

# Optimal Delegation with Multi-Dimensional Decisions <sup>\*</sup>

Frédéric KOESSLER <sup>†</sup>

David MARTIMORT <sup>‡</sup>

June 9, 2010

## Abstract

This paper investigates optimal communication mechanisms in a framework with a two-dimensional decision space and no monetary transfers. If the conflicts of interests between the principal and the agent are different on each dimension, the principal can better extract information from the agent by using the spread between the two decisions as a costly and type-dependent screening device. In that situation, delegation sets no longer trade off inflexible rules and full discretion but instead take more complex shapes. We use advanced results from the calculus of variations to ensure existence of a solution and derive sufficient and necessary conditions for optimality. The optimal mechanism is continuous and deterministic. The agent's informational rent, the average decision and its spread are strictly monotonic in the agent's type. The comparison of the optimal mechanism with the standard one-dimensional mechanisms explains how competition between different principals controlling various dimensions of the agent's activities impedes information revelation.

KEYWORDS: Communication; Delegation; Mechanism Design; Multi-Dimensional Decision.

JEL CLASSIFICATION: D82; D86.

## 1 Introduction

Consider an informed agent who contracts with an uninformed principal. When the principal's and the agent's interests are conflicting, the principal may want to exert some ex ante control on the agent by restricting the decision set from which the agent may pick actions. Examples of such constrained delegation abound across all fields of economics and political science. A firm's CEO controls division managers by designing capital budgeting rules and allocating decision rights among unit managers.<sup>1</sup> Many different aspects of the firm's decisions related to product design and quality, prices, or polluting emissions are instead scrutinized by regulators. Lastly, Congress Committees exert ex ante control on better informed regulatory agencies by designing various administrative procedures and rules that limit bureaucratic drift.<sup>2</sup>

---

<sup>\*</sup>We thank Thomas Palfrey for helpful discussions at the early stage of this project, Ricardo Alonso, Wouter Dessein, Jeffrey Ely, Larry Samuelson, Aggey Semenov, Shan Zhao and several anonymous referees for useful comments and suggestions. The usual disclaimer applies.

<sup>†</sup>Paris School of Economics – CNRS.

<sup>‡</sup>Toulouse School of Economics – EHESS.

<sup>1</sup>Harris and Raviv (1996) and Alonso et al. (2008).

<sup>2</sup>McCubbins et al. (1987), Huber and Shipan (2002), Epstein and O'Halloran (1999).

Those examples share the common feature that principals hardly use monetary transfers to control their agents. Following the seminal works of Holmström (1984) and Melumad and Shibano (1991), those settings are fruitfully analyzed as mechanism design problems in which the principal can commit to a decision rule but cannot use monetary transfers to implement that rule.<sup>3</sup> With no transfers and when actions lie in a one-dimensional set, optimal communication mechanisms look crude enough. Quite intuitively, the principal finds it hard to induce information revelation and align conflicting objectives when he controls only a single action of the agent. In a one-dimensional setting, an optimal mechanism balances the flexibility gains of letting the agent choose freely this action according to his own private information and the agency cost coming from the fact that the principal and the agent might have conflicting objectives.

The first major result provided by the existing literature highlights the trade-off between *rules* and *discretion* that arises in such contexts. Inflexible rules allow the principal to choose his most preferred policy although they make no use of private information. Leaving discretion to the agent allows to implement state dependent actions but those choices reflect now only the agent's preferences and not those of the principal. The second important result pushed forward by the literature is that the optimal mechanism (when continuous) can be implemented by means of simple delegation sets which put bounds on the agent's action. This is an important theoretical insight because it reduces the design of the mechanism to finding those bounds. This simplification has also a great value in view of the implementation of the optimal mechanism in practice.

The objective of this paper is to study how those results and optimal communication mechanisms are modified when several of the agent's activities can be controlled by a principal or, equivalently, when several principals, each controlling a single decision of the agent, can cooperate in designing a common communication mechanism. First, one may wonder whether the trade-off between rules and discretion remains. Clearly screening possibilities improve and rules seem less attractive but by how much? Second, in a multi-dimensional context, the agency problem between the principal and his agent may not only be related to their average conflict of interests over all dimensions but also to the distribution of those conflicts across the different activities. The extent by which it is so must also be clarified. These questions are highly relevant not only from a pure theoretical viewpoint but, as we saw above, many real world problems involve indeed the control by a principal of an agent of several of his activities. In these contexts, it is important to understand whether and how looking at each dimension separately rules out important possibilities to limit the agent's informational advantage.

To give a first pass on these questions, the present paper investigates the form of optimal communication mechanisms in a simple two-dimensional setting with quadratic and separable payoffs and a uniform types distribution. We characterize the optimal communication mechanism in such an environment and show that it tights the agent's choices on each dimension through a

---

<sup>3</sup>Armstrong (1994), Baron (2000), Martimort and Semenov (2006), Alonso and Matouschek (2008), Goltsman et al. (2009), and Kovac and Mylovanov (2009), among others.

smooth and deterministic delegation set. When the conflicts of interests between the principal and the agent on both dimensions differ, the optimal delegation set never crosses the agent's ideal points and never exhibits any pooling. Hence, the trade-off between rules and discretion highlighted by the existing literature disappears.

Intuitively, what matters from the incentive viewpoint is, on the one hand, the average decision that the principal would like to implement and, on the other hand, its spread, i.e., how far apart the levels of each activity are one from the other. That spread plays a role similar albeit somewhat different than what transfers play in standard models with quasi-linear preferences and monetary payments. The similarity comes from the fact that, as transfers in usual screening models, using a type-dependent spread as a screening device facilitates information revelation. Suppose indeed that the agent would ideally like to choose the same decision on dimensions 1 and 2 of his activity but, on average, prefers lower levels of those decisions compared with the principal's ideal points. In such settings, the principal wants to control the agent to avoid information manipulations that aim at implementing such lower levels of activities. To limit those incentives to claim for lower average decisions, the principal might pull decisions 1 and 2 further apart one from the other following such claims making it more costly for the agent to call for lower levels of activity. Instead, the principal might also reduce the spread when the agent reports some information that aim at implementing higher levels of activities. Incentives punishments and rewards are possible by playing on the spread. The spread, or transfer, is indeed strictly positive and monotonic in the agent's type whenever the conflicts of interests between the principal and the agent are different on each dimension.

Compared with a setting where monetary transfers are available, one util left to the agent no longer costs one util to the principal. Implementing such spread in the decisions introduces also some nonlinear costs and benefits for the principal. Suppose indeed that the principal would ideally prefer one more unit of activity 2 for any realization of the agent's private information, his ideal spread is just one and, on top, both activities 1 and 2 give more return to the principal. In that case, the agent may want to pretend that lower average activities should be implemented. Increasing the spread between those activities above 1 for low average activity levels and at the same time decreasing it below 1 for higher activity levels is of course costly for the principal but it also facilitates screening.

From a technical viewpoint, this non-linearity due to the absence of monetary payments makes the characterization of the optimal communication mechanism quite complex.<sup>4</sup> We use results from the calculus of variations (Clarke, 1990) to ensure existence of a solution and derive sufficient and necessary conditions for optimality.

---

<sup>4</sup>The absence of monetary transfers makes the contracting problem looks a bit like those that are solved in the optimal taxation literature (Mirrlees, 1971) where also utility functions are not quasi-linear. The techniques of this literature cannot nevertheless be directly imported into our framework. Even with quadratic payoffs the principal's objective function may not be everywhere Lipschitz-continuous contrary to what is supposed in the optimal taxation literature.

*Related Literature.* Melumad and Shibano (1991) provided a significant analysis of the delegation problem with quadratic payoffs and a uniform types distribution in contexts where no transfers are available and where the uninformed party (the principal) commits to a communication mechanism with the uninformed party (the agent). Martimort and Semenov (2006) and Alonso and Matouschek (2008) characterized settings where simple connected delegation sets are optimal, a feature that was a priori assumed in Holmström (1984), Armstrong (1994) and Baron (2000) for instance. Alonso and Matouschek (2007) have brought the standard delegation model to a dynamic context where the principal and the agent repeatedly interact. Focusing on dominant strategy to get a sharp characterization of the set of incentive feasible allocations, Martimort and Semenov (2008) have instead extended this mechanism design approach to the case of multiple privately informed agents (lobbyists) dealing with a single principal (a Legislature) in a political economy context where the principal chooses a one-dimensional policy.<sup>5</sup> Farrell and Gibbons (1989) and Goltsman and Pavlov (2009) have analyzed private and public communication with a single informed agent and two decision-makers. As in our model, the decisions on each dimension enter separately into the agent’s payoff function and are strategically independent across the two decision-makers. None of these papers has addressed the design of multi-dimensional communication with commitment. In that respect, the closest paper to ours might be Ambrus and Egorov (2010). These authors introduce the possibility that the principal “burns money” or imposes costly activities for the agent in an otherwise standard delegation set-up. Money burning offers a second instrument that facilitates screening although the impact of those new screening possibilities on the principal’s payoff is different than in our paper.

*Organization of the paper.* Section 2 presents the model and the by-now standard result where a single activity of the agent is controlled by the principal. Section 3 presents some preliminary results on incentive compatibility and assesses the performances of simple and intuitive mechanisms that take into account the new possibilities that incentive compatibility in multi-dimensional environments open. Section 4 is the core of the paper. We formulate the design problem using advanced tools of the calculus of variations and we derive the optimal multi-dimensional mechanism. Some robustness checks are provided in Section 5. Section 6 concludes and paves the way for future research. Proofs are relegated to the Appendix.

## 2 The Model

A principal controls two actions,  $x_1$  and  $x_2$ , undertaken by a single agent on his behalf. We denote by  $(x_1, x_2)$  the bi-dimensional vector of those actions. For simplicity, those actions lie in a compact set  $\mathcal{K} = [-K, K] \subseteq \mathbb{R}$  for  $K$  large enough. Utility functions are single-peaked, quadratic and

---

<sup>5</sup>Austen-Smith (1993), Battaglini (2002, 2004), Krishna and Morgan (2001), Levy and Razin (2007), and Ambrus and Takahashi (2008) have instead considered cheap talk settings with multiple privately informed senders.

respectively given for the principal and his agent by:<sup>6</sup>

$$V(x_1, x_2, \theta) = -\frac{1}{2} \sum_{i=1}^2 (x_i - \theta - \delta_i)^2 \quad (1)$$

and

$$U(x_1, x_2, \theta) = -\frac{1}{2} \sum_{i=1}^2 (x_i - \theta)^2. \quad (2)$$

With those preferences, the agent's ideal point on each dimension is  $x_A = \theta$  whereas the principal has an ideal point located at  $x_{iP} = \theta + \delta_i$  in dimension  $i$ ,  $i = 1, 2$ . The principal is biased in the same direction on both dimensions but his preferences may be more or less congruent with those of the agent, i.e.,  $0 \leq \delta_1 \leq \delta_2$ . For further references, we denote by  $\Delta \equiv \delta_2 - \delta_1$  the difference in biases between the two dimensions and by  $\delta \equiv \frac{\delta_1 + \delta_2}{2}$  the average bias.

The agent has private information on his ideal point  $\theta$  (or type), that is drawn from a uniform distribution on  $\Theta = [0, 1]$ .<sup>7</sup> The principal is not informed about the agent's type.

The principal controls the whole vector of the agent's activities  $(x_1, x_2)$ . From the Revelation Principle (Myerson, 1982), there is no loss of generality in restricting the analysis to direct communication mechanisms stipulating (maybe stochastic<sup>8</sup>) decisions as functions of the agent's report on his type. Any deterministic communication mechanism is a mapping  $x(\cdot) = \{x_1(\cdot), x_2(\cdot)\} : \Theta \rightarrow \mathcal{K}^2$ .

The model can also be interpreted as a situation with two principals,  $P_1$  and  $P_2$ , a common agent, and the following utility function for principal  $P_i$ ,  $i = 1, 2$ :

$$V_i(x_i, \theta) = -\frac{1}{2} (x_i - \theta - \delta_i)^2.$$

Under a non-cooperative design and private communication between the agent and each principal, principals independently choose their communication spaces with the agent and design their own mechanism. Since there is no externality between principals (i.e., in our context, each principal  $P_i$ 's utility only depends on the decision  $x_i$  and the agent's utility function is separable in the decisions controlled by each principal), each principal offers the same communication mechanism as if he was alone contracting with the agent. If principals cooperate in designing a communication mechanism and have equal bargaining powers, then the merged principal objective function is exactly the one of Equation (1).

For further references, let us consider the case where the principal controls a single decision  $x_i$ , for some  $i \in \{1, 2\}$  or, with the interpretation above, principals independently choose their communication spaces with the agent and design their own mechanism.

---

<sup>6</sup>The choice of quadratic utility functions is standard in the literature. This assumption is reasonable if one sees it as a Taylor approximation of more general utility functions in a context where actions would not move much around the agent's ideal point.

<sup>7</sup>The characterization of the contractual outcomes would be untractable if we were assuming other distributions.

<sup>8</sup>We postpone to Subsection 5.1 the analysis of stochastic mechanisms proving there their suboptimality.

**Proposition 1 (One-Dimensional Activity)** *In the one-dimensional case, the optimal communication mechanism  $x_i^O(\cdot)$  is given by:*

$$x_i^O(\theta) = \max \{ \theta, \theta_i^O \}, \text{ where } \theta_i^O = 2\delta_i < 1. \quad (3)$$

When a principal controls a single dimension of the agent's activity, the optimal communication mechanism has a simple structure: The optimal action corresponds to the agent's ideal point if it is large enough and is otherwise independent of the agent's type. This outcome can be easily achieved by means of a simple *delegation set*. Instead of using a direct revelation mechanism and communicating with the agent, the principal could as well offer a menu of options  $\mathcal{D}_i = [\theta_i^O, +\infty)$  and let the agent freely choose within this set. When the floor  $\theta_i^O$  is not binding, the agent is not constrained by the principal and everything happens as if he had full discretion in choosing his own ideal point. When the floor is instead binding, the agent is constrained and cannot choose his bliss point which is too low compared with what the principal would implement himself.

The optimal communication mechanism trades off the benefits of flexibility (the agent choosing sometimes a state-dependent action) against the loss of control it implies (this state-dependent action being different from the principal's ideal point). Setting a floor  $\theta_i^O$  limits the agent's discretion and reduces the loss of control. Clearly,  $\theta_i^O$  increases with  $\delta_i$  meaning that a less rigid rule is chosen when the conflict of interests between the principal and the agent is less pronounced.

### 3 Preliminary Results

In the multi-dimensional case, incentive compatibility constraints can be written as:

$$\theta \in \arg \max_{\hat{\theta} \in \Theta} -\frac{1}{2} \sum_{i=1}^2 (x_i(\hat{\theta}) - \theta)^2.$$

**Lemma 1** *The necessary and sufficient condition for incentive compatibility is that  $\sum_{i=1}^2 x_i(\theta)$  is non-decreasing in  $\theta$  and thus a.e. differentiable in  $\theta$ . At any differentiability point, we have:*

$$\sum_{i=1}^2 \dot{x}_i(\theta) \geq 0, \quad (4)$$

$$\sum_{i=1}^2 \dot{x}_i(\theta)(x_i(\theta) - \theta) = 0. \quad (5)$$

In this multi-dimensional world, the principal can now use both  $x_1(\cdot)$  and  $x_2(\cdot)$  to screen the agent's preferences. To understand how it can be so, it is useful first to observe that the principal could at least offer the optimal one-dimensional communication mechanisms he would offer for each dimension, namely the pair of mechanisms described in Proposition 1. Although communication mechanisms that would satisfy (3) for  $i = 1, 2$  also satisfy (4) and (5), more communication

mechanisms are now incentive compatible. By trading off distortions along each dimension or by choosing actions that vary in opposite directions on each dimension as the agent's type changes, the principal can introduce countervailing incentives which might facilitate information revelation.<sup>9</sup>

This characterization of incentive compatible allocations already gives some powerful insights on the properties of mechanisms by looking at a couple of simple two-dimensional communication mechanisms.

**Example 1** Consider the linear communication mechanism  $\{x_1^\alpha(\theta), x_2^\alpha(\theta)\}_{\theta \in \Theta}$  such that  $x_1^\alpha(\theta) = \theta - \alpha$  and  $x_2^\alpha(\theta) = \theta + \alpha$  where  $\alpha$  is a fixed number. This mechanism is incentive compatible since it satisfies both (4) and (5). The best of such communication mechanisms maximizes the principal's profit, i.e.,  $\alpha$  should be optimally chosen so that any concession made by the principal on  $x_1$  by moving this decision closer to the agent's own ideal point is compensated by an equal shift in  $x_2$  in the direction of the principal's ideal point. Typically,  $\alpha = \frac{\Delta}{2}$  does the trick since

$$\arg \min_{\alpha} \int_0^1 \left( \sum_{i=1}^2 (x_i^\alpha(\theta) - \theta - \delta_i)^2 \right) d\theta = \arg \min_{\alpha} (\alpha + \delta_1)^2 + (\alpha - \delta_2)^2 = \frac{\Delta}{2}.$$

Still the principal can find that such decisions are too close to the agent's ideal points. The mechanism  $\{x_1^\alpha(\theta), x_2^\alpha(\theta)\}_{\theta \in \Theta}$  can be improved upon by introducing a pooling area as in the one-dimensional case. This trick is used in the next example.

**Example 2** Consider the incentive compatible mechanism  $\{\tilde{x}_1(\theta), \tilde{x}_2(\theta)\}_{\theta \in \Theta}$  defined as:

$$\tilde{x}_1(\theta) = \begin{cases} \tilde{\theta} - \frac{\Delta}{2} & \text{if } \theta \leq \tilde{\theta} \\ \theta - \frac{\Delta}{2} & \text{otherwise} \end{cases} \quad \text{and} \quad \tilde{x}_2(\theta) = \begin{cases} \tilde{\theta} + \frac{\Delta}{2} & \text{if } \theta \leq \tilde{\theta} \\ \theta + \frac{\Delta}{2} & \text{otherwise} \end{cases} \quad (6)$$

This new mechanism is obtained by piecing together a floor on policy for  $\theta \leq \tilde{\theta}$ . The optimal mechanism within this class, that we denote thereafter  $\{\tilde{x}_1^*(\theta), \tilde{x}_2^*(\theta)\}_{\theta \in \Theta}$ , is such that

$$\tilde{\theta}^* = \arg \min_{\tilde{\theta} \in \Theta} \int_0^1 \left( \sum_{i=1}^2 (\tilde{x}_i(\theta) - \theta - \delta_i)^2 \right) d\theta = \arg \min_{\tilde{\theta} \in \Theta} \int_0^{\tilde{\theta}} (\tilde{\theta} - \theta - \delta)^2 d\theta + \int_{\tilde{\theta}}^1 \delta^2 d\theta = 2\delta.$$

This mechanism has a non-trivial pooling area since  $2\delta < 1$ . The principal limits now the pooling area to an average between the pooling areas that he would choose when designing an optimal mechanism on each dimension separately.

Example 2 is instructive because it stresses two aspects of optimal mechanisms that our more general analysis will confirm. First, the principal trades off distortions on each dimension by introducing a spread between  $x_1$  and  $x_2$ . Second, decision rules should be rather flat on the lower

---

<sup>9</sup>The literature on countervailing incentives (Lewis and Sappington, 1989a,b and Laffont and Martimort, 2002, Chapter 3 among others) has been developed in settings with monetary transfers. Sometimes those models generate pooling as an optimal response to incentives to over- and under-report types as in Lewis and Sappington (1989b). On the contrary, in our model pooling is never an issue as we show below.

tail of the distribution. However, contrary to the simple mechanism  $\{\tilde{x}_1^*(\theta), \tilde{x}_2^*(\theta)\}_{\theta \in \Theta}$ , the optimal mechanism will not exhibit any pooling, the spread between the two dimensions will not be constant with the agent's type, and one action will be decreasing on the upper and lower tails of the distribution.

## 4 Optimal Multi-Dimensional Mechanism

To characterize the optimal mechanism, it is useful to use a new set of variables to reparameterize our problem. This transformation shall not only bring new insights on the nature of the economic problem but it will also allow us to easily show later on that stochastic or discontinuous mechanisms are not optimal. Consider thus the following two extra auxiliary variables which are the *average* decision and a measure of the *spread* of those decisions:

$$x(\theta) \equiv \frac{1}{2} \sum_{i=1}^2 x_i(\theta) \text{ and } t(\theta) \equiv \frac{1}{2} \sum_{i=1}^2 (x_i(\theta) - x(\theta))^2 = \frac{1}{4} (x_2(\theta) - x_1(\theta))^2. \quad (7)$$

Note that, under complete information, the principal would like to choose an average decision  $x_P(\theta) = \theta + \delta$  and an optimal spread  $t_P(\theta) = \frac{\Delta^2}{4}$ . These two quantities differ from those that would be ideally chosen by the agent on his own, namely,  $x_A(\theta) = \theta$  and  $t_A(\theta) = 0$ .

Solving this system of equations for  $x_1(\theta)$  and  $x_2(\theta)$  yields immediately:

$$x_1(\theta) = x(\theta) - \sqrt{t(\theta)} \text{ and } x_2(\theta) = x(\theta) + \sqrt{t(\theta)}. \quad (8)$$

Define now the agent's non-positive *information rent*  $U(\theta)$  as:

$$U(\theta) \equiv \max_{\hat{\theta} \in \Theta} -\frac{1}{2} \left( \sum_{i=1}^2 (x_i(\hat{\theta}) - \theta)^2 \right).$$

Using (8) and incentive compatibility, we rewrite:

$$U(\theta) = -(x(\theta) - \theta)^2 - t(\theta) = \max_{\hat{\theta} \in \Theta} -(x(\hat{\theta}) - \theta)^2 - t(\hat{\theta}). \quad (9)$$

With this formulation, the agent's rent depends now only on the screening variables through the average decision  $x(\theta)$  and the spread  $t(\theta)$ . This utility function becomes "quasi-linear" with the spread or "transfer"  $t(\theta)$  measuring the cost for the agent of choosing different decisions along each dimension. Note that  $t(\theta) \geq 0$  so that imposing a spread on the decision has a cost for the agent. Varying that cost with the realization of his private information certainly eases screening. The technical difficulty that we will face in the sequel comes from the fact that this transfer does not enter linearly into the principal's objective.

The average decision  $x(\theta)$  has an impact on the agent's marginal utility which depends on his realized type. It can thus be used as a screening variable as in standard screening models. Clearly,

an agent with type  $\theta$  may be tempted to lie downward to move the average decision closer to his own ideal point. The principal can make that strategy less attractive by increasing the spread between decisions for the lowest types.<sup>10</sup>

As usual in screening problems with quasi-linear utility functions, the incentive compatibility conditions (4) and (5) can be restated in terms of the properties of the pair  $(U(\theta), x(\theta))$ .

**Lemma 2** *The information rent  $U(\theta)$  is absolutely continuous with a first derivative defined almost everywhere and, at any differentiability point:*

$$\dot{U}(\theta) = 2(x(\theta) - \theta). \quad (10)$$

*The average decision  $x(\theta)$  is non-decreasing and thus almost everywhere differentiable with, at any differentiability point:*

$$\dot{x}(\theta) = \frac{\ddot{U}(\theta)}{2} + 1 \geq 0. \quad (11)$$

Note that the non-negativity of the spread implies

$$t(\theta) = -U(\theta) - \frac{\dot{U}^2(\theta)}{4} \geq 0, \quad (12)$$

with an equality only when  $x_1(\theta) = x_2(\theta) = x(\theta)$ , i.e., when both decisions are equal. The fact that (12) will not be an equality with the optimal mechanism reflects the multi-dimensionality aspect of the screening problem. It means that playing on this spread is a useful screening device for the principal.

With the new set of variables, we rewrite the principal's loss in each state of nature  $\theta$  as:

$$\frac{1}{2} \sum_{i=1}^2 (x_i(\theta) - \theta - \delta_i)^2 \equiv L_{\Delta}(U(\theta), \dot{U}(\theta)) = -U(\theta) - \delta \dot{U}(\theta) - \Delta \sqrt{-U(\theta) - \frac{\dot{U}^2(\theta)}{4}} + \delta^2 + \frac{\Delta^2}{4}.$$

From this, we get the following expression of the principal's relaxed problem neglecting for the time being the monotonicity condition on  $x(\theta)$  that will be checked ex post:

$$(\mathcal{P}_{\Delta}) : \quad \min_{U \in W^{1,1}(\Theta)} \int_0^1 L_{\Delta}(U(\theta), \dot{U}(\theta)) d\theta,$$

where  $W^{1,1}(\Theta)$  denotes the set of absolutely continuous arcs on  $\Theta$ . In the parlance of the calculus of variations,  $(\mathcal{P}_{\Delta})$  is actually a Bolza problem with free end-points (see Clarke, 1990, Chapter 4). It is non-standard because the functional  $L_{\Delta}(s, v)$  even though it is continuous and strictly convex

---

<sup>10</sup>The principal-agent literature has stressed that a principal can use the agent's risk-aversion to ease incentives (see, e.g., Arnott and Stiglitz, 1988) by for instance using stochastic mechanisms. Introducing some spread in the agent's decisions in a model with quadratic payoffs has a similar flavor. Subsection 5.1 shows that stochastic mechanisms are suboptimal in our framework.

in  $(s, v)$  is not everywhere differentiable (or even Lipschitz), especially at points where  $-s - \frac{v^2}{4} = 0$  if any such point exists on an admissible curve where  $v(\theta) = \dot{U}(\theta)$  and  $s(\theta) = U(\theta)$ .

We now proceed as follows. First, we prove existence of an optimal arc in  $W^{1,1}(\Theta)$ . Second, we characterize this arc by means of a second-order Euler-Lagrange equation. Third, a first quadrature tells us that such solution solves a first-order differential equation known up to a constant. Finally, we impose conditions on that constant so that the monotonicity condition (11) always holds.

**Lemma 3** *A solution  $U^*(\cdot)$  to  $(\mathcal{P}_\Delta)$  exists.*

Once such solution is known, the pair of decision rules  $\{x_1^*(\cdot), x_2^*(\cdot)\}$  is recovered using the formulae:

$$x_1^*(\theta) = \theta + \frac{\dot{U}^*(\theta)}{2} - \sqrt{-U^*(\theta) - \frac{(\dot{U}^*(\theta))^2}{4}} \text{ and } x_2^*(\theta) = \theta + \frac{\dot{U}^*(\theta)}{2} + \sqrt{-U^*(\theta) - \frac{(\dot{U}^*(\theta))^2}{4}}. \quad (13)$$

**Proposition 2** *An optimal arc  $U^*(\cdot)$  is such that:*

- *The following Euler-Lagrange equation holds at any interior point of differentiability:*

$$\frac{\partial L_\Delta}{\partial U}(U^*(\theta), \dot{U}^*(\theta)) = \frac{d}{d\theta} \left( \frac{\partial L_\Delta}{\partial \dot{U}}(U^*(\theta), \dot{U}^*(\theta)) \right); \quad (14)$$

- *The following free end-points conditions hold on the boundaries of the interval  $[0, 1]$ :*

$$\frac{\partial L_\Delta}{\partial \dot{U}}(U^*(\theta), \dot{U}^*(\theta))|_{\theta=0} = \frac{\partial L_\Delta}{\partial \dot{U}}(U^*(\theta), \dot{U}^*(\theta))|_{\theta=1} = 0; \quad (15)$$

- *$U^*(\theta)$  is continuously differentiable, and thus  $x_1^*(\theta)$  and  $x_2^*(\theta)$  are continuous.*

The next proposition investigates the nature of the solution to the second-order ordinary differential equation (14) by obtaining a first quadrature parameterized by some integration constant  $\lambda \in \mathbb{R}$ . This constant must be non-positive to ensure that the second-order condition (11) holds.

**Proposition 3** *For each solution  $U(\theta, \lambda)$  to (14) which is everywhere negative and satisfies (11), there exists  $\lambda \in \mathbb{R}_-$  such that<sup>11</sup>*

$$\dot{U}(\theta, \lambda) = 2\sqrt{-U(\theta, \lambda) - \Delta^2 \left( \frac{U(\theta, \lambda)}{U(\theta, \lambda) + \lambda} \right)^2}, \quad (16)$$

---

<sup>11</sup>It is important to note that the differential equation (16) may a priori have a singularity and more than one solution going through a given point. This might be the case when, for such solution, there exists  $\theta_0$  such that  $U(\theta_0, \lambda) + \Delta^2 \left( \frac{U(\theta_0, \lambda)}{U(\theta_0, \lambda) + \lambda} \right)^2 = 0$ . Indeed, the right-hand side of (16) fails to be Lipschitz at such a point. It turns out that this possibility does not arise for the optimal mechanism described below because a careful choice of  $\lambda$  ensures that the condition (17) holds everywhere on the optimal path.

and

$$(U(\theta, \lambda) + \lambda)^2 + \Delta^2 U(\theta, \lambda) > 0 \text{ for all } \theta \in \Theta. \quad (17)$$

We are now ready to characterize the optimal mechanism in the multi-dimensional case.

**Theorem 1 (Two-Dimensional Activity.)** *Assume that a single principal controls the two decisions  $x_1$  and  $x_2$  of the agent. When  $\Delta > 0$  the optimal communication mechanism entails the following properties.*

- *Optimal decisions on each dimension are never equal to the agent's ideal points:*

$$x_1^*(\theta) = x^*(\theta) - \frac{\Delta U^*(\theta)}{U^*(\theta) + \lambda^*} < x_2^*(\theta) = x^*(\theta) + \frac{\Delta U^*(\theta)}{U^*(\theta) + \lambda^*}. \quad (18)$$

with  $\lambda^* \in (-\frac{\Delta^2}{4} - \delta^2, -\frac{\Delta^2}{4})$  and  $x^*(\theta) = \theta + \frac{\dot{U}^*(\theta)}{2}$ ;

- *The rent profile  $U^*(\theta)$  is everywhere negative, strictly increasing and solves (16) for  $\lambda^*$ .*
- *There is no pooling area. Monotonicity conditions are satisfied everywhere:  $\dot{x}^*(\theta) > 0$ .*

*Preliminary remarks.* When  $\Delta = 0$ , the optimal mechanism coincides with that described in Proposition 1 for the one-dimensional problem, with

$$U_0^*(\theta) = -(\min\{\theta - 2\delta, 0\})^2, \quad x_0^*(\theta) = \max\{\theta, 2\delta\} \quad \text{and} \quad \lambda^* = 0. \quad (19)$$

Exactly as in Examples 1 and 2, when  $\delta_1 = \delta_2$  there is no gain for the principal to trade off distortions on each dimension because there is no conflict of interest between the principal and the agent concerning their ideal spread. It is therefore costly for the principal to use a spread on decisions.

These remarks being made, we now turn to a more precise analysis of the distortions and provide intuition for our result. That analysis is cast in terms of average distortions and spread that are implemented at the optimal mechanism.

*Average decision.* Beyond the special case where  $\Delta = 0$ , several features of the optimal mechanism are worth being stressed when  $\Delta > 0$ , i.e., when there is a conflict between the principal and the agent on what should be the optimal spread.

First, even when the agent's ideal point is on the lower tail of the distribution, there is no need to offer a pooling contract;  $x^*(\theta)$  is monotonically increasing everywhere. This stands in sharp contrast with the one-dimensional case.

The next corollary shows that the average decision lies systematically in a greater interval than if the principal was restricted to offer the simple mechanism  $\{\tilde{x}_1^*(\theta), \tilde{x}_2^*(\theta)\}_{\theta \in \Theta}$  or the naive mechanism  $\{x_1^O(\theta), x_2^O(\theta)\}_{\theta \in \Theta}$ . It also shows that the benefits of fine-tuning the distortions on decisions allows the principal to move up the average decision further away from the agent's ideal points.

**Corollary 1** For any  $\Delta > 0$ , we have  $[2\delta, 1] \subsetneq [x^*(0), x^*(1)]$ . More precisely, there exists  $\theta^*(\Delta) \in (0, 2\delta)$  such that:

$$x^*(\theta) < 2\delta \text{ if and only if } \theta \leq \theta^*(\Delta).$$

Moreover, we have also:

$$x^*(\theta) > \theta \text{ for all } \theta \in \Theta.$$

These features are illustrated in Figure 1 which compares the average naive decision rule  $\tilde{x}^*(\theta) = \frac{1}{2}(\tilde{x}_1^*(\theta) + \tilde{x}_2^*(\theta)) = \max\{2\delta, \theta\}$  with the optimal average decision rule  $x^*(\theta)$  for a fixed average bias  $\delta$  and different values of  $\Delta$ . The agent's information rent under the optimal mechanism, which is strictly increasing when  $\Delta > 0$ , is represented in Figure 2.

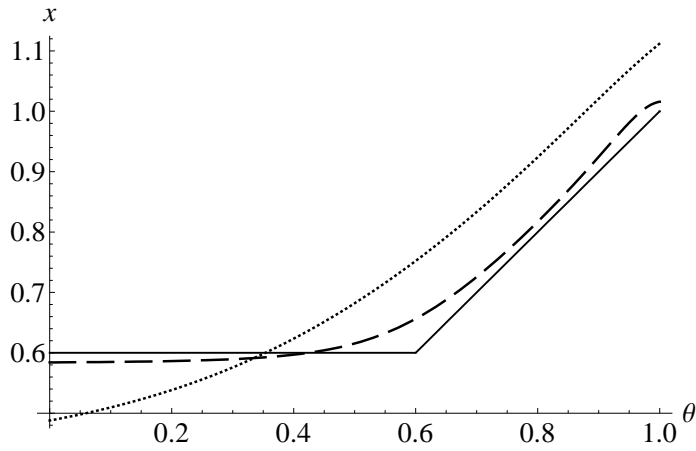


Figure 1: Average decision  $x^*(\theta)$  when  $\delta = 0.3$ , and  $\Delta = 0.6$  (dotted line),  $\Delta = 0.2$  (dashed line) and  $\Delta = 0$  (plain line, which coincides with  $\tilde{x}^*(\theta)$ ).

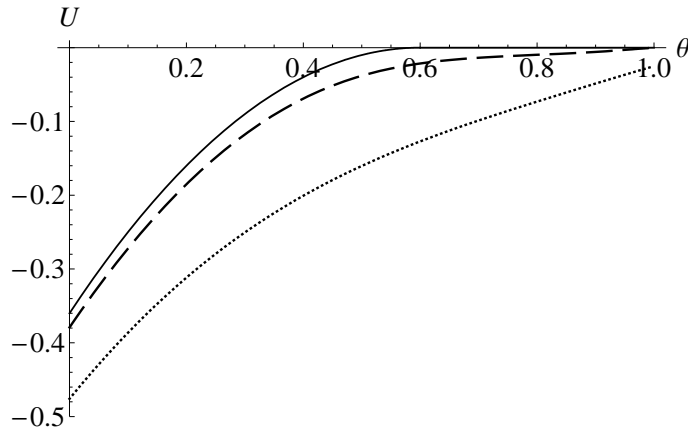


Figure 2: Agent's information rent  $U^*(\theta)$  when  $\delta = 0.3$ , and  $\Delta = 0.6$  (dotted line),  $\Delta = 0.2$  (dashed line), and  $\Delta = 0$  (plain line).

In sharp contrast with the one-dimensional case, the agent's ideal points are never chosen at the optimal mechanism. When the ideal spread of the principal,  $\Delta$ , is strictly positive, the

principal is always able to induce truth-telling without making the agent residual claimant for those decisions.

The distortions on each dimension are indeed quite complex, as illustrated by Figure 3. While  $x_2^*(\theta)$  is always strictly greater than  $\theta$ , it is not always increasing, while  $x_1^*(\theta)$  is strictly increasing over  $[0, 1]$  but not always greater than  $\theta$ . In addition, for the incentive compatibility constraint (5) to be satisfied,  $x_2^*(\theta)$  should be strictly decreasing if and only if  $x_1^*(\theta)$  is larger than  $\theta$ . This feature of the optimal mechanism is general, and is summarized in the next corollary.

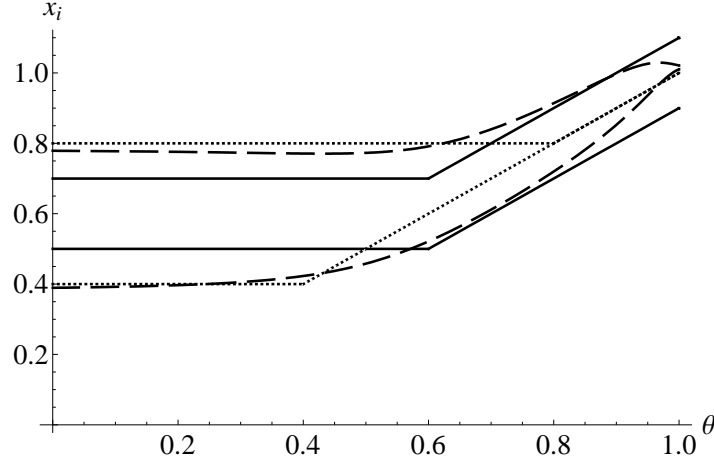


Figure 3: Decisions  $x_i^*(\theta)$  (dashed lines),  $\tilde{x}_i^*(\theta)$  (plain lines) and  $x_i^O(\theta)$  (dotted line), for  $i = 1, 2$ , when  $\delta_1 = 0.2$  and  $\delta_2 = 0.4$ .

**Corollary 2** Assume that  $\Delta > 0$ .

1. For every  $\theta \in [0, 1]$ , we have  $\dot{x}_1^*(\theta) > 0$  and  $x_2^*(\theta) > \theta$ ;
2. For every  $\theta \in [0, 1]$ , we have  $x_1^*(\theta) > \theta$  if and only if  $\dot{x}_2^*(\theta) < 0$ ;
3.  $x_1^*(0) \geq 0$  and  $\dot{x}_2^*(0) \leq 0$  (with strict inequalities when  $\delta_1 > 0$ );
4.  $x_1^*(1) \geq 1$  and  $\dot{x}_2^*(1) \leq 0$  (with strict inequalities when  $\delta_1 > 0$ ).

While the principal distorts the decision on each dimension like in Examples 1 and 2, Figure 3 also illustrates that, contrary to those simple mechanisms, the optimal spread between the two dimensions is not constant; it is actually strictly decreasing in  $\theta$  whenever  $\Delta > 0$ , as shown in the next corollary. This strictly monotonic spread allows the principal to extract information from the agent at a lower cost. The intuition is that the principal can transfer some value from the agent by making activities on both dimensions more disperse when the realized state is lower. In that way, when the agent claims the state is lower (he is biased to do so to get a smaller average decision  $x^*(\theta)$ ) he is punished by having a larger gap between the actions. When  $\Delta = 0$  this spread is not used anymore, and the optimal mechanism coincides with the one dimensional mechanism

because there is no longer any conflict of interest between the principal and the agent concerning the ideal spread (they both want that  $x_2(\theta) = x_1(\theta)$  even if their most preferred value for that decision diverge). It is therefore equally costly for the principal and the agent to introduce a distortion between both dimensions.

Taking a broader perspective, let us think of two different principals  $P_1$  and  $P_2$  as controlling each a single dimension of the agent's activity, say  $x_i$  ( $i = 1, 2$ ), and having objective  $V_i(x_i, \theta) = -\frac{1}{2} \sum_{i=1}^2 (x_i - \theta - \delta_i)^2$ . The intuition of our result becomes then quite obvious. There is no gain in jointly maximizing the sum of the two principals' payoffs when their preferences coincide ( $\delta_1 = \delta_2$ ) since such joint design would only replicate what each of them individually would like to do.

*Optimal spread.* We now turn to a more precise study that uncovers further properties of the optimal spread  $t^*(\theta)$ .

**Corollary 3** *For any  $\Delta > 0$ , the optimal spread  $t^*(\theta) = \frac{1}{4}(x_2^*(\theta) - x_1^*(\theta))^2$  is continuous and strictly decreasing in  $\theta$  with*

$$\Delta^2 > t^*(0) > \frac{\Delta^2}{4} > t^*(1) > 0.$$

A rough intuition could suggest that the optimal spread lies somewhere in between the respective ideal spreads of the principal ( $t_P(\theta) = \frac{\Delta^2}{4}$ ) and the agent ( $t_A(\theta) = 0$ ) to achieve some kind of compromise. This Corollary shows that this is actually not the case. Although the optimal spread is positive which always hurts the agent, it may be significantly beyond the principal's ideal spread for  $\theta$  close enough to zero. For  $\theta$  close enough to one, increasing the spread above the agent's ideal point but still keeping it lower than  $t_P(\theta)$  relaxes the agent's incentive constraint but that moves also goes in the direction of increasing the principal's payoff. However, for  $\theta$  close to zero, the principal "overshoots" and is ready to push the optimal spread beyond his own ideal one.

*Delegation sets.* Even if simple delegation sets trading off inflexible rules and full discretion are no longer optimal, it is still true that a version of the Taxation Principle holds. The principal can implement the optimal communication mechanism by offering an indirect mechanism, i.e., a (continuous) curve in the  $(x_1, x_2)$  space constructed from the parametrization  $\{x_1^*(\theta), x_2^*(\theta)\}_{\theta \in \Theta}$  and letting the agent free to pick any point on this curve. On Figure 4 we have represented in the  $(x_1, x_2)$  space such curve corresponding to the optimal mechanism. At the same time, this figure also features the indirect mechanisms corresponding to the simple mechanism  $\{\tilde{x}_1^*(\theta), \tilde{x}_2^*(\theta)\}_{\theta \in \Theta}$  and the naive mechanism  $\{x_1^O(\theta), x_2^O(\theta)\}_{\theta \in \Theta}$ . Observe that the slope of the optimal indirect mechanism is lower than the slope of the simple indirect mechanism. This is a general feature that can be directly deduced from Corollary 3. Indeed, the slope of the simple curve is one, while the slope of the optimal curve is strictly smaller than one when the spread is strictly decreasing.

More generally, simple duality arguments give us a little bit more information on the shape of those delegation sets. Let define  $T(x)$  as  $T(x) = t^*(\theta)$  for  $x = x^*(\theta)$ .  $T(\cdot)$  is thus the nonlinear

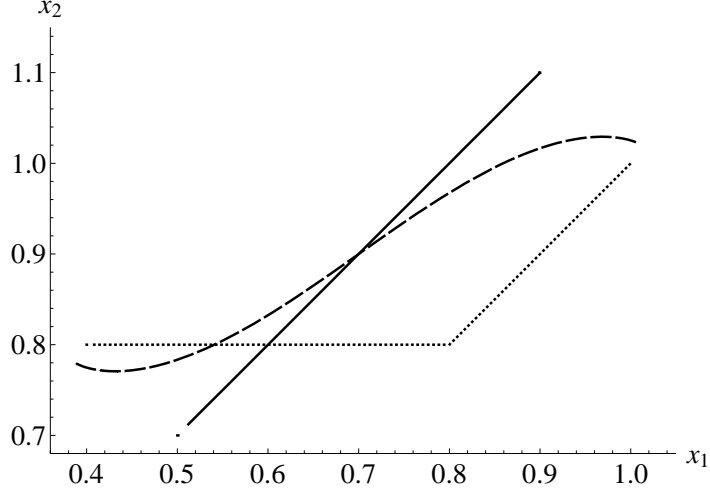


Figure 4: Delegation sets for the optimal mechanism (dashed lines), the simple mechanism  $(\tilde{x}_1^*(\cdot), \tilde{x}_2^*(\cdot))$  (plain lines), and the naive mechanism  $(x_1^O(\cdot), x_2^O(\cdot))$  (dotted lines) when  $\delta_1 = 0.2$  and  $\delta_2 = 0.4$ .

“tax” paid in terms of the spread on decisions when the average decision is  $x$ . By definition of the agent’s optimality conditions, we have:

$$\mathcal{U}^*(\theta) = \max_x 2\theta x - \mathcal{T}(x)$$

where  $\mathcal{U}^*(\theta) = U^*(\theta) + \theta^2$  and  $\mathcal{T}(x) = T(x) + x^2$ .  $\mathcal{U}^*(\theta)$  is convex as a maximum of linear functions. Therefore  $U^*(\theta)$  is the difference of two convex functions.

By duality, we have also:

$$\mathcal{T}(x) = \max_{\theta} 2\theta x - \mathcal{U}^*(\theta).$$

Hence is also  $\mathcal{T}(x)$  is also convex and  $T(x)$  is the difference of two convex functions.

*Simple mechanisms.* Since the design of the optimal mechanism looks rather complex, one may wonder whether simple mechanisms perform well and under which circumstances. The intuition is that, although the optimal mechanism requests full separation of types, it is only marginally so on the lower tail of the type distribution. In this respect, our next proposition shows that the simple mechanism  $\{\tilde{x}_1^*(\theta), \tilde{x}_2^*(\theta)\}_{\theta \in \Theta}$  of Example 2 or the naive mechanism  $\{x_1^O(\theta), x_2^O(\theta)\}_{\theta \in \Theta}$  that simply replicates the one-dimensional mechanisms perform quite well when  $\Delta$  is small enough. When two principals control each dimension of the agent’s activity, this result means that the gain for cooperation between the principals is only significant for a large enough difference between their ideal actions.

**Proposition 4** *The principal’s loss from using the simple mechanism  $\{\tilde{x}_1^*(\theta), \tilde{x}_2^*(\theta)\}_{\theta \in \Theta}$  or the naive mechanism  $\{x_1^O(\theta), x_2^O(\theta)\}_{\theta \in \Theta}$  instead of the optimal mechanism  $\{x_1^*(\theta), x_2^*(\theta)\}_{\theta \in \Theta}$  is of order at most 2 in  $\Delta$ :*

$$\int_0^1 L_{\Delta}(\tilde{U}^*(\theta), \dot{\tilde{U}}^*(\theta)) d\theta - \int_0^1 L_{\Delta}(U^*(\theta), \dot{U}^*(\theta)) d\theta \leq \frac{3\Delta^2}{4}, \quad (20)$$

$$\int_0^1 L_\Delta(U^O(\theta), \dot{U}^O(\theta))d\theta - \int_0^1 L_\Delta(U^*(\theta), \dot{U}^*(\theta))d\theta \leq (1 - \delta)\Delta^2. \quad (21)$$

## 5 Extensions

This section develops some extensions of our basic framework and shows the robustness of some of our results.

### 5.1 Non-Optimality of Stochastic Mechanisms

Kovac and Mylovanov (2009) showed that the restriction to deterministic mechanisms is without loss of generality in the case of quadratic payoffs and a one-dimensional activity. This result clearly extends in our framework when  $\Delta = 0$  *mutatis mutandis*. However, it holds also in our multi-dimensional context when  $\Delta > 0$ .

**Proposition 5** *The optimal deterministic mechanism characterized in Theorem 1 cannot be improved by stochastic mechanisms.*

To relax incentive compatibility, the principal could a priori use random allocations and play on the variance of each decision, i.e., choose how decisions move around their expected values to threaten the agent with some risk in case he reports low values of  $\theta$ . Of course, the principal can still play on top on how decisions are spread as in our analysis of deterministic mechanisms. The second of those strategies has already been shown useful above. The first is suboptimal. The intuition is straightforward, the principal and the agent are both equally averse to such randomizations in allocations and there is no gain from using those that could not already have been achieved by playing on the spread between decisions alone.

### 5.2 Leaving No Discretion Is Generic

Our no-discretion result is highly robust. Indeed, the next proposition shows that the principal never finds optimal to leave full discretion to the agent, leaving him free of choosing his ideal points, on a subset  $I$  with a no-empty interior whatever the everywhere positive and atomless density  $f(\theta)$  on  $\Theta$ .

**Proposition 6** *Assume any everywhere positive and atomless density  $f(\theta)$  on  $\Theta$ . When  $\Delta > 0$ , the optimal deterministic mechanism has never  $x_1^*(\theta) = x_2^*(\theta) = \theta$  on any subset  $I$  with a no-empty interior.*

The intuition is straightforward. Suppose the contrary. The principal could, as in Example 1, move down  $x_1$  and up  $x_2$  by the same small amount on that interval still keeping the same average decision so that incentives for truthtelling are unchanged. Doing so yields a strict benefit to the principal who enjoys having decisions spread apart more than his agent.

## 6 Conclusion

Optimal multi-dimensional communication mechanisms are quite different from the simple delegation sets found in the one-dimensional case. The principal and the agent may now differ not only on their most preferred average decision but also on the distribution of those decisions. The possibility of trading off distortions along each dimension of the agent's activities eases screening and leads to fully separating allocations. Simple delegation sets, which trade off inflexible rules and full discretion, are no longer optimal. The spread on decisions that is necessary to induce cheaper information revelation is a decreasing function of the average decision taken by the agent.

Such extended possibilities for screening offer a strong justification for principals controlling different dimensions of the agent's activities to merge and jointly design contracts. At worst, the analysis of the cooperative contracting design undertaken in this paper characterizes an upper bound on the benefits of merging controls in settings where divided control is often pervasive. Regulation by different agencies and bureaucratic oversights by different Legislative committees are two examples in order. From a theoretical viewpoint, such comparison may depend on fine details of the contracting possibilities available under a non-cooperative design. For instance, the non-cooperative outcome may depend on whether principals observe or not decisions that they do not directly control<sup>12</sup> and whether the agent's messages towards each principal either private or public, the latter case being a priori closer to the cooperative outcome developed in this paper.<sup>13</sup>

It would also be worth investigating optimal mechanisms in more complex environments allowing more general utility functions, more than two decisions and more general type distributions. Some relatively easy extensions should be to investigate optimal mechanisms when the principal and the agent value differently the losses on each dimension still keeping the quadratic structure. We conjecture that the simple decomposition in terms of average decisions and spread would generalize and would still be useful in characterizing optimal mechanisms. More dispersion in decisions is certainly needed when the agent makes decisions that goes counter to what the principal would like on average.

Finally, it would be also interesting to extend our approach by allowing for multi-dimensional preferences, as in the cheap talk framework developed in Battaglini (2002, 2004), but with only one informed player: the agent's bliss points on each dimension of his activity being not necessarily perfectly correlated. This extension is also likely to meet strong technical difficulties, both in the cheap talk and the mechanism design framework, but certainly deserves some attention.<sup>14</sup> For

---

<sup>12</sup>Martimort (2007) coined this situation as a case of *public agency*.

<sup>13</sup>A previous version of this paper (Koessler and Martimort, 2008) analyzed public communication with two principals (see also Goltsman and Pavlov, 2009 for further analysis of the cheap talk setting combining both public and private messages). It was shown that there is no non-cooperative equilibrium with continuous and deterministic action rules. The characterization of the equilibrium mechanisms with public communication remains an interesting open problem.

<sup>14</sup>The literature on multi-dimensional screening has already stressed that pooling allocations are pervasive in non-

all those cases, we conjecture that the decomposition between the average decision and its spread will play a crucial role for contract design.

## Appendix

**Proof of Proposition 1.** See Melumad and Shibano (1991). ■

**Proof of Lemma 1. Necessity:** Incentive compatibility implies for all pairs  $(\theta, \hat{\theta}) \in \Theta^2$ :

$$\sum_{i=1}^2 (x_i(\hat{\theta}) - \theta)^2 \geq \sum_{i=1}^2 (x_i(\theta) - \theta)^2 \text{ and } \sum_{i=1}^2 (x_i(\theta) - \hat{\theta})^2 \geq \sum_{i=1}^2 (x_i(\hat{\theta}) - \hat{\theta})^2. \quad (\text{A.1})$$

Summing those inequalities yields:

$$\sum_{i=1}^2 (x_i(\theta) - x_i(\hat{\theta}))(\theta - \hat{\theta}) \geq 0. \quad (\text{A.2})$$

Hence,  $\sum_{i=1}^2 x_i(\theta)$  is non-decreasing in  $\theta$ . Therefore, it is almost everywhere differentiable with, at any differentiability point, a derivative such that (4) holds. At such a point, an incentive compatible mechanism must also satisfy the first-order condition of the agent's revelation problem, namely (5). Moreover, using (A.1), we get:

$$\sum_{i=1}^2 x_i^2(\theta) - \sum_{i=1}^2 x_i^2(\hat{\theta}) \geq 2\hat{\theta} \left( \sum_{i=1}^2 x_i(\theta) - \sum_{i=1}^2 x_i(\hat{\theta}) \right).$$

Hence,  $\sum_{i=1}^2 x_i^2(\theta)$  is non-decreasing in  $\theta$  when  $\sum_{i=1}^2 x_i(\theta)$  is itself non-decreasing.  $\sum_{i=1}^2 x_i^2(\theta)$  is thus almost everywhere differentiable.

*Sufficiency:* That  $\sum_{i=1}^2 x_i(\theta)$  is non-decreasing in  $\theta$  is then also a sufficient condition for optimality.<sup>15</sup> Indeed, since  $\sum_{i=1}^2 x_i^2(\theta)$  and  $\sum_{i=1}^2 x_i(\theta)$  are both non-decreasing in  $\theta$  and thus almost everywhere differentiable with, at any differentiability point, a derivative which is measurable, Theorem 3 in Royden (1988, p. 100) implies:

$$\begin{aligned} \sum_{i=1}^2 (x_i(\hat{\theta}) - \theta)^2 - \sum_{i=1}^2 (x_i(\theta) - \theta)^2 &\geq \sum_{i=1}^2 \int_{\theta}^{\hat{\theta}} \dot{x}_i(s)(x_i(s) - \theta) ds \\ &= \sum_{i=1}^2 \int_{\theta}^{\hat{\theta}} \dot{x}_i(s)(x_i(s) - s + s - \theta) ds = \sum_{i=1}^2 \int_{\theta}^{\hat{\theta}} \dot{x}_i(s)(s - \theta) ds \geq 0, \end{aligned}$$

where the last equality follows from (5) and the last inequality from (4). ■

**Proof of Lemma 2.** The proof is standard and follows Milgrom and Segal (2002). ■

---

linear pricing environments (Armstrong, 1996; Fang and Norman, 2008; Rochet and Choné, 1998). We conjecture that pooling may be even more pervasive in our setting with no transfer.

<sup>15</sup>Garcia (2005) provides an analysis of the multi-dimensional adverse selection model in a framework with quasi-linear utility functions but focuses a priori on differentiable mechanisms.

**Proof of Lemma 3.** We proceed along the lines of Clarke (1990, Chapter 4). Let us first define the extended-value Lagrangian

$$L_{\Delta}^*(s, v) = \begin{cases} L_{\Delta}(s, v) & \text{if } s \leq -\frac{v^2}{4}, \\ +\infty & \text{otherwise.} \end{cases}$$

As requested in Clarke (1990, p. 167), we observe that:

1.  $L_{\Delta}^*(s, v)$  is  $\mathcal{B}$ -measurable where  $\mathcal{B}$  denotes the  $\sigma$ -algebra of subsets of  $\mathbb{R} \times \mathbb{R}$ ;
2.  $L_{\Delta}^*(s, v)$  is lower-semi continuous;
3.  $L_{\Delta}^*(s, v)$  is convex in  $v$ .

Define now the Hamiltonian as  $H(s, p) = \sup_{v \in \mathbb{R}} \{pv - L_{\Delta}^*(s, v)\}$ . When  $s \leq -\frac{v^2}{4}$ ,  $L_{\Delta}^*(s, v) = L_{\Delta}(s, v)$  is strictly convex in  $v$  and the maximum above is achieved for

$$p = \frac{\partial L_{\Delta}}{\partial v}(s, v). \quad (\text{A.3})$$

This yields the maximand

$$v^* = 4(p + \delta) \sqrt{\frac{-s}{4(p + \delta)^2 + \Delta^2}},$$

which gives

$$H(s, p) = \begin{cases} s + \sqrt{-s(4(p + \delta)^2 + \Delta^2)} - \delta^2 - \frac{\Delta^2}{4} & \text{if } s \leq 0, \\ -\infty & \text{otherwise.} \end{cases}$$

Note that  $H(s, p)$  is differentiable on  $(-\infty, 0] \times \mathbb{R}$ . We get the following inequality:

$$H(s, p) \leq |s| + \Delta \sqrt{|s|} + 2|p + \delta| \sqrt{|s|} - \delta^2 - \frac{\Delta^2}{4}.$$

Using now that  $\sqrt{|s|} \leq 1 + \frac{|s|}{2}$  and that  $|p + \delta| \leq |p| + \delta$ , we obtain finally:

$$H(s, p) \leq \Delta + 2\delta - \delta^2 - \frac{\Delta^2}{4} + 2|p| + |s| \left(1 + \delta + \frac{\Delta}{2} + |p|\right) \leq 2 + 2|p| + |s| \left(1 + \delta + \frac{\Delta}{2} + |p|\right). \quad (\text{A.4})$$

This is a ‘‘growth’’ condition on the Hamiltonian as requested in Clarke (1990, Theorem 4.1.3). ■

**Lemma 4** Clarke (1990). Assume that  $L_{\Delta}^*(\cdot)$  satisfies conditions 1. to 3. above, that  $H(\cdot)$  satisfies the ‘‘growth’’ equation (A.4) and that  $\int_0^1 L_{\Delta}^*(U_0(\theta), \dot{U}_0(\theta)) d\theta$  is finite for at least one admissible arc  $U_0(\theta)$ . Then, problem  $\mathcal{P}_{\Delta}$  has a solution.

It remains to show that  $\int_0^1 L_{\Delta}^*(U_0(\theta), \dot{U}_0(\theta)) d\theta$  is finite for at least one admissible arc  $U_0(\theta)$ . Take  $U_0(\theta) = 0$  which corresponds to decisions  $x_{10}(\theta) = x_{20}(\theta) = \theta$ . This arc does the job and yields  $\int_0^1 L_{\Delta}^*(U_0(\theta), \dot{U}_0(\theta)) d\theta = \int_0^1 L_{\Delta}^*(U_0(\theta), \dot{U}_0(\theta)) d\theta = \delta^2 + \frac{\Delta^2}{4}$ . ■

**Proof of Proposition 2. Preliminaries:** We say that  $H$  satisfies the *strong Lipschitz condition* near an arc  $U$  if there exists  $\epsilon > 0$  and a constant  $k$  such that for all  $p \in \mathbb{R}$  and for all  $(s_1, s_2) \in \mathbb{T}(U, \epsilon)$  the tube of radius  $\epsilon$  centered on the arc  $U$ , the following inequality holds:

$$|H(s_1, p) - H(s_2, p)| \leq k(1 + |p|)|s_1 - s_2|. \quad (\text{A.5})$$

This property holds in our context when there exists  $\eta > 0$  such that  $U(\theta) < -\eta$  for all  $\theta$  (i.e.,  $U(\theta)$  is bounded away from zero which will be the case for the solution we exhibit below). Indeed, we have over the relevant range where  $s_i \leq 0$ :

$$|H(s_1, p) - H(s_2, p)| = |s_1 - s_2 + (\sqrt{-s_1} - \sqrt{-s_2})\sqrt{\Delta^2 + 4(p + \delta)^2}|.$$

Note that  $|\sqrt{-s_1} - \sqrt{-s_2}| = \frac{|s_1 - s_2|}{2\sqrt{-s_0}}$  for some  $s_0 \in \mathbb{T}(U, \epsilon)$  from the Mean-Value Theorem. Therefore,  $|\sqrt{-s_1} - \sqrt{-s_2}| \leq \frac{|s_1 - s_2|}{2\sqrt{\eta - \epsilon}}$  for  $\epsilon$  small enough. Hence, we get:

$$\begin{aligned} |H(s_1, p) - H(s_2, p)| &\leq |s_1 - s_2| \left( 1 + \frac{\sqrt{\Delta^2 + 4(p + \delta)^2}}{2\sqrt{\eta - \epsilon}} \right) \leq |s_1 - s_2| \left( 1 + \frac{\Delta + 2(\delta + |p|)}{2\sqrt{\eta - \epsilon}} \right) \\ &\leq \max \left\{ 1 + \frac{\Delta + 2\delta}{2\sqrt{\eta - \epsilon}}, \frac{1}{\sqrt{\eta - \epsilon}} \right\} |s_1 - s_2|(1 + |p|), \end{aligned}$$

which is (A.5) with  $k = \max \left\{ 1 + \frac{\Delta + 2\delta}{2\sqrt{\eta - \epsilon}}, \frac{1}{\sqrt{\eta - \epsilon}} \right\}$ .

*Euler equation and boundaries conditions:* From Clarke (1990, Theorem 4.2.2, p.169), and since  $L_\Delta^*(\cdot)$  satisfies conditions 1., 2., and 3. above and  $H(\cdot)$  satisfies the strong Lipschitz condition (A.5), there exists an absolutely continuous arc  $p(\cdot)$  such that the following conditions hold for the optimal arc  $U^*(\theta)$ .

- Optimality conditions for the Hamiltonian  $H(\cdot)$ :

$$-\dot{p}(\theta) = \frac{\partial H}{\partial s}(U^*(\theta), p(\theta)), \quad (\text{A.6})$$

$$\dot{U}^*(\theta) = \frac{\partial H}{\partial p}(U^*(\theta), p(\theta)). \quad (\text{A.7})$$

- Boundary conditions:

$$p(0) = p(1) = 0. \quad (\text{A.8})$$

Using (A.3) yields  $p(\theta) = \frac{\partial L_\Delta}{\partial U}(U^*(\theta), \dot{U}^*(\theta))$ . Differentiating with respect to  $\theta$ , inserting into (A.6) and observing that  $\frac{\partial H}{\partial s}(U^*(\theta), p(\theta)) = -\frac{\partial L_\Delta}{\partial U}(U^*(\theta), \dot{U}^*(\theta))$  yields (14). Finally, using again  $p(\theta) = \frac{\partial L_\Delta}{\partial U}(U^*(\theta), \dot{U}^*(\theta))$  yields (15).

*Continuity:* First observe that, a.e. on  $\Theta$ , we have by definition

$$H(U^*(\theta), p(\theta)) = p(\theta)\dot{U}^*(\theta) - L_\Delta(U^*(\theta), \dot{U}^*(\theta)) \geq p(\theta)v - L_\Delta(U^*(\theta), v), \quad \forall v \leq 2\sqrt{-U^*(\theta)}. \quad (\text{A.9})$$

If  $\dot{U}$  is not continuous at some  $\theta_0 \in (0, 1)$ , there exists an increasing sequence  $\theta_n^-$  and a decreasing sequence  $\theta_n^+$  ( $n \geq 1$ ) both converging towards  $\theta_0$ , such that (A.9) applies at  $\theta_n^-$ ,  $\theta_0$  and  $\theta_n^+$ , and (using monotonicity to get the strict inequality):

$$\lim_{n \rightarrow +\infty} \dot{U}^*(\theta_n^-) = \dot{U}^*(\theta_0^-) < \dot{U}^*(\theta_0^+) = \lim_{n \rightarrow +\infty} \dot{U}^*(\theta_n^+).$$

Because  $L_\Delta(s, v)$  is continuous in  $(s, v)$  and  $U^*(\theta)$  is absolutely continuous and thus continuous at  $\theta_0$ , we have:

$$L_\Delta(U^*(\theta_0), v) = \lim_{n \rightarrow +\infty} L_\Delta(U^*(\theta_n^-), v) \text{ and } L_\Delta(U^*(\theta_0), \dot{U}^*(\theta_0^-)) = \lim_{n \rightarrow +\infty} L_\Delta(U^*(\theta_n^-), \dot{U}^*(\theta_n^-)). \quad (\text{A.10})$$

Taking  $\theta = \theta_n^-$  into (A.9) and passing to the limit, using the continuity of  $p(\theta)$ , yields

$$p(\theta_0)\dot{U}^*(\theta_0^-) - L_\Delta(U^*(\theta_0), \dot{U}^*(\theta_0^-)) \geq p(\theta_0)v - L_\Delta(U^*(\theta_0), v) \quad \forall v \leq 2\sqrt{-U^*(\theta)}.$$

Using similar arguments with the sequence  $\theta_n^+$ , we also get

$$p(\theta_0)\dot{U}^*(\theta_0^+) - L_\Delta(U^*(\theta_0), \dot{U}^*(\theta_0^+)) \geq p(\theta_0)v - L_\Delta(U^*(\theta_0), v) \quad \forall v \leq 2\sqrt{-U^*(\theta)}.$$

Hence, the function  $v \rightarrow p(\theta_0)v - L_\Delta(U^*(\theta_0), v)$  defined for  $v \leq 2\sqrt{-U^*(\theta_0)}$  achieves its maxima at both  $\dot{U}^*(\theta_0^+)$  and  $\dot{U}^*(\theta_0^-)$ . Since it is strictly concave, we get  $\dot{U}^*(\theta_0^+) = \dot{U}^*(\theta_0^-)$ . From this contradiction, we conclude that any arbitrary  $\theta \in \Theta$  is contained in a relatively open interval on which  $\dot{U}^*$  is almost everywhere equal to a continuous function.  $\dot{U}^*$  and thus  $x_1^*$  and  $x_2^*$  are continuous on  $\Theta$ .  $\blacksquare$

**Proof of Proposition 3.** Since the functional  $L_\Delta(\cdot)$  does not depend on  $\theta$ , we can obtain a first quadrature of (14) on any interval where  $U(\theta) + \frac{\dot{U}^2(\theta)}{4} < 0$  as:

$$L_\Delta(U(\theta, \lambda), \dot{U}(\theta, \lambda)) - \dot{U}(\theta, \lambda) \frac{\partial L_\Delta}{\partial \dot{U}}(U(\theta, \lambda), \dot{U}(\theta, \lambda)) = \lambda + \delta^2 + \frac{\Delta^2}{4}, \quad (\text{A.11})$$

where a priori  $\lambda \in \mathbb{R}$  and where we make explicit the dependence of the solution on this parameter. We obtain immediately:

$$U(\theta, \lambda) + \delta \dot{U}(\theta, \lambda) + \Delta \sqrt{-U(\theta, \lambda) - \frac{\dot{U}^2(\theta, \lambda)}{4}} - \dot{U}(\theta, \lambda) \left( \delta - \frac{\Delta \dot{U}(\theta, \lambda)}{4 \sqrt{-U(\theta, \lambda) - \frac{\dot{U}^2(\theta, \lambda)}{4}}} \right) = -\lambda.$$

Simplifying yields:

$$U(\theta, \lambda) \left( 1 - \frac{\Delta}{\sqrt{-U(\theta, \lambda) - \frac{\dot{U}^2(\theta, \lambda)}{4}}} \right) = -\lambda.$$

Solving for  $\dot{U}(\theta, \lambda)$  yields

$$\dot{U}^2(\theta, \lambda) = -4 \left( U(\theta, \lambda) + \Delta^2 \left( \frac{U(\theta, \lambda)}{U(\theta, \lambda) + \lambda} \right)^2 \right), \quad (\text{A.12})$$

which requires  $-\Delta^2 \left( \frac{U(\theta, \lambda)}{U(\theta, \lambda) + \lambda} \right)^2 \geq U(\theta, \lambda)$  or  $(U(\theta, \lambda) + \lambda)^2 + \Delta^2 U(\theta, \lambda) \geq 0$  given that  $U(\theta, \lambda) \leq 0$  since by definition the agent's information rent is negative. Solving the second-order equation (A.12) and keeping the positive root only,<sup>16</sup> we get (16). When  $\dot{U}(\theta, \lambda) > 0$ , differentiating (16) with respect to  $\theta$  yields

$$\ddot{U}(\theta, \lambda) + \frac{\dot{U}(\theta, \lambda) \left( 1 + 2\lambda\Delta^2 \frac{U(\theta, \lambda)}{(U(\theta, \lambda) + \lambda)^3} \right)}{\sqrt{-U(\theta, \lambda) - \Delta^2 \left( \frac{U(\theta, \lambda)}{U(\theta, \lambda) + \lambda} \right)^2}} = \ddot{U}(\theta, \lambda) + 2 \left( 1 + 2\lambda\Delta^2 \frac{U(\theta, \lambda)}{(U(\theta, \lambda) + \lambda)^3} \right) = 0 \quad (\text{A.13})$$

Hence, on any interval where  $\dot{U}(\theta, \lambda) > 0$ , the second-order condition (11) can be written as

$$0 \leq \ddot{U}(\theta, \lambda) + 2 = -4\lambda\Delta^2 \frac{U(\theta, \lambda)}{(U(\theta, \lambda) + \lambda)^3}. \quad (\text{A.14})$$

Since  $U(\theta, \lambda) \leq 0$  holds,  $\lambda \leq 0$  implies also  $U(\theta, \lambda) + \lambda \leq 0$  and then (A.14) holds. This imposes the requested restriction on the admissible solutions to (16). Finally, note that the second-order condition (11) holds obviously on any interval where instead  $\dot{U}(\theta, \lambda) = 0$ . ■

**Proof of Theorem 1.** The structure of the proof is as follows. First, we derive from the necessary free end-point conditions (15) some properties of the boundary values of  $U^*$  that are used to find  $\lambda^*$ . Sufficiency follows.

*Necessity:* Define the function  $P(x) = \frac{-x((x+\lambda)^2 + \Delta^2 x)}{(x+\lambda)^2}$ . For  $x < 0$ ,  $P(x) > 0$  if and only if the second degree polynomial  $(x + \lambda)^2 + \Delta^2 x$  is everywhere positive. This is so when  $\lambda < -\frac{\Delta^2}{4}$ . When that condition holds, the differential equation (16) is Lipschitz at any point where  $U(\theta, \lambda) < 0$  and thus it has a single solution at any such point. Moreover, a solution  $U(\theta, \lambda)$  is then everywhere increasing on the whole domain where  $U(\theta, \lambda) < 0$ . As a result, the differential equation (16) is everywhere Lipschitz when  $U(1, \lambda) < 0$  which turns out to be the case for the path we derive below.

The necessary free end-points conditions (15) can be rewritten for an optimal path as:

$$\left( -\delta + \Delta \frac{\dot{U}^*(\theta)}{4\sqrt{-U^*(\theta) - \frac{(\dot{U}^*(\theta))^2}{4}}} \right) |_{\theta=0,1} = 0.$$

Using (16) to express  $\dot{U}^*(\theta)$ , those conditions can be simplified so that  $U^*(0)$  and  $U^*(1)$  solve indeed the following second-order equation in  $U$ :

$$(U + \lambda^*)^2 = -(\Delta^2 + 4\delta^2)U, \quad (\text{A.15})$$

where  $\lambda^*$  is the value of  $\lambda$  for the optimal arc  $U^*$ . Assuming now that  $\lambda^* > -\frac{\Delta^2}{4} - \delta^2$  (a condition checked below), (A.15) admits two solutions respectively given by

$$U^*(0) = -\lambda^* - \frac{1}{2} \left( \Delta^2 + 4\delta^2 + \sqrt{(\Delta^2 + 4\delta^2)^2 + 4\lambda^*(\Delta^2 + 4\delta^2)} \right), \quad (\text{A.16})$$

<sup>16</sup>Since it corresponds to an average decision  $x(\theta)$  biased towards the principal, namely  $x(\theta) \geq \theta$  (see equation (10)).

$$U^*(1) = -\lambda^* - \frac{1}{2} \left( \Delta^2 + 4\delta^2 - \sqrt{(\Delta^2 + 4\delta^2)^2 + 4\lambda^*(\Delta^2 + 4\delta^2)} \right). \quad (\text{A.17})$$

Note in particular that (A.15) implies that both  $U^*(0)$  and  $U^*(1)$  are negative.

The last step is to show that there exists  $\lambda^* \in (-\frac{\Delta^2}{4} - \delta^2, -\frac{\Delta^2}{4})$  such that the corresponding path  $U^*(\theta) = U(\theta, \lambda^*)$  solving (16) and starting from

$$U(0, \lambda) = -\lambda - \frac{1}{2} \left( \Delta^2 + 4\delta^2 + \sqrt{(\Delta^2 + 4\delta^2)^2 + 4\lambda(\Delta^2 + 4\delta^2)} \right),$$

reaches

$$U(1, \lambda) = -\lambda - \frac{1}{2} \left( \Delta^2 + 4\delta^2 - \sqrt{(\Delta^2 + 4\delta^2)^2 + 4\lambda(\Delta^2 + 4\delta^2)} \right).$$

This requires to find a solution  $\lambda^*$  to the equation  $\varphi(\lambda) = \psi(\lambda)$ , with

$$\varphi(\lambda) = U(1, \lambda) - U(0, \lambda) = \sqrt{(\Delta^2 + 4\delta^2)^2 + 4\lambda(\Delta^2 + 4\delta^2)},$$

and

$$\psi(\lambda) = \int_0^1 \dot{U}(\theta, \lambda) d\theta = \int_0^1 2\sqrt{-U(\theta, \lambda) - \Delta^2} \left( \frac{U(\theta, \lambda)}{U(\theta, \lambda) + \lambda} \right)^2 d\theta,$$

where the path  $U(\theta, \lambda)$  starts from the initial condition  $U(0, \lambda)$ . Note that both  $\varphi(\cdot)$  and  $\psi(\cdot)$  are continuous in  $\lambda$ . It is clear that  $\varphi(\lambda)$  is strictly increasing in  $\lambda$  with, for  $\lambda_1 = -\frac{\Delta^2}{4} - \delta^2$  and  $\lambda_2 = -\frac{\Delta^2}{4}$ ,

$$\varphi(\lambda_1) = 0 < 2\delta\sqrt{\Delta^2 + 4\delta^2} = \varphi(\lambda_2).$$

On the other hand, note that

$$\psi(\lambda_1) > 0 = \varphi(\lambda_1), \quad (\text{A.18})$$

since the path  $U(\theta, \lambda_1)$  starting from  $U(0, \lambda_1)$  is strictly increasing. Moreover, for  $\lambda_2$ , (16) can be rewritten as:

$$\dot{U}(\theta, \lambda_2) = 2\sqrt{-U(\theta, \lambda_2)} \frac{|U(\theta, \lambda_2) - \lambda_2|}{|U(\theta, \lambda_2) + \lambda_2|}. \quad (\text{A.19})$$

The path solving (A.19) and starting at  $U(0, \lambda_2)$  (note that  $U(0, \lambda_2) < \lambda_2 < U(1, \lambda_2)$ ) is strictly increasing everywhere and cannot cross the boundary  $U = \lambda_2$  because the only solution to (A.19) such that  $U(\theta_1) = \lambda_2$  for a given  $\theta_1 > 0$  is such that  $U(\theta) = \lambda_2$  for all  $\theta$  since the right-hand side of (A.19) satisfies a Lipschitz condition at any point  $U(\theta, \lambda_2)$  away from zero; a contradiction with  $U(0, \lambda_2) < \lambda_2$ . From that, we deduce  $U(\theta, \lambda_2) < \lambda_2$  for all  $\theta$ . Hence, the following sequence of inequalities holds:

$$\psi(\lambda_2) = \int_0^1 \dot{U}(\theta, \lambda_2) d\theta < \lambda_2 - U(0, \lambda_2) = \lambda_2 - U(1, \lambda_2) + \varphi(\lambda_2).$$

Finally, we get:

$$\psi(\lambda_2) < \varphi(\lambda_2). \quad (\text{A.20})$$

Gathering Equations (A.18) and (A.20) yields the existence of  $\lambda^* \in (\lambda_1, \lambda_2)$  such that  $\varphi(\lambda^*) = \psi(\lambda^*)$ . For such  $\lambda^* \in (-\frac{\Delta^2}{4} - \delta^2, -\frac{\Delta^2}{4})$  we have  $U(1, \lambda^*) < 0$  and thus  $U^*(\theta) = U(\theta, \lambda^*) < 0$  for all

$\theta$ . From Proposition 3 this implies  $\dot{U}^*(\theta) > 0$  for all  $\theta$ . Finally, using  $2\dot{x}^*(\theta) = \dot{U}^*(\theta) + 2$  and (A.14) with  $U^*(\theta) < 0$  and  $\lambda^* < 0$  we get  $\dot{x}^*(\theta) > 0$  for all  $\theta$ .

*Sufficiency:* Sufficiency follows from Clarke (1990, Chapter 4, Corollary p. 179) when noticing that  $L_\Delta(\cdot)$  satisfies the convexity assumption and the function  $s \rightarrow H(s, p(\theta))$  is concave in  $s$ . ■

**Proof of Corollary 1.** First observe that

$$x^*(0) = \frac{\dot{U}^*(0)}{2} = \frac{\sqrt{-U^*(0)((U^*(0) + \lambda^*)^2 + \Delta^2 U^*(0))}}{|U^*(0) + \lambda^*|}.$$

Using (A.16), we get:

$$x^*(0) = 2\delta \frac{|U^*(0)|}{|U^*(0) + \lambda^*|} < 2\delta.$$

Finally, we obviously have  $x^*(\theta) - \theta = \frac{\dot{U}^*(\theta)}{2} > 0$  for all  $\theta$ . ■

**Proof of Corollary 2.** The first property follows from (18). Now, from the incentive constraint (5) and the first property of the corollary we get the second property. Next, using (18) we have  $x_1^*(0) \geq 0$  if and only if

$$\frac{\dot{U}^*(0)}{2} \geq \Delta \left( \frac{U^*(0)}{U^*(0) + \lambda^*} \right).$$

Using (16) and simplifying we get  $2\Delta^2 U^*(0) \geq -(U^*(0) + \lambda^*)^2$ , i.e., by (A.15),  $2\Delta^2 \leq \Delta^2 + 4\delta^2$ , which is always satisfied (with a strict inequality when  $\delta_1 > 0$ ).  $\dot{x}_2^*(0) \leq 0$  follows now from the second property of the corollary. The last property is proved similarly. ■

**Proof of Corollary 3.** From Theorem 1, Equation (18), we have:

$$t^*(\theta) = \Delta^2 \left( \frac{U^*(\theta)}{U^*(\theta) + \lambda^*} \right)^2.$$

Therefore, we get:

$$\dot{t}^*(\theta) = 2\Delta^2 \lambda \dot{U}^*(\theta) \frac{U^*(\theta)}{(U^*(\theta) + \lambda^*)^3} < 0$$

since  $\lambda^* < 0$  and  $\dot{U}^*(\theta) < 0$  for all  $\theta$ . Using (A.16), we get

$$t^*(0) = \Delta^2 \left( \frac{U^*(0)}{U^*(0) + \lambda^*} \right)^2 < \Delta^2$$

$$\Leftrightarrow 1 > \frac{U^*(0)}{U^*(0) + \lambda^*} = 1 + \frac{2\lambda^*}{\Delta^2 + 4\delta^2 + \sqrt{(\Delta^2 + 4\delta^2)^2 + 4\lambda^*(\Delta^2 + 4\delta^2)}},$$

which holds since  $\lambda^* < 0$ .

Still using (A.16), we also get:

$$t^*(0) > \frac{\Delta^2}{4} \Leftrightarrow \frac{1}{2} < \frac{U^*(0)}{U^*(0) + \lambda^*} = 1 + \frac{2\lambda^*}{\Delta^2 + 4\delta^2 + \sqrt{(\Delta^2 + 4\delta^2)^2 + 4\lambda^*(\Delta^2 + 4\delta^2)}} \quad (\text{A.21})$$

$$\Leftrightarrow \Delta^2 + 4\delta^2 + 4\lambda^* + \sqrt{(\Delta^2 + 4\delta^2)^2 + 4\lambda^*(\Delta^2 + 4\delta^2)} > 0,$$

which holds since  $\lambda^* > -\frac{\Delta^2}{4} - \delta^2$ .

Now using (A.17), we get:

$$\begin{aligned} t^*(1) &= \Delta^2 \left( \frac{U^*(1)}{U^*(1) + \lambda^*} \right)^2 < \frac{\Delta^2}{4} \tag{A.22} \\ \Leftrightarrow \frac{1}{2} &> \frac{U^*(1)}{U^*(1) + \lambda^*} = 1 + \frac{2\lambda^*}{\Delta^2 + 4\delta^2 - \sqrt{(\Delta^2 + 4\delta^2)^2 + 4\lambda^*(\Delta^2 + 4\delta^2)}} \\ &\Leftrightarrow \Delta^2 + 4\delta^2 + 4\lambda^* < \sqrt{(\Delta^2 + 4\delta^2)^2 + 4\lambda^*(\Delta^2 + 4\delta^2)} \end{aligned}$$

which holds again since  $\lambda^* > -\frac{\Delta^2}{4} - \delta^2$  and  $\lambda^* < 0$ . ■

**Proof of Proposition 4.** We have

$$\begin{aligned} \int_0^1 L_\Delta(U^*(\theta), \dot{U}^*(\theta)) d\theta &\geq \int_0^1 \left( L_0(U^*(\theta), \dot{U}^*(\theta)) - \Delta \sqrt{-U^*(\theta) - \frac{(\dot{U}^*(\theta))^2}{4} + \frac{\Delta^2}{4}} \right) d\theta \\ &\geq \int_0^1 \left( L_0(U_0^*(\theta), \dot{U}_0^*(\theta)) - \Delta \sqrt{-U^*(\theta) - \frac{(\dot{U}^*(\theta))^2}{4} + \frac{\Delta^2}{4}} \right) d\theta \\ &= \delta^2 + \frac{\Delta^2}{4} - \frac{4\delta^3}{3} - \Delta^2 \int_0^1 \frac{U^*(\theta)}{U^*(\theta) + \lambda^*} d\theta, \end{aligned}$$

where  $U_0^*(\theta)$  is given in (19) and the last inequality follows from (16). This implies:

$$\int_0^1 L_\Delta(U^*(\theta), \dot{U}^*(\theta)) d\theta \geq \delta^2 - \frac{3\Delta^2}{4} - \frac{4\delta^3}{3}.$$

We also have

$$\int_0^1 L_\Delta(\tilde{U}^*(\theta), \dot{\tilde{U}}^*(\theta)) d\theta = \delta^2 - \frac{4\delta^3}{3},$$

and

$$\int_0^1 L_\Delta(U^O(\theta), \dot{U}^O(\theta)) d\theta = \delta^2 + \frac{\Delta^2}{4} - \frac{2\delta_1^3}{3} - \frac{2\delta_2^3}{3},$$

which gives the required inequalities. ■

**Proof of Proposition 5.** A stochastic direct mechanism is a mapping  $\mu(\cdot|\cdot) : \Theta \rightarrow \Delta(\mathcal{K} \times \mathcal{K})$  where  $\Delta(\mathcal{K} \times \mathcal{K})$  is the set of measures on  $\mathcal{K} \times \mathcal{K}$ . For further references, we define the mean and variance of such stochastic mechanism as

$$\bar{x}_i(\hat{\theta}) = \int_{\mathcal{K} \times \mathcal{K}} x_i d\mu(x_1, x_2|\hat{\theta}) \text{ and } \sigma_i^2(\hat{\theta}) = \int_{\mathcal{K} \times \mathcal{K}} (x_i - \bar{x}_i(\hat{\theta}))^2 d\mu(x_1, x_2|\hat{\theta}) \geq 0.$$

Boundedness of  $\mathcal{K}$  ensures that such moments exist. Note that deterministic mechanisms are such that  $\sigma_i^2(\hat{\theta}) \equiv 0$ . For further references also, denote  $\bar{x}(\hat{\theta}) = \frac{1}{2} \sum_{i=1}^2 \bar{x}_i(\hat{\theta})$  the average decision and

$\bar{y}(\hat{\theta}) = \bar{x}_2(\hat{\theta}) - \bar{x}_1(\hat{\theta})$  the spread of those average decisions. In this context, incentive compatibility can be written as:

$$U(\theta) = \max_{\hat{\theta} \in \Theta} \int_{\mathcal{K} \times \mathcal{K}} \left( \sum_{i=1}^2 -\frac{1}{2}(x_i - \theta)^2 \right) d\mu(x_1, x_2 | \hat{\theta}).$$

Taking expectations, we get:

$$U(\theta) = \max_{\hat{\theta} \in \Theta} -\frac{1}{2} \left( \sum_{i=1}^2 (\bar{x}_i(\hat{\theta}) - \theta)^2 \right) - z(\hat{\theta}) = \max_{\hat{\theta} \in \Theta} -(\bar{x}(\hat{\theta}) - \theta)^2 - \frac{\bar{y}^2(\hat{\theta})}{4} - z(\hat{\theta}),$$

where  $z(\hat{\theta}) = \frac{1}{2} \sum_{i=1}^2 \sigma_i^2(\hat{\theta}) \geq 0$ . From this, it immediately follows that  $U(\cdot)$  is absolutely continuous with a derivative defined almost everywhere defined as

$$\dot{U}(\theta) = 2(\bar{x}(\hat{\theta}) - \theta), \quad (\text{A.23})$$

with

$$z(\theta) \leq -U(\theta) - \frac{\dot{U}^2(\theta)}{4} - \frac{\bar{y}^2(\theta)}{4}. \quad (\text{A.24})$$

Similarly, the expected payoff of the principal with such a stochastic mechanism can be written as:

$$\int_0^1 \left( \int_{\mathcal{K} \times \mathcal{K}} \left( \sum_{i=1}^2 -\frac{1}{2}(x_i - \theta - \delta_i)^2 \right) d\mu(x_1, x_2 | \theta) \right) d\theta = \int_0^1 \left( (U(\theta) + \delta \dot{U}(\theta) + \frac{\Delta}{2} \bar{y}(\theta) - \delta^2 - \frac{\Delta^2}{4}) \right) d\theta.$$

The principal problem when stochastic mechanisms are allowed can be written as:

$$(\mathcal{P}_\Delta^s) : \min_{\{U \in W^{1,1}(\Theta), z \geq 0\}} \int_0^1 L_\Delta^s(U(\theta), \dot{U}(\theta), z(\theta)) d\theta,$$

where

$$L_\Delta^s(U(\theta), \dot{U}(\theta), z(\theta)) = -U(\theta) - \delta \dot{U}(\theta) - \Delta \sqrt{-U(\theta) - \frac{\dot{U}^2(\theta)}{4} - z(\theta) + \delta^2 + \frac{\Delta^2}{4}}.$$

Clearly, the pointwise solution to this problem when  $\Delta > 0$  is achieved for  $z(\theta) = 0$ , i.e., for deterministic mechanisms. ■

**Proof of Proposition 6.** Suppose that the optimal solution  $U^*$  is such that  $U^*(\theta) = 0$  on an interval  $I$  with non-empty interior, i.e.,  $x_i^*(\theta) = \theta$  on that interval. Consider now the new utility profile  $\tilde{U}$  obtained by leaving the decisions  $x_i^*(\theta)$  unchanged on  $I^c$  but choosing  $\tilde{x}_1(\theta) = \theta - \epsilon$  and  $\tilde{x}_2(\theta) = \theta + \epsilon$  on  $I$ . Note that  $\dot{\tilde{U}}(\theta) = \dot{U}^*(\theta)$  both on  $I$  and  $I^c$ . Observe that

$$\begin{aligned} & \int_0^1 L_\Delta(\tilde{U}(\theta), \dot{\tilde{U}}(\theta)) f(\theta) d\theta - \int_0^1 L_\Delta(U^*(\theta), \dot{U}^*(\theta)) f(\theta) d\theta \\ &= \int_I (L_\Delta(\tilde{U}(\theta), \dot{\tilde{U}}(\theta)) - L_\Delta(U^*(\theta), \dot{U}^*(\theta))) f(\theta) d\theta = -(\epsilon \Delta - \epsilon^2) \int_I f(\theta) d\theta < 0, \end{aligned}$$

for  $\epsilon$  small enough, a contradiction with the optimality of  $U^*$ . ■

## References

- ALONSO, N. AND R. MATOUSCHEK (2007): "Relational Delegation," *Rand Journal of Economics*, 38, 70–89.
- (2008): "Optimal Delegation," *Review of Economic Studies*, 75, 259–293.
- ALONSO, R., W. DESSEIN, AND N. MATOUSCHEK (2008): "When Does Coordination Require Centralization?" *American Economic Review*, 98, 145–179.
- AMBRUS, A. AND G. EGOROV (2010): "Delegation and Nonmonetary Incentives," Tech. rep., Working Paper, Harvard University.
- AMBRUS, A. AND S. TAKAHASHI (2008): "Multi-Sender Cheap Talk with Restricted State Space," *Theoretical Economics*, 3, 1–27.
- ARMSTRONG, M. (1994): "Delegation and Discretion," Discussion Papers in Economics and Econometrics 9421, University of Southampton.
- (1996): "Multiproduct Nonlinear Pricing," *Econometrica*, 64, 51–75.
- ARNOTT, R. AND J. STIGLITZ (1988): "Randomization with Asymmetric Information," *Rand Journal of Economics*, 19, 344–362.
- AUSTEN-SMITH, D. (1993): "Interested Experts and Policy Advice: Multiple Referrals under Open Rule," *Games and Economic Behavior*, 5, 3–44.
- BARON, D. (2000): "Legislative Organization with Informational Committees," *American Journal of Political Sciences*, 44, 485–505.
- BATTAGLINI, M. (2002): "Multiple Referrals and Multidimensional Cheap Talk," *Econometrica*, 70, 1379–1401.
- (2004): "Policy Advice with Imperfectly Informed Experts," *Advances in Theoretical Economics*, 4.
- CLARKE, F. H. (1990): *Optimization and Nonsmooth Analysis*, SIAM. Philadelphia.
- EPSTEIN, D. AND S. O'HALLORAN (1999): *Delegating Powers*, Cambridge University Press.
- FANG, H. AND P. NORMAN (2008): "Optimal Provision of Multiple Excludable Public Goods," .
- FARRELL, J. AND R. GIBBONS (1989): "Cheap Talk with Two Audiences," *American Economic Review*, 79, 1214–1223.
- GARCIA, D. (2005): "Monotonicity in Direct Revelation Mechanisms," *Economics Letters*, 88, 21–26.

- GOLTSMAN, M., J. HÖRNER, G. PAVLOV, AND F. SQUINTANI (2009): "Mediation, Arbitration and Negotiation," *Journal of Economic Theory*, 144, 1397–1420.
- GOLTSMAN, M. AND G. PAVLOV (2009): "How to Talk to Multiple Audiences," mimeo.
- HARRIS, M. AND A. RAVIV (1996): "The Capital Budgeting Process, Incentives and Information," *Journal of Finance*, 51, 1139–1174.
- HOLMSTRÖM, B. (1984): "On the Theory of Delegation," in *Bayesian Models in Economic Theory*, ed. by M. Boyer and R. Kihlstrom, Elsevier Science B. V.
- HUBER, J. AND C. SHIPAN (2002): *Deliberative Discretion: The Institutional Foundations of Bureaucratic Autonomy*, Cambridge University Press.
- KOESSLER, F. AND D. MARTIMORT (2008): "Multidimensional Communication Mechanisms: Cooperative and Conflicting Designs," PSE Working Paper 2008-07.
- KOVAC, E. AND T. MYLOVANOV (2009): "Stochastic Mechanisms in Settings without Monetary Transfers: The Regular Case," *Journal of Economic Theory*, 144, 1373–1395.
- KRISHNA, V. AND J. MORGAN (2001): "A Model of Expertise," *Quarterly Journal of Economics*, 116, 747–775.
- LAFFONT, J.-J. AND D. MARTIMORT (2002): *The Theory of Incentives: The Principal-Agent Model*, Princeton University Press.
- LEVY, G. AND R. RAZIN (2007): "On the Limits of Communication in Multidimensional Cheap Talk: A comment," *Econometrica*, 75, 885–894.
- LEWIS, T. R. AND D. E. M. SAPPINGTON (1989a): "Countervailing Incentives in Agency Problems," *Journal of Economic Theory*, 49, 294–313.
- (1989b): "Inflexible Rules in Incentive Problems," *American Economic Review*, 79, 69–84.
- MARTIMORT, D. (2007): "Multi-Contracting Mechanism Design," in *Advances in Economic Theory, Proceedings of the 2005 World Congress of the Econometric Society*, ed. by R. Blundell, W. Newey, and T. Person, Cambridge University Press, 56–101.
- MARTIMORT, D. AND A. SEMENOV (2006): "Continuity in Mechanism Design without Transfers," *Economic Letters*, 93, 182–189.
- (2008): "The Informational Effects of Competition and Collusion in Legislative Politics," *Journal of Public Economics*, 92, 1541–1563.
- MCCUBBINS, M., R. NOLL, AND B. WEINGAST (1987): "Administrative Procedures as Instruments of Political Control," *The Journal of Law, Economics and Organization*, 3, 243–277.

- MELUMAD, N. D. AND T. SHIBANO (1991): "Communication in Settings with no Transfers," *Rand Journal of Economics*, 22, 173–198.
- MILGROM, P. AND I. SEGAL (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601.
- MIRRELES, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38, 175–208.
- MYERSON, R. B. (1982): "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," *Journal of Mathematical Economics*, 10, 67–81.
- ROCHET, J. AND P. CHONÉ (1998): "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, 66, 783–826.
- ROYDEN, H. (1988): *Real Analysis*, 3rd Edition, Prentice Hall.