journal of Economics*, 46(4), 872–890.

Krähmer, D. (2020). Information Disclosure and Full Surplus Extraction in Mechanism Design. *Journal of Economic Theory*, 187, 105020.

Laffont, J. J. and Maskin, E. (1980). A Differential Approach to Dominant Strategy Mechanisms. *Econometrica*, 48(6), 1507–1520.

Lang, M. (2023). Stochastic Contracts and Subjective Evaluations. *The RAND Journal of Economics*, 54(1), 104–134.

Levin, J. (2003). Relational Incentive Contracts. *American Economic Review*, 93(3), 835–857.

Lewis, T. and Sappington, D. (1989a). Countervailing Incentives in Agency Problems. *Journal of Economic Theory*, 49, 294–313.

Lewis, T. and Sappington, D. (1989b). Inflexible Rules in Incentive Problems. *The American Economic Review*, 79, 69–84.

Lopomo, G., Rigotti, L., and Shannon, C. (2022). Detectability, Duality, and Surplus Extraction. *Journal of Economic Theory*, 204, 105465.

MacLeod, W. B. (2003). Optimal Contracting with Subjective Evaluation. *American Economic Review*, 93, 216–240.

Martimort, D. and Stole, L. (2022). Participation Constraints in Discontinuous Adverse Selection Models. *Theoretical Economics*, 17, 1145–1181.

McAfee, R. P. and Reny, P. (1992). Correlated Information and Mechanism Design. *Econometrica*, 60(2), 395–421.

Milgrom, P. and Segal, I. (2002). Envelope Theorems for Arbitrary Choice Sets. *Econometrica*, 70, 583–601.

Rahman, D. (2012a). Surplus Extraction on Arbitrary Type Spaces. Unpublished manuscript.

Rahman, D. (2012b). But Who Will Monitor the Monitor? *American Economic Review*, 102(6), 2767–2797.

Rajan, M. V. and Reichelstein, S. (2006). Subjective Performance Indicators and Discretionary Bonus Pools. *Journal of Accounting Research*, 44(3), 585–618.

Riordan, M. and Sappington, D. (1988). Optimal Contracts with Public Ex Post Information. *Journal of Economic Theory*, 45, 189–199.

Strausz, R. (2006). Interim Information in Long-Term Contracts. *Journal of Economics and Management Science*, 15, 1041–1067.

Zábojník, J. (2014). Subjective Evaluations with Performance Feedback. *The RAND Journal of Economics*, 45(2), 341–369.

Appendix A. Main Proofs

Proof of Proposition 1. Consider the payments as in (3.3) with, in addition,

$$z(\hat{\theta},\sigma) = k(\hat{\theta})\frac{g_\theta(\sigma|\hat{\theta})}{g(\sigma|\hat{\theta})} \quad \forall\hat{\theta}\in\Theta, \forall\sigma\in\Sigma \tag{A.1}$$

with a convenient function k to be found below.

Incentive compatibility constraints (4.2) are still valid when σ is verifiable. Those constraints thus rewrite as

$$U(\theta) = \max_{\hat{\theta}\in\Theta} S(q^{fb}(\hat{\theta})) - \theta q^{fb}(\hat{\theta}) - W^{fb}(\hat{\theta}) + k(\hat{\theta})\mathbb{E}_\sigma\left(\frac{g_\theta(\sigma|\hat{\theta})}{g(\sigma|\hat{\theta})}|\theta\right), \quad \forall\theta\in\Theta. \tag{A.2}$$

By incentive compatibility plus the fact that $\mathbb{E}_\sigma\left(\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}|\theta\right) = 0$, we check that there is full extraction of the agent's surplus when the agent tells the truth on his type, namely

$$U(\theta) = S(q^{fb}(\theta)) - \theta q^{fb}(\theta) - W^{fb}(\theta) = 0, \quad \forall\theta\in\Theta. \tag{A.3}$$

Observe that the maximand in (A.2) is absolutely continuous in θ when $\frac{g_\theta(\sigma|\hat{\theta})}{g(\sigma|\hat{\theta})}$ is so as we assumed. We can thus apply the Envelope Theorem in Milgrom and Segal (2002, Theorem 2), to show that U is absolutely continuous with a derivative (which is actually defined everywhere in the case under scrutiny) which is now worth

$$\dot{U}(\theta) = -q^{fb}(\theta) + k(\theta)\mathbb{E}_\sigma\left(\left(\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\right)^2|\theta\right), \quad \forall\theta\in\Theta. \tag{A.4}$$

Identifying this envelope condition with the derivative of (A.3), we must thus have

$$k(\theta) = \frac{q^{fb}(\theta)}{\mathbb{E}_\sigma\left(\left(\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\right)^2|\theta\right)}, \quad \forall\theta\in\Theta. \tag{A.5}$$

Taking together (A.2) and (A.3), global incentive compatibility holds when

$$(\theta-\hat{\theta})q^{fb}(\hat{\theta}) \geq k(\hat{\theta})\mathbb{E}_\sigma\left(\frac{g_\theta(\sigma|\hat{\theta})}{g(\sigma|\hat{\theta})}|\theta\right), \quad \forall(\theta,\hat{\theta})\in\Theta^2$$

or, using (A.4),

$$(\theta-\hat{\theta})\mathbb{E}_\sigma\left(\left(\frac{g_\theta(\sigma|\hat{\theta})}{g(\sigma|\hat{\theta})}\right)^2|\hat{\theta}\right) \geq \mathbb{E}_\sigma\left(\frac{g_\theta(\sigma|\hat{\theta})}{g(\sigma|\hat{\theta})}|\theta\right), \quad \forall(\theta,\hat{\theta})\in\Theta^2. \tag{A.6}$$

Let us now define

$$\psi(\theta,\hat{\theta}) = \mathbb{E}_\sigma\left(\frac{g_\theta(\sigma|\hat{\theta})}{g(\sigma|\hat{\theta})}|\theta\right).$$

We may compute

$$\psi_\theta(\hat{\theta},\hat{\theta}) = \mathbb{E}_\sigma\left(\left(\frac{g_\theta(\sigma|\hat{\theta})}{g(\sigma|\hat{\theta})}\right)^2|\hat{\theta}\right), \psi_{\theta\theta}(\theta,\hat{\theta}) = \mathbb{E}_\sigma\left(\frac{g_\theta(\sigma|\hat{\theta})g_{\theta\theta}(\sigma|\theta)}{g(\sigma|\hat{\theta})g(\sigma|\theta)}|\theta\right).$$

Integrating by parts yields

$$\psi_{\theta\theta}(\theta,\hat{\theta}) = -\int_{\underline{\sigma}}^{\overline{\sigma}}\frac{d}{d\sigma}\left(\frac{g_\theta(\sigma|\hat{\theta})}{g(\sigma|\hat{\theta})}\right)G_{\theta\theta}(\sigma|\theta)d\sigma.$$

When Assumptions 2 and 3 both hold, $\psi(\theta,\hat{\theta})$ is concave with respect to θ and thus

$$\psi(\theta,\hat{\theta}) \leq \psi(\hat{\theta},\hat{\theta}) + \psi_\theta(\hat{\theta},\hat{\theta})(\theta-\hat{\theta})$$

which, taking into account $\psi(\hat{\theta},\hat{\theta}) = 0$, amounts to (A.6). □

Proof of Lemma 1. Observe that the maximand in (A.2) is absolutely continuous in θ for any $\hat{\theta}$ when g is twice continuously differentiable in θ as assumed. We can thus apply the Envelope Theorem in Milgrom and Segal (2002, Theorem 2), to show that U is absolutely continuous with a derivative, when it exists, which is defined as in (4.7). □

Proof of Proposition 2. We prove each item in turn.

1. Suppose that the first-best output profile $q^{fb}(\theta)$ is implementable without money burning. Then, using (4.5), there would exist a rent profile U and a function $H(\theta)$ such that

   (A.7) $$U(\theta) = \max_{\hat{\theta}\in\Theta} S(q^{fb}(\hat{\theta})) - \theta q^{fb}(\hat{\theta}) - H(\hat{\theta}), \quad \forall\theta\in\Theta.$$

   We can again apply Lemma 1 to show that U is absolutely continuous with a derivative (almost everywhere) worth

   $$\dot{U}(\theta) = -q^{fb}(\theta) \quad \forall\theta\in\Theta.$$

   With full surplus extraction, we should have $U(\theta) = U^{fb}(\theta) = 0$ and thus $q^{fb}(\theta) = 0$ a.e.; a contradiction with the fact that output remains positive in (3.1).

2. Suppose that the first-best output allocation $(q^{fb}(\theta), U^{fb}(\theta))$ is implementable with money burning. Then, using (4.5), there would exist a function $H(\theta)$ such that

   (A.8) $$U^{fb}(\theta) = 0 = \max_{\hat{\theta}\in\Theta} S(q^{fb}(\hat{\theta})) - \theta q^{fb}(\hat{\theta}) - H(\hat{\theta}) - \mathbb{E}_\sigma\left(b(\hat{\theta},\sigma)|\theta\right), \quad \forall\theta\in\Theta.$$

We can again apply the Envelope Theorem in Milgrom and Segal (2002, Theorem 2) to get

$$\dot{U}^{fb}(\theta) = 0 = -q^{fb}(\theta) - \mathbb{E}_\sigma\left(b(\theta,\sigma)\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}|\theta\right), \quad \forall\theta \in \Theta.$$

With full surplus extraction, we should thus have

$$-q^{fb}(\theta) - \mathbb{E}_\sigma\left(b(\theta,\sigma)\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}|\theta\right) = 0, \quad \forall\theta \in \Theta. \tag{A.9}$$

From (A.8), global incentive compatibility holds when

$$0 \geq S(q^{fb}(\hat{\theta})) - \theta q^{fb}(\hat{\theta}) - H(\hat{\theta}) - \mathbb{E}_\sigma\left(b(\hat{\theta},\sigma)|\theta\right), \quad \forall(\theta,\hat{\theta}) \in \Theta^2$$

or,

$$(\theta - \hat{\theta})q^{fb}(\hat{\theta}) \geq \mathbb{E}_\sigma\left(b(\hat{\theta},\sigma)|\hat{\theta}\right) - \mathbb{E}_\sigma\left(b(\hat{\theta},\sigma)|\theta\right), \quad \forall(\theta,\hat{\theta}) \in \Theta^2. \tag{A.10}$$

Consider the principal's optimization problem when willing to implement the first-best allocation $(q^{fb}(\theta), U^{fb}(\theta) = 0)$. It consists in pointwise maximizing with respect to $b(\theta,\sigma) \in [0, B]$ the following objective

$$\mathbb{E}_\theta\left(S(q^{fb}(\theta)) - \theta q^{fb}(\hat{\theta}) - \mathbb{E}_\sigma\left(b(\hat{\theta},\sigma)|\theta\right)\right)$$

subject to (A.9) only. Provided that (A.10) holds for the solution to this relaxed problem, we shall have find the optimal mechanism that extracts all surplus from the agent whatever his type and implements the first-best output.

Let denote by $\lambda(\theta)$ the Lagrange multiplier for (A.9). Because the corresponding Lagrangean is linear in $b(\theta,\sigma)$, the solution is bang-bang with

$$b(\theta,\sigma)\begin{cases} = B & \text{if } g(\sigma|\theta)f(\theta) + \lambda(\theta)g_\theta(\sigma|\theta) < 0, \\ \in [0,B] & \text{if } g(\sigma|\theta)f(\theta) + \lambda(\theta)g_\theta(\sigma|\theta) = 0, \\ = 0 & \text{if } g(\sigma|\theta)f(\theta) + \lambda(\theta)g_\theta(\sigma|\theta) > 0. \end{cases} \tag{A.11}$$

First, notice that $\lambda(\theta) = 0$ cannot be; because, otherwise, $b(\theta,\sigma) = 0$ for all $\sigma \in \Sigma$ and (A.9) cannot hold. Second, suppose now that $\lambda(\theta) < 0$. Then, Assumption 2 would imply that $b(\theta,\sigma) = B$ for $\sigma \in \left[\sigma^{fb}(\theta), \overline{\sigma}\right]$ where $\sigma^{fb}(\theta)$ is defined as

$$\frac{g_\theta(\sigma^{fb}(\theta)|\theta)}{g(\sigma^{fb}(\theta)|\theta)} = -\frac{f(\theta)}{\lambda(\theta)}. \tag{A.12}$$

Moreover, $g_\theta(\sigma|\theta) > 0$ for all $\sigma \in (\sigma^{fb}(\theta), \overline{\sigma}]$. But then, the left-hand side of (A.9) is necessarily negative; a contradiction. Hence, $\lambda(\theta) > 0$. Then, Assumption 2 implies that $b(\theta,\sigma) = B$ for $\sigma \in \left[\underline{\sigma}, \sigma^{fb}(\theta)\right]$. Now, observe that we have $g_\theta(\sigma|\theta) < 0$ for all $\sigma \in \left[\underline{\sigma}, \sigma^{fb}(\theta)\right)$. Finally, $\sigma^{fb}(\theta)$ is defined as (5.2). In fact, Assumption 2 implies that $\int_{\underline{\sigma}}^{\sigma} g_\theta(\tilde{\sigma}|\theta)d\tilde{\sigma} = G_\theta(\sigma|\theta)$ is quasi-convex in σ and thus minimum at $\sigma_0(\theta)$ such that $g_\theta(\sigma_0(\theta)|\theta) = 0$. Condition (5.2) has thus a solution $\sigma^{fb}(\theta) \in [\underline{\sigma}, \sigma_0(\theta)]$ when B is large enough, namely

$$B \geq \max_{\theta\in\Theta} \left\{ \frac{-q^{fb}(\theta)}{G_\theta(\sigma_0(\theta)|\theta)} \right\}.$$

Let us check that (A.10) holds for the solution so obtained. First, notice that H is defined as

$$\text{(A.13)} \qquad H(\theta) = S(q^{fb}(\theta)) - \theta q^{fb}(\theta) - BG_\theta(\sigma^{fb}(\theta)|\theta), \quad \forall\theta \in \Theta.$$

Second, global incentive compatibility amounts to checking that

$$\text{(A.14)} \qquad (\theta - \hat{\theta})q^{fb}(\hat{\theta}) \geq B \left(\int_{\underline{\sigma}}^{\sigma^{fb}(\hat{\theta})} (g(\sigma|\hat{\theta}) - g(\sigma|\theta))d\sigma \right), \quad \forall(\theta, \hat{\theta}) \in \Theta^2.$$

Using

$$-q^{fb}(\theta) = B \int_{\underline{\sigma}}^{\sigma^{fb}(\theta)} g_\theta(\sigma|\theta)d\sigma = BG_\theta(\sigma^{fb}(\theta)|\theta),$$

and inserting into (A.14) allows us to rewrite this condition as

$$\text{(A.15)} \qquad (\hat{\theta} - \theta)G_\theta(\sigma^{fb}(\hat{\theta})|\hat{\theta}) \geq G(\sigma^{fb}(\hat{\theta})|\hat{\theta}) - G(\sigma^{fb}(\hat{\theta})|\theta)), \quad \forall(\theta, \hat{\theta}) \in \Theta^2.$$

This condition holds under Assumption 3.

□

Proof of Proposition 3. When the incentive problem is regular, we can recover from the binding participation constraint for the worst type $\overline{\theta}$ (5.3) and the Envelope Condition (4.7), the expression of the rent profile $U(\theta)$ in terms of the output allocation as (5.8). From there, the fee $H(\theta)$ is obtained as

$$\text{(A.16)} \qquad U(\theta) = \mathbb{E}_\sigma \left(S(q(\theta, \sigma)) - \theta q(\theta, \sigma) - b(\theta, \sigma)|\theta \right) - H(\theta).$$

Inserting (5.8) into the maximand ($\mathcal{P}^r$) and integrating by parts yields that $q^{sb}(\theta, \sigma)$ and $b(\theta, \sigma)$ should now be pointwise-optimizers for

$$\text{(A.17)} \qquad \mathbb{E}_{\theta,\sigma} \left(\left(1 + \frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)} \frac{F(\theta)}{f(\theta)} \right) (S(q(\theta, \sigma)) - \theta q(\theta, \sigma) - b(\theta, \sigma)) - \frac{F(\theta)}{f(\theta)} q(\theta, \sigma) \right).$$

From there, the following consequences are immediately drawn.

1. Observe that Condition (5.5) implies that the coefficient of the surplus $S(q(\theta, \sigma)) - \theta q(\theta, \sigma))$ in this maximand remains positive. Hence, the optimization problem remains strictly concave. Pointwise optimization of this objective then yields the expression of

the second-best output in (5.6). Inada conditions on S, as assumed, imply that $q^{sb}(\theta,\sigma)$ remains positive.

2. Because Condition (5.5) holds and the objective is linear in $b(\theta,\sigma)$, (5.7) immediately follows.

MONOTONICITY. Condition (5.4) implies that $q^{sb}(\theta,\sigma)$ is non-increasing in θ. That $q^{sb}(\theta,\sigma)$ is non-decreasing in σ follows from Assumption 2.

REGULARITY. Regularity follows from checking that the necessary condition (4.7) suffices for (4.5) while the participation constraint (4.1) follows from (5.3). We check each item in turn.

1. We rewrite the incentive compatibility conditions (4.5) using (5.7) as

$$U^{sb}(\theta) \geq \mathbb{E}_\sigma\left(S(q^{sb}(\hat{\theta},\sigma)) - \theta q^{sb}(\hat{\theta},\sigma)|\theta\right) - H^{sb}(\hat{\theta})$$

or

$$U^{sb}(\theta)-U^{sb}(\hat{\theta}) \geq \mathbb{E}_\sigma\left(S(q^{sb}(\hat{\theta},\sigma)) - \theta q^{sb}(\hat{\theta},\sigma)|\theta\right)-\mathbb{E}_\sigma\left(S(q^{sb}(\hat{\theta},\sigma)) - \hat{\theta} q^{sb}(\hat{\theta},\sigma)|\hat{\theta}\right), \forall(\theta,\hat{\theta}) \in \Theta^2.$$

Using the necessary condition (4.7), we rewrite the left-hand side above and get

$$\text{(A.18)} \qquad \int_{\hat{\theta}}^{\theta} \mathbb{E}_\sigma\left(-q^{sb}(\tilde{\theta},\sigma) + \left(S(q^{sb}(\tilde{\theta},\sigma)) - \tilde{\theta} q^{sb}(\tilde{\theta},\sigma)\right)\frac{g_\theta(\sigma|\tilde{\theta})}{g(\sigma|\tilde{\theta})}|\tilde{\theta}\right)d\tilde{\theta}$$

$$\geq \mathbb{E}_\sigma\left(S(q^{sb}(\hat{\theta},\sigma)) - \theta q^{sb}(\hat{\theta},\sigma)|\theta\right) - \mathbb{E}_\sigma\left(\left(S(q^{sb}(\hat{\theta},\sigma)) - \hat{\theta} q^{sb}(\hat{\theta},\sigma)\right)|\hat{\theta}\right), \quad \forall(\theta,\hat{\theta}) \in \Theta^2.$$

Define now ψ as

$$\text{(A.19)} \qquad \psi(\theta,\hat{\theta}) = \mathbb{E}_\sigma\left(S(q^{sb}(\hat{\theta},\sigma)) - \theta q^{sb}(\hat{\theta},\sigma)|\theta\right).$$

We rewrite (A.18) as

$$\text{(A.20)} \qquad \int_{\hat{\theta}}^{\theta} \psi_\theta(\tilde{\theta},\tilde{\theta})d\tilde{\theta} \geq \psi(\hat{\theta},\hat{\theta}) - \psi(\theta,\hat{\theta}) \quad \forall(\theta,\hat{\theta}) \in \Theta^2.$$

Condition (A.20) amounts to

$$\int_{\hat{\theta}}^{\theta} [\psi_\theta(\tilde{\theta},\tilde{\theta}) - \psi_\theta(\tilde{\theta},\hat{\theta})]d\tilde{\theta} \geq 0$$

$$\int_{\hat{\theta}}^{\theta}\int_{\hat{\theta}}^{\tilde{\theta}} \psi_{\theta\hat{\theta}}(\tilde{\theta},x)dxd\tilde{\theta} \geq 0$$

It immediately follows that (A.18) is satisfied when

$$\text{(A.21)} \qquad \psi_{\theta\hat{\theta}}(\theta,\hat{\theta}) \geq 0 \quad \forall(\theta,\hat{\theta}) \in \Theta^2.$$

Using (A.19), we compute

$$\psi_{\theta\hat{\theta}}(\theta,\hat{\theta}) = \int_{\underline{\sigma}}^{\overline{\sigma}} q_\theta^{sb}(\hat{\theta},\sigma)\left(-1+(S'(q^{sb}(\hat{\theta},\sigma))-\theta)\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\right)g(\sigma|\theta)d\sigma. \tag{A.22}$$

Using (5.6), observe that

$$-1+(S'(q^{sb}(\theta,\sigma))-\theta)\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)} = -\frac{1}{1+\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\frac{F(\theta)}{f(\theta)}} < 0$$

Taken this with the monotonicity condition $q_\theta^{sb} < 0$, we get $\psi_{\theta\hat{\theta}}(\theta,\theta) > 0$ and so (A.21) holds when $\Delta = \overline{\theta} - \underline{\theta}$ is small enough.

2. We now demonstrate that $U^{sb}(\theta)$ is decreasing and thus the conjecture that (4.1) can be replaced by (5.3) is correct. To this end, we compute

$$\mathbb{E}_\sigma\left((S(q^{sb}(\theta,\sigma))-\theta q^{sb}(\theta,\sigma))\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}|\theta\right) = \int_{\underline{\sigma}}^{\overline{\sigma}}(S(q^{sb}(\theta,\sigma))-\theta q^{sb}(\theta,\sigma)g_\theta(\sigma|\theta)d\sigma$$

$$= \left[S(q^{sb}(\theta,\sigma))-\theta q^{sb}(\theta,\sigma)G_\theta(\sigma|\theta)\right]_{\underline{\sigma}}^{\overline{\sigma}} - \int_{\underline{\sigma}}^{\overline{\sigma}}(S'(q^{sb}(\theta,\sigma))-\theta)q_\sigma^{sb}(\theta,\sigma)G_\theta(\sigma|\theta)d\sigma.$$

Using $G_\theta(\underline{\sigma}|\theta) = G_\theta(\overline{\sigma}|\theta) = 0$ and (5.6), we rewrite the right-hand side as

$$-\int_{\underline{\sigma}}^{\overline{\sigma}} \frac{\frac{F(\theta)}{f(\theta)}}{1+\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\frac{F(\theta)}{f(\theta)}} q_\sigma^{sb}(\theta,\sigma)G_\theta(\sigma|\theta)d\sigma.$$

The right-hand side of (4.7) thus becomes

$$-\int_{\underline{\sigma}}^{\overline{\sigma}} q^{sb}(\theta,\sigma)g(\sigma|\theta)\left(\frac{1}{1+\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\frac{F(\theta)}{f(\theta)}} + \frac{\frac{G_\theta(\sigma|\theta)}{g(\sigma|\theta)}\left(\frac{F(\theta)}{f(\theta)}\right)^2}{\left(1+\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\frac{F(\theta)}{f(\theta)}\right)^2}\frac{d}{d\sigma}\left(\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\right)\right)d\sigma. \tag{A.23}$$

Hence, U^{sb} is indeed certainly decreasing when Condition (5.9) holds.

□

Proof of Proposition 4. Remember from the Proof of Proposition 3 that, when looking for a solution to $(\mathcal{P}^r)$, the principal must end up maximizing pointwise the expression (A.17). Suppose that (5.5) fails at some pair (θ,σ). From Assumption 2, this condition also fails for all $\sigma' \leq \sigma$. In other words,

$$1+\frac{g_\theta(\sigma'|\theta)}{g(\sigma'|\theta)}\frac{F(\theta)}{f(\theta)} < 0, \quad \forall \sigma' \in \left[\underline{\sigma}, \sigma^{sb}(\theta)\right] \tag{A.24}$$

where $\sigma^{sb}(\theta) > \underline{\sigma}$ is uniquely defined (again from Assumption 2) by (5.13). when

$$1 + \frac{g_\theta(\underline{\sigma}|\theta)}{g(\underline{\sigma}|\theta)} \frac{F(\theta)}{f(\theta)} < 0. \tag{A.25}$$

Moreover, this latter condition itself requires $\theta \geq \theta^*$ for some $\theta^* > \underline{\theta}$ that solves (5.12), again from Assumption 5.

Pointwise optimization for (A.17) with respect to $b(\theta, \sigma)$ then leads to (5.15). The optimality condition (5.14) is also straightforward. □

PROOF OF PROPOSITION 5. We first consider the relaxed problem $(\mathcal{P}^r)$. Following previous steps, the principal must end up maximizing pointwise the following expression (where we take into account the fact that output is no longer contingent on signal):

$$\mathbb{E}_\theta \left(S(q(\theta)) - \left(\theta + \frac{F(\theta)}{f(\theta)} \right) q(\theta) - \mathbb{E}_\sigma \left(b(\theta, \sigma) \left(1 + \frac{F(\theta)}{f(\theta)} \frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)} \right) |\theta \right) \right). \tag{A.26}$$

It immediately follows that the optimal output is given by (6.3) while, thanks to the bang-bang nature of the problem, the optimal level of money burned is characterized in (6.4).

REGULARITY. Given the structure of the contract laid out in (6.3) and (6.4), we may rewrite the incentive compatibility conditions (4.5) as

$$U^{sb}(\theta) = S(q^{bm}(\theta)) - \theta q^{bm}(\theta) - BG(\sigma^{sb}(\theta)|\theta) - H^{bm}(\theta)$$

$$\geq S(q^{bm}(\hat{\theta})) - \theta q^{bm}(\hat{\theta}) - BG(\sigma^{sb}(\hat{\theta})|\theta) - H^{bm}(\hat{\theta})$$

or

$$U^{sb}(\theta) - U^{sb}(\hat{\theta}) \geq (\hat{\theta} - \theta) q^{bm}(\hat{\theta}) - B(G(\sigma^{sb}(\hat{\theta})|\theta) - G(\sigma^{sb}(\hat{\theta})|\hat{\theta})),\ \forall (\theta, \hat{\theta}) \in \Theta^2.$$

Using the necessary condition (6.2), we rewrite the right-hand side above and get

$$\int_\theta^{\hat{\theta}} \left(q^{bm}(\tilde{\theta}) + BG_\theta(\sigma^{sb}(\tilde{\theta})|\tilde{\theta}) \right) d\tilde{\theta} \geq (\hat{\theta} - \theta) q^{bm}(\hat{\theta}) - B(G(\sigma^{sb}(\hat{\theta})|\theta) - G(\sigma^{sb}(\hat{\theta})|\hat{\theta})),\ \forall (\theta, \hat{\theta}) \in \Theta^2. \tag{A.27}$$

Note that q^{bm} is non-increasing and thus

$$\int_\theta^{\hat{\theta}} q^{bm}(\tilde{\theta}) d\tilde{\theta} \geq (\hat{\theta} - \theta) q^{bm}(\hat{\theta}), \quad \forall (\theta, \hat{\theta}) \in \Theta^2. \tag{A.28}$$

A sufficient condition for (A.27) to hold is thus:

$$\int_\theta^{\hat{\theta}} G_\theta(\sigma^{sb}(\tilde{\theta})|\tilde{\theta}) d\tilde{\theta} \geq G(\sigma^{sb}(\hat{\theta})|\hat{\theta}) - G(\sigma^{sb}(\hat{\theta})|\theta), \quad \forall (\theta, \hat{\theta}) \in \Theta^2. \tag{A.29}$$

Condition (A.29) amounts to

$$\int_{\theta}^{\hat{\theta}} [G_\theta(\sigma^{sb}(\tilde{\theta})|\tilde{\theta}) - G_\theta(\sigma^{sb}(\hat{\theta})|\tilde{\theta})] d\tilde{\theta} \geq 0$$

$$\int_{\theta}^{\hat{\theta}} \int_{\sigma^{sb}(\hat{\theta})}^{\sigma^{sb}(\tilde{\theta})} g_\theta(\sigma|\tilde{\theta}) d\sigma d\tilde{\theta} \geq 0$$

It immediately follows that (A.29) is satisfied when σ^{sb} is non-decreasing and $g_\theta(\sigma|\theta) \leq 0$ for $\sigma \in \left[\sigma^{sb}(\hat{\theta}), \sigma^{sb}(\theta)\right]$ if $\hat{\theta} \geq \theta$ (or $\sigma \in \left[\sigma^{sb}(\theta), \sigma^{sb}(\hat{\theta})\right]$ if $\theta \geq \hat{\theta}$). This last condition holds since Assumption 2 with the definition of $\sigma^*(\theta)$ in (5.13) altogether imply that

$$\text{(A.30)} \qquad \frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)} < -\frac{f(\theta)}{F(\theta)} < 0, \quad \forall \theta \in \left[\theta^*, \overline{\theta}\right], \forall \sigma \leq \sigma^{sb}(\theta).$$

It remains to prove that σ^{sb} is non-decreasing, let us first differentiate (5.13) to get

$$\text{(A.31)} \qquad \frac{\partial}{\partial \sigma}\left(\frac{g_\theta(\sigma^{sb}(\theta)|\theta)}{g(\sigma^{sb}(\theta)|\theta)}\right)\dot{\sigma}^{sb}(\theta) = -\frac{d}{d\theta}\left(\frac{f(\theta)}{F(\theta)}\right) - \frac{\partial}{\partial \theta}\left(\frac{g_\theta(\sigma^{sb}(\theta)|\theta)}{g(\sigma^{sb}(\theta)|\theta)}\right).$$

Observe that the coefficient on the left-hand side is non-negative from Assumption 2 while the right-hand side, is non-negative from Assumption 5 and (5.13).[16] Hence, $\dot{\sigma}^{sb}(\theta) \geq 0$ as requested.

We now demonstrate that $U^{sb}(\theta)$ is decreasing and thus the conjecture that (4.1) can be replaced by (5.3) is correct. It is true when

$$q^{bm}(\theta) \geq -BG_\theta(\sigma^{sb}(\theta)|\theta), \quad \forall \theta \in \Theta;$$

a condition that follows from (6.5).

□

[16] Indeed, we have

$$-\frac{d}{d\theta}\left(\frac{f(\theta)}{F(\theta)}\right) = \frac{\frac{d}{d\theta}\left(\frac{F(\theta)}{f(\theta)}\right)}{\left(\frac{F(\theta)}{f(\theta)}\right)^2}$$

and

$$-\frac{\partial}{\partial \theta}\left(\frac{g_\theta(\sigma^{sb}(\theta)|\theta)}{g(\sigma^{sb}(\theta)|\theta)}\right) = -\frac{\partial}{\partial \theta}\left(\frac{1}{\frac{g(\sigma^{sb}(\theta)|\theta)}{g_\theta(\sigma^{sb}(\theta)|\theta)}}\right) = \frac{\frac{\partial}{\partial \theta}\left(\frac{g(\sigma^{sb}(\theta)|\theta)}{g_\theta(\sigma^{sb}(\theta)|\theta)}\right)}{\left(\frac{g(\sigma^{sb}(\theta)|\theta)}{g_\theta(\sigma^{sb}(\theta)|\theta)}\right)^2}.$$

From (5.13), it follows that $\left(\frac{g(\sigma^{sb}(\theta)|\theta)}{g_\theta(\sigma^{sb}(\theta)|\theta)}\right)^2 = \left(\frac{F(\theta)}{f(\theta)}\right)^2$, and thus:

$$-\frac{d}{d\theta}\left(\frac{f(\theta)}{F(\theta)}\right) - \frac{\partial}{\partial \theta}\left(\frac{g_\theta(\sigma^{sb}(\theta)|\theta)}{g(\sigma^{sb}(\theta)|\theta)}\right) = \frac{\frac{\partial}{\partial \theta}\left(\frac{F(\theta)}{f(\theta)} + \frac{g(\sigma^{sb}(\theta)|\theta)}{g_\theta(\sigma^{sb}(\theta)|\theta)}\right)}{\left(\frac{F(\theta)}{f(\theta)}\right)^2} \geq 0,$$

which is non-negative by Assumption 5.

PROOF OF PROPOSITION 6. We use Jullien (2000)'s and Martimort and Stole (2022)'s techniques to characterize the solution (denoted as with a supercript $r\prime$) to $(\mathcal{P}^{r\prime})$. We thus associate to the participation constraint (4.1), a probability measure μ defined over the Borel subsets of Θ with an associated adjoint function, $M : \Theta \to [0, 1]$, defined by $M(\underline{\theta}) = 0$, $M(\overline{\theta}^{+}) = 1$ and

$$M(\theta) = \int_{[\underline{\theta},\theta)} \mu(d\tilde{\theta}) \text{ for } \theta > \underline{\theta}.$$

Necessary and sufficient conditions (thanks to strict concavity of the maximand) are then stated in terms of thus probability measure which serves to first express a complementary slackness condition:

$$\int_{\underline{\theta}}^{\overline{\theta}} U^{r\prime}(\tilde{\theta})\mu(d\tilde{\theta}) = 0. \tag{A.32}$$

This condition states that the probability measure μ has support $\Omega = \{\theta : U^{r\prime}(\theta) = 0\}$. When $\overset{\circ}{\Omega} \neq \emptyset$, we may differentiate $U^{r\prime}(\theta) = 0$ with respect to θ and, using (6.2), get (6.14).

Second, the optimality conditions write as

(A.33)

$$(q^{r\prime}(\theta), b^{r\prime}(\theta,\sigma)) \in \arg\max_{(q,b(\cdot))} \Big\{ f(\theta)\left(S(q) - \theta q - \mathbb{E}_{\sigma}\left(b(\sigma)|\theta\right)\right) - (F(\theta) - M(\theta))\left(q + \mathbb{E}_{\sigma}\left(b(\sigma)\frac{g_{\theta}(\sigma|\theta)}{g(\sigma|\theta)}|\theta\right)\right)\Big\} \text{ for a.e. } \theta \in \Theta.$$

The first-order optimality condition with respect to output q is given by (6.8). Because $M(\theta) \geq 0$, it immediately follows that (6.9) holds.

Given the linearity of the objective function on the right-hand side of (A.33) in $b(\cdot)$, the optimality condition with respect to the amount of money burned is given by then bang-bang solution (6.10). Observe that (6.14) requires $\sigma^{r\prime}(\theta) > \underline{\sigma}$ for all $\theta \in \overset{\circ}{\Omega} \neq \emptyset$. Thus, (6.13) must necessarily hold.

Finally, when $q^{r\prime}(\theta) \leq q^{fb}(\theta)$ implies $M(\theta) \leq F(\theta)$. Hence, (6.12) implies

$$\frac{g_{\theta}(\sigma^{r\prime}(\theta)|\theta)}{g(\sigma^{r\prime}(\theta)|\theta)} = -\frac{f(\theta)}{F(\theta) - M(\theta)} < -\frac{f(\theta)}{F(\theta)} = \frac{g_{\theta}(\sigma^{sb}(\theta)|\theta)}{g(\sigma^{sb}(\theta)|\theta)}, \quad \forall\theta \in \left[\theta^{*}, \overline{\theta}\right]. \tag{A.34}$$

Assumption 2 then implies (6.12).

INCENTIVE COMPATIBILITY. The proof of that (4.5) hold provided that $q^{r\prime}$ is non-increasing and $\sigma^{r\prime}$ non-decreasing is similar to that made in the Proof of Proposition 5 and is thus omitted. □

Appendix B. A Revelation Principle

We provide here a proof of the Revelation Principle that is used in the main text. It follows quite closely the one formulated in Appendix B in Dequiedt and Martimort (2015). A general mechanism in this case consists in a message spaces $\mathcal{M}_A$ and $\mathcal{M}_P$ respectively for the agent and the principal, and an allocation rule x that maps the set $\mathcal{M}_A \times \mathcal{M}_P$ into the set $\Delta(\mathcal{Q} \times \mathcal{T} \times \mathcal{B})$ of probability distributions over $\mathcal{Q} \times \mathcal{T} \times \mathcal{B}$, where we remind the reader that $\mathcal{Q}$ is the set of possible quantities and $\mathcal{T}$ is the set of possible transfers. The agent's strategy maps Θ into the set $\Delta(\mathcal{M}_A)$ of probability distributions over $\mathcal{M}_A$. The principal's strategy maps $\Sigma \times \mathcal{M}_a$ into the set $\Delta(\mathcal{M}_P)$ of probability distributions over $\mathcal{M}_P$. Fixing the mechanism allocation x, a message pair (m_A, m_P) induces the following utilities for principal and agent respectively:

$$\mathcal{V}(x(m_A, m_P)) = \int_{\mathcal{Q}\times\mathcal{T}\times\mathcal{B}} v(q, t+b) dx(q, t, b | m_A, m_P)$$

$$\mathcal{U}(x(m_A, m_P); \theta) = \int_{\mathcal{Q}\times\mathcal{T}\times\mathcal{B}} u(q, t; \theta) dx(q, t, b | m_A, m_P)$$

The timing is the following:

1. The principal proposes a mechanism $\{\mathcal{M}_A, \mathcal{M}_P, x\}$
2. The agent and the principal respectively learn θ and σ.
3. The agent decides whether to accept the mechanism. In the affirmative case, the agent sends a message $m_A \in \mathcal{M}_A$. Otherwise, both get their reservation values that are normalised to zero.
4. The principal observes the message m_a, and then send a message $m_P \in \mathcal{M}_P$.
5. The allocation rule x inputs (m_A, m_P) and outputs $(q, t, b) \in \mathcal{Q} \times \mathcal{T} \times \mathcal{B}$.
6. Payoffs are realised, and the interaction ends.

The appropriate equilibrium concept for the induced game after acceptance is that of Perfect Bayesian Equilibrium. Thus, after observing σ and m_A, the principal updates her beliefs about θ according to Bayes' rule whenever applicable. Similarly, the agent updates his beliefs about σ after observing θ. The message strategies must be optimal for each player:

(B.1)

$$\forall \sigma \in \Sigma : \ \forall m_A \in \mathcal{M}_A : \ \forall m_P \in \mathbf{supp}\left(m_P^*(m_A, \sigma)\right) : \ m_P \in \arg \max_{\hat{m}^P \in \mathcal{M}_P} \mathbb{E}_\theta \left(\mathcal{V}(x(m_A, \hat{m}_P)) \mid m_A, \sigma\right)$$
$$= \arg \max_{\hat{m}_P \in \mathcal{M}_P} \mathcal{V}(x(m_A, \hat{m}_P))$$

(B.2)

$$\forall \theta \in \Theta : \forall m_A \in \mathbf{supp}(m_A^*(\theta)) : m_A \in \arg \max_{\hat{m}_A \in \mathcal{M}_A} \mathbb{E}_\sigma \left(\int_{\mathcal{M}_P} \mathcal{U}(x(\hat{m}_A, m_P); \theta) dm_P^*(m_P | \ \hat{m}_A, \sigma) \middle| \theta \right)$$

Where the second line follows because we are in a private values context in which θ does not enter principal's utility. Moreover, because we are in the continuation game after acceptance, we must have:

$$\text{(B.3)} \quad \forall \theta \in \Theta: \ \forall m_A \in \mathbf{supp}(m_A^*(\theta)): \ \mathbb{E}_\sigma\left(\int_{\mathcal{M}_P} \mathcal{U}(x(m_A, m_P); \theta) dm_P^*(m_P | \ \hat{m}_A, \sigma) \,\middle|\, \theta\right) \geq 0$$

We now consider the direct revelation mechanism $\{\Theta, \Sigma, \tilde{x}\}$, where Θ is the set of reports of the agent, Σ is the set of reports of the principal, and $\tilde{x}$ is the allocation rule defined as:

$$\text{(B.4)} \quad \forall \hat{\theta} \in \Theta: \ \forall \hat{\sigma} \in \Sigma: \ \tilde{x}(\hat{\theta}, \hat{\sigma}) = x(m_A^*(\hat{\theta}), m_P^*(m_A^*(\hat{\theta}), \hat{\sigma}))$$

PROPOSITION 7. *Any allocation rule achieved in a continuation equilibrium* (m_A^*, m_P^*) *with the offer and acceptance of* $(\mathcal{M}_A, \mathcal{M}_P, x)$ *can be replicated with a direct revelation mechanism* $\{\Theta, \Sigma, \tilde{x}\}$ *satisfying (B.4), acceptance and truthful reporting:*

$$\text{(B.5)} \quad \forall \theta \in \Theta: \ \mathbb{E}_\sigma\left(\mathcal{U}\left(\tilde{x}(\theta, \sigma); \theta\right) | \ \theta\right) \geq 0$$

$$\text{(B.6)} \quad \forall \theta \in \Theta: \ \mathbb{E}_\sigma\left(\mathcal{U}\left(\tilde{x}(\theta, \sigma); \theta\right) | \ \theta\right) = \max_{\hat{\theta} \in \Theta} \mathbb{E}_\sigma\left(\mathcal{U}\left(\tilde{x}(\hat{\theta}, \sigma); \theta\right) \,\middle|\, \theta\right)$$

$$\text{(B.7)} \quad \forall \theta \in \Theta: \forall \sigma \in \Sigma: \ \mathcal{V}\left(\tilde{x}(\theta, \sigma)\right) = \max_{\hat{\sigma} \in \Sigma} \mathcal{V}\left(\tilde{x}(\theta, \hat{\sigma})\right)$$

Proof. We use the following notations:

$$\mathcal{U}\left(x(m_A^*(\theta), m_P^*(m_A^*(\theta), \sigma)); \theta\right) = \int_{\mathcal{M}_A} \int_{\mathcal{M}_P} \mathcal{U}\left(x(m_A, m_P); \theta\right) dm_P^*(m_P | m_A, \sigma) dm_A^*(m_A | \theta)$$

$$\mathcal{V}\left(x(m_A^*(\theta), m_P^*(m_A^*(\theta), \sigma))\right) = \int_{\mathcal{M}_A} \int_{\mathcal{M}_P} \mathcal{V}\left(x(m_A, m_P)\right) dm_P^*(m_P | m_A, \sigma) dm_A^*(m_A | \theta)$$

Condition (B.5) follows straightforwardly from the combination of (B.3) and (B.4):

$$\mathbb{E}_\sigma\left(\mathcal{U}\left(\tilde{x}(\theta, \sigma); \theta\right) | \ \theta\right) \underbrace{=}_{\text{by (B.4)}} \mathbb{E}_\sigma\left(\mathcal{U}\left(x(m_A^*(\theta), m_P^*(m_A^*(\theta), \sigma)); \theta\right) | \ \theta\right) \underbrace{\geq}_{\text{by (B.3)}} 0$$

Condition (B.6) is implied by (B.2) and (B.4):

$$\begin{aligned} \mathbb{E}_\sigma\left(\mathcal{U}\left(\tilde{x}(\theta, \sigma); \theta\right) | \ \theta\right) &\underbrace{=}_{\text{by (B.4)}} \mathbb{E}_\sigma\left(\mathcal{U}\left(x(m_A^*(\theta), m_P^*(m_A^*(\theta), \sigma)); \theta\right) | \ \theta\right) \\ &\underbrace{\geq}_{\text{by (B.2)}} \mathbb{E}_\sigma\left(\mathcal{U}\left(x(m_A^*(\hat{\theta}), m_P^*(m_A^*(\hat{\theta}), \sigma)); \theta\right) \,\middle|\, \theta\right) \\ &\underbrace{=}_{\text{by (B.4)}} \mathbb{E}_\sigma\left(\mathcal{U}\left(\tilde{x}(\hat{\theta}, \sigma); \theta\right) \,\middle|\, \theta\right) \end{aligned}$$

where we took $\hat{m}_A \in \mathbf{supp}\left(m_A^*(\hat{\theta})\right)$ in the second line. Finally, condition (B.7) is implied by conditions (B.1) and (B.4):

$$\begin{aligned}\mathcal{V}\left(\tilde{x}(\theta,\sigma)\right) &\underbrace{=}_{\text{by (B.4)}} \mathcal{V}\left(x(m_A^*(\theta), m_P^*(m_A^*(\theta),\sigma))\right)\\ &\underbrace{\geq}_{\text{by (B.1)}} \mathcal{V}\left(x(m_A^*(\theta), m_P^*(m_A^*(\theta),\hat{\sigma}))\right)\\ &\underbrace{=}_{\text{by (B.4)}} \mathcal{V}\left(\tilde{x}(\theta,\hat{\sigma})\right)\end{aligned}$$

where we took $\hat{m}_P \in \mathbf{supp}\left(m_P(m_A^*(\theta),\hat{\sigma})\right)$ in the second line.

□

Appendix C. Participation Constraints

Propositions 3 and 4 are valid provided that $\dot{U}(\theta) < 0$ for all $\theta < \overline{\theta}$, and that global IC constraints are all met. However, the informativeness of the signal may be strong enough to bring the rents of some positive mass of types to zero. Thus, in a similar fashion to that in Proposition 6, we consider a partially relaxed problem where the participation constraint writes as (4.1), a pure state constraint, while incentive compatibility constraints (4.5) are still reduced to the Envelope Condition (4.7). The solution to this problem is now characterized.

Proposition 4B. *The solution to the relaxed problem $(\mathcal{P}^{r\prime})$ entails the following properties.*

1. *There exists a probability measure μ whose support is $\Omega = \{\theta : U^{r\prime}(\theta) = 0\}$ such that*
(C.1)
$$\begin{cases} q^{r\prime}(\theta,\sigma) = 0 & \text{if } 1 + \frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\frac{F(\theta)-M(\theta)}{f(\theta)} \leq 0 \text{ and } F(\theta) \geq M(\theta),\\ S'(q^{r\prime}(\theta)) = \theta + \frac{\frac{F(\theta)-M(\theta)}{f(\theta)}}{1+\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\frac{F(\theta)-M(\theta)}{f(\theta)}}, & \text{otherwise.}\end{cases}$$
where $M(\theta) = \int_{[\underline{\theta},\theta)} \mu(d\tilde{\theta})$.

2. *Money is burned following very bad news only*
(C.2)
$$b^{r\prime}(\theta,\sigma) = \begin{cases} B & \text{if } 1 + \frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\frac{F(\theta)-M(\theta)}{f(\theta)} \leq 0,\\ 0 & \text{otherwise.}\end{cases}$$

3. *And*
(C.3)
$$\mathbb{E}_\sigma\left(q^{r\prime}(\theta,\sigma)\big|\theta\right) = \mathbb{E}_\sigma\left(\left(S(q^{r\prime}(\theta,\sigma)) - \theta q^{r\prime}(\theta,\sigma) - b^{r\prime}(\theta,\sigma)\right)\frac{g_\theta(\sigma|\theta)}{g(\sigma|\theta)}\middle|\theta\right), \quad \forall\theta \in \mathring{\Omega}.$$

Proof of Proposition 4B. The proof replicates that of Proposition 6, and is therefore omitted. The main difference stems from the objective function of the Principal's problem:

(C.4)

$$(q^{r\prime}(\theta,\sigma), b^{r\prime}(\theta,\sigma)) \in \arg\max_{(q,b)} \Big\{ f(\theta)g(\sigma|\theta)\left(S(q)-\theta q - b\right) + \left[F(\theta)-M(\theta)\right]\left[-qg(\sigma|\theta) + g_\theta(\sigma|\theta)\left(S(q)-\theta q - b\right)\right] \Big\} \text{ for a.e. } \theta \in \Theta.$$

□