

On Transparency in Organizations

PHILIPPE JEHIEL

Paris School of Economics (Ecole des Ponts, Paritech) and University College London

First version received November 2012; final version accepted October 2014 (Eds.)

When is it best for the Principal to commit to not disclosing all that he/she knows in moral hazard interactions? I show that whenever the Agent would choose the same action under complete information at several distinct states, then full transparency can generically be improved upon. This implies that full transparency is generically suboptimal whenever the dimension of the information held by the Principal exceeds the dimension of the Agent's action. In a simple class of problems, I make further progress on the best disclosure policy assuming the Principal can either disclose fully the state or remain silent.

Key words: Transparency, Information Disclosure, Dimensionality

JEL Codes: D21, D84, D86

1. INTRODUCTION

A central question of economics is how to provide the best incentives to agents in an attempt to improve the working of organizations. Most of the literature in contract theory concerned with this question has focused on how to use monetary instruments so as to best align the objectives of the agents with those of the organization at a minimal cost. But, an equally important design question concerns the distribution of information in organizations insofar as information affects the perceived consequences of actions and thus the incentives of agents.

Consider a Principal–Agent relationship in which the Agent must exert some effort so as to increase the chances of success of the task. The Principal does not observe effort directly. He/She is informed at the time the task is to be completed about how difficult it is to achieve a successful outcome and how much uncertainty there is in measuring success in a reasonable time scale. The Agent is assumed to enjoy some exogenously set bonus when a positive signal about success is received *ex post* by the Principal.

A possible disclosure policy that I refer to as full transparency requires that the Agent be informed before making his/her effort decision about the monitoring technology in the hands of the Principal and how difficult the task is. Alternative less transparent disclosure policies would require that some aspects of the information held by the Principal be kept unknown to the Agent, at least until the completion of the task. The question studied in this article is whether full transparency should be expected to be the disclosure policy that is most preferred by the Principal in such situations.

At some naive level, it would seem that irrespective of what the Principal observes about the difficulty of the task, it is a good idea for the Principal when the monitoring technology is poor to not let the Agent know about it, given that a poor monitoring technology translates into low incentives to exert effort for the Agent. But, if the Agent infers from the absence of communication

on the monitoring technology that monitoring is poor, the Agent may not be induced to work more. It is then unclear whether such an opaque disclosure policy may be effective.

The main result of this article will imply that in situations like the moral hazard one just described in which the information held by the Principal varies over more dimensions than the action of the Agent and the Agent is assumed to make rational inferences from whatever is being disclosed to him, full transparency is generically not the optimal disclosure policy. It should be stressed that the conclusion that full transparency is not optimal in that moral hazard scenario crucially relies on the possibility for the Principal to hide aspects that concern both the quality of the monitoring technology and the difficulty of the task. It would not necessarily hold if the information held by the Principal were restricted to the quality of the monitoring technology, in contrast to what is suggested by the above naive approach.¹

To be more specific, I consider the transparency question in the context of the following general abstract model. An Agent is engaged in a moral hazard interaction parameterized by a state θ . He/She must choose an action a . Both the state θ and the action a can be varied locally. In state θ , action a results in an expected pay-off $u(a; \theta)$ to the Agent and an expected pay-off $\pi(a; \theta)$ to the Principal. At the time the Agent must choose his/her action, the Principal knows the realization of the state θ . Full transparency would require that the Agent be fully informed of the state θ before making his/her choice of action a . The question I am interested in is whether other less transparent disclosure policies could be preferable for the Principal. Or to put it differently, whether in expectation the Principal can achieve a better outcome by having the Agent be incompletely rather than completely informed of the state θ .

In addressing the above transparency question, I have in mind that the disclosure policy regarding θ is decided at an *ex ante* stage before the realization of θ is known to the Principal. Thus, the choice of disclosure policy does not signal anything about which θ is being observed by the Principal in the current interaction of interest.² I believe such a scenario fits in well with the application to organizations in which due to a constant flow of new pieces of information the design of how information is distributed is better thought of from an *ex ante* perspective. I also have in mind that the Agent is fully rational, in particular regarding the inference he/she makes from whatever is being disclosed to him. I also assume that the Principal and the Agent share a common (correct) view about how the parameters are distributed in the economy.

In principle, one could consider arbitrary disclosure policies when addressing the above transparency question, but in some applications, some choices of disclosure policy would seem harder to implement (because they may require information devices that sound less natural). Having this constraint in mind, I consider in some parts of the article the case in which the Principal has only one option: either disclose fully the state θ to the Agent or remain silent and not say anything about the state.

The main result of this article can be stated as follows. For generic organizational objectives π , the Principal can do strictly better than full transparency, as long as there exist at least two distinct states θ_1 and θ_2 such that the action of the Agent would be the same at these two states in the fully transparent benchmark. Such a strict improvement over full transparency can be achieved even

1. In particular, full transparency would be optimal in such a case with a fully rational agent whenever the objective of the Principal is a sufficiently convex function of the effort of the Agent.

2. If the disclosure policy were chosen after the Principal observes the state θ , the work of Crawford and Sobel (1982) on strategic information transmission suggests that not every information held by the Principal would be transmitted to the Agent, as soon as there is some conflict of interest between the Principal and the Agent (the Sender and the Receiver in their words). As I observe later, when addressing the transparency question from an *ex ante* perspective as in this article, full transparency is optimal in the leading quadratic example considered by Crawford and Sobel.

if the Principal is restricted to either disclosing fully the state or remaining completely silent for a well-chosen subset of states.

Before I elaborate on the result, note that in the Principal–Agent scenario considered above, the effort level in the fully transparent case is positively affected by a better monitoring technology or by a more productive task. Thus, the same effort level can be achieved by having a more productive task and a less good monitoring technology for well-chosen magnitudes, thereby ensuring that the main non-transparency result applies to this scenario.

More generally, there will typically be several distinct states that would induce the same choice of action in the fully transparent benchmark, as long as the dimension of the state space exceeds the dimension of the action space of the Agent. For such situations, my main result says that full transparency is not optimal, and that full transparency can be improved upon using a fairly simple disclosure device that only requires that the Principal remains silent about the realization of the state for a well-chosen subset of states.

Importantly, the same non-transparency result extends to situations in which in addition to the disclosure policy, the Principal can use monetary instruments to incentivize the Agent (as most of the contract theory assumes) and to situations with more than one agent, covering applications such as moral hazard in teams. In all cases, full transparency can be improved upon, as long as there exist at least two distinct states that would induce the same choice of actions of the agents in the fully transparent benchmark.

To understand the main result of this article, consider two distinct states θ_1 and θ_2 at which the same action would be chosen in the full transparency benchmark. Assume, for simplicity, that the action of the Agent—referred to as effort—varies over one dimension, and consider a state θ_2^+ in the vicinity of θ_2 such that in the complete information benchmark the effort at θ_2^+ is slightly larger than that at θ_2 and θ_1 .

There are two basic forces that would lead the Principal to prefer that the Agent be uninformed as to whether the state is θ_1 or θ_2^+ . The first force is related to the relative sensitivity of the Principal's objective to the Agent's action at θ_1 and θ_2 : whenever the Principal cares more about the Agent exerting more effort at θ_1 than at θ_2 (and the utility function of the Agent is equally concave at θ_1 and θ_2^+), keeping the Agent in the dark is good for the Principal because it will typically average the effort made by the Agent in the two states, and thus, compared with the fully transparent benchmark, the Agent will make more effort in state θ_1 and less effort in state θ_2^+ , which is overall beneficial because the Principal cares more about effort in state θ_1 than in state θ_2^+ . The second force is related to the relative concavity of the Agent's utility in effort at θ_1 and θ_2^+ : whenever the utility function of the Agent is sufficiently more concave in effort at θ_2^+ than at θ_1 , keeping the Agent in the dark is good for the Principal because deviation from the optimal effort is more costly to the Agent at θ_2^+ than at θ_1 and thus the opaque disclosure policy leads to an effort level (at both θ_1 and θ_2^+) closer to that at θ_2^+ , which is favourable to the Principal.

I provide later a sense of how the relative sensitivity of the Principal's objective to the Agent's effort on the one hand and the relative concavity in effort of the Agent's utility on the other should be aggregated for the Principal to prefer that the Agent be uninformed as to whether the state is θ_1 or θ_2^+ . Whenever this condition holds, it is good not to let the Agent know whether the state is θ_1 or θ_2^+ . And whenever the condition does not hold, I show that for generic specifications of the Principal's objective, the Principal finds it strictly beneficial not to let the Agent know whether θ is θ_1 or θ_2^- where θ_2^- is defined so that θ_2 is the middle point between θ_2^- and θ_2^+ . Thus, for generic organizational objectives, the Principal is strictly better off either when the Agent does not know whether $\theta = \theta_1$ or θ_2^+ or when he/she does not know whether $\theta = \theta_1$ or θ_2^- as compared with the full transparency benchmark. Completing the proof that the Principal can gain in expected terms

over the full transparency benchmark requires that the Agent be kept in the dark over a positive measure of states, and such a result can be obtained using simple continuity arguments. I note that the gain can be achieved even if the Principal is restricted to either disclosing fully the state or remaining completely silent.

The above result does not say how much can be gained over full transparency. I illustrate through the above moral hazard scenario that the gains over full transparency can be arbitrarily large in relative terms. In the final part of the article, I also characterize when the Principal should optimally remain silent about the state in the context of a simple class of problems that includes the above moral hazard application. I use these results to shed some light on how to define the breadth of task assignments in organizations.

The transparency question addressed in this article is closely related to the optimal information disclosure question considered in Rayo and Segal (2010) or Tamura (2012) and to the Bayesian persuasion question considered in Kamenica and Gentzkow (2011). These authors (mostly) consider Sender–Receiver interactions without monetary instruments in which the Sender possesses private information and the Receiver chooses an action based on the information he/she has (or infers from the Sender’s communication). These papers ask in specific contexts: which disclosure policy should the Sender commit to so as to maximize his/her expected pay-off? It should be noted that these papers do not consider the possibility that the only option for the Sender is to remain silent or disclose fully the state.

Rayo and Segal (2010) observe in their set-up that transparency is not best when there are no monetary instruments. Their result can be viewed as providing an illustration of the main non-transparency result of this article. In Rayo and Segal, the action of the Receiver can be described as the probability of accepting the project proposed by the Sender (even though in their model this action is derived from the random realization of the Receiver’s outside option), and the state can be described as the profile of pay-offs the Sender and the Receiver would receive in case the project would be adopted. For a given Receiver’s pay-off of implementing the project, the probability of acceptance of the Receiver is the same irrespective of how valuable the project is to the Sender. Thus, the key condition of the main Proposition of this article is satisfied, thereby explaining why full transparency can be improved upon in Rayo and Segal’s set-up. In a later part of their paper, Rayo and Segal observe that when the Sender can use monetary transfers, full transparency becomes optimal. Yet, the transparency result they obtain when side-payments are allowed would generically not hold when perturbing the Sender’s pay-off, as the non-transparency result derived in Section 5 shows. Although Rayo and Segal have offered an insightful exploration of the optimal disclosure policy in an interesting example, I believe the general treatment offered in this article provides a better account of why full transparency should be expected to be suboptimal in general moral hazard interactions with or without monetary instruments when the dimension of the state space exceeds the dimension of the action space.³

Kamenica and Gentzkow (2011) mostly consider situations in which the state and the action take values over discrete realizations, and they note that the optimal disclosure policy typically involves a well-adjusted noisiness in the precision of the transmission so as to optimally let the Agent be indifferent over several actions in some instances. Although Kamenica and Gentzkow (2011) provide a general way to cope with such investigations, I note that my set-up makes an

3. In Section 5, I also discuss why the insight about the optimality of full disclosure in exchange for well-adjusted side-payments obtained by Eso and Szentes (2007) in an auction set-up does not apply in general here (essentially because the full information disclosure would not in general result in the maximization of total welfare in typical moral hazard situations unlike in the auction set-up). In a related vein, note that Sobel (1993) observes that in the context of Principal–Agent interactions with optimally adjusted monetary instruments, full disclosure need not be optimal if the Agent is not risk neutral.

essential use of the possibility that the action and the state be locally varied, which plays no role in Kamenica and Gentzkow. I also note that their focus on Bayesian persuasion or on when it is best for the Sender to disclose some information (as opposed to none) to the Receiver lies at the other extreme of the non-transparency question considered in this article.

The question addressed in this article is also tangentially related to other strands of literature. For example, some papers have shown in adverse selection environments such as auctions or monopoly that it is best for the designer to transmit as much information as possible to the agent(s) whenever the information of the designer is affiliated with the information of the agent(s).⁴ The main non-transparency result of this article suggests that, in moral hazard environments, when the information of the designer has more dimensions than the action of the agent, there is no way (natural or not) to make full transparency optimal.

The rest of the article is structured as follows. A general Principal–Agent framework is presented in Section 2. Preliminary considerations together with a preliminary investigation of a leading moral hazard example are gathered in Section 3. Section 4 states, and discusses, the main non-transparency result. Section 5 considers the extensions to multi-agent settings and to the case in which side-payments are allowed. Section 6 considers the optimal way to stay silent in a simple class of scenarios that includes Rayo–Segal’s model as well as the above moral hazard environment. Section 7 concludes.

2. A PRINCIPAL–AGENT FRAMEWORK

I consider a family of moral hazard problems with one agent parameterized by a state variable $\theta \in \mathbb{R}^m$. The state θ is assumed to be distributed according to a smooth (*i.e.* twice continuously differentiable) density $p(\theta)$ that is strictly positive on some open bounded subset of \mathbb{R}^m . In every state θ , the Agent chooses an action a in A , an open subset of \mathbb{R}^n .

In state θ , the expected pay-off to the Agent is $u(a; \theta)$ when the Agent chooses a . The corresponding expected pay-off to the Principal is denoted by $\pi(a; \theta)$. I assume that u is a concave function of a that varies smoothly with θ . Moreover, I assume that whatever θ , the function $a \rightarrow u(a; \theta)$ is always maximized in a bounded subset of A .⁵ The function π is also assumed to be smooth (*i.e.* twice continuously differentiable).

The general theme I wish to explore is whether it is in the interest of the Principal that the Agent be informed of θ whatever its realization before he/she makes his/her choice of action a . Specifically, I ask:

Question 1. When is it beneficial for the Principal that the Agent be incompletely informed of θ ? Or to put it differently, when is some form of non-transparency desirable?

When some form of non-transparency is desirable, it would be of interest to know more about the best disclosure policy for the Principal. In some parts of the article (when putting more structure on the problem), I address the following question:

Question 2. Assuming the Principal can *ex ante* commit for each realization of θ to either telling what the realization of the state θ is or remaining silent about this, what is the best strategy for the Principal?

4. The relevant reference for auctions is Milgrom and Weber (1982), and for monopoly it is Ottaviani and Prat (2001).

5. Such an assumption will typically guarantee that one can work with first-order conditions to deal with the maximization problem of the agent.

When addressing the above questions, observe that I am assuming that the Principal can commit in advance (before knowing the realization of θ) to the chosen disclosure policy. This implies that the choice of information disclosure policy cannot help the Agent in his/her estimate of θ , since that choice is made before the Principal knows θ . This is the same *ex ante* view as the one adopted in Rayo–Segal (2010), Kamenica–Gentzkow (2011), and it seems appropriate to deal with organizations in which there is enough time to commit in advance (before the realization of θ is known) to whatever disclosure policy sounds best.

Although question 2 does not deal with the most general form of communication strategy, I believe that the option of remaining silent is a common (and widespread) communication device. It is also fairly easy to implement. Of course, the premise of the analysis is that the Agent would make correct inferences as to how the state θ is distributed from whatever is being disclosed to him, including when the Principal remains silent.

3. PRELIMINARIES

3.1. Preference alignments

Clearly, if the Principal and the Agent have the same preferences, $\pi = u$, then full transparency is best because in a decision problem one can only benefit from having more information. In such a case, the best disclosure policy is for the Principal to tell the Agent the realization θ .

At the other extreme, if the Principal and the Agent have completely opposed preferences in the sense that there exists a constant c such that for all a and θ , $u(a; \theta) + \pi(a; \theta) = c$, then remaining completely silent as to the realization of θ is the best disclosure policy for the Principal, given that the more informed the Agent, the better for him and hence the worse for the Principal in this zero-sum context.

From these two basic observations, it would seem that the preference alignment between the Principal and the Agent is the main driving force behind the desirability of transparency. Yet, consider the quadratic example of Crawford and Sobel (1982) in which preferences can be represented as:

$$\begin{aligned} u(a; \theta) &= -(a - \theta - b)^2 \\ \pi(a; \theta) &= -(a - \theta)^2 \end{aligned}$$

The Principal would like the action a of the Agent to be as close as possible to the state θ whereas the Agent would like this action to be as close as possible to $\theta + b$ where b can be interpreted as measuring the size of the bias of the Agent or the degree of misalignment of the preferences of the Agent and the Principal.

Interestingly, as also noted in Kamenica–Gentzkow (2011), in this quadratic example, the best disclosure policy for the Principal is full transparency no matter how big b is, the reason being that anyway the Agent will apply the bias b to his/her choice of a and given the concavity of $a \rightarrow -(a - \theta)^2$ it is never in the interest of the Principal that the Agent be incompletely informed about θ .⁶

6. When several θ belong to the same information set I , the Agent picks $a^{CI} = E(\theta | \theta \in I) + b$ and conditional on $\theta \in I$ the Principal gets $-E[(a^{CI} - \theta)^2 | \theta \in I]$ which is less (by Jensen's inequality) than $-(a^{CI} - E(\theta | \theta \in I))^2 = -b^2$. Given that $-b^2$ is what the Principal gets under full transparency, we get the desired conclusion that transparency is optimal whatever b (and whatever the distribution of θ).

3.2. Basic considerations

The Crawford–Sobel quadratic example casts doubt as to whether the misalignment of preferences of the Principal and the Agent is the key driving force behind the desirability of non-transparency. In this subsection, I suggest that two forces may lead the Principal to prefer a non-fully transparent disclosure policy. These two forces will later be used to interpret most of the results derived in subsequent sections.

To formulate these two forces, I consider the case of the bundling of two states θ_1 and θ_2 into one information set $I = \{\theta_1, \theta_2\}$ for the Agent, and I illustrate two extreme cases in which the Principal may prefer this to letting the Agent know whether $\theta = \theta_1$ or θ_2 . The first of these cases is related to the relative concavity of the Agent's utility function with respect to his/her own action at these two states, whereas the second of these cases is related to the relative sensitivity of the Principal's utility function with respect to the Agent's action.⁷

To present this most simply, assume that the action space A is one-dimensional. When the Agent knows the state θ , he/she picks action

$$a(\theta) = \arg \max_a u(a; \theta).$$

When the Agent does not know whether θ is θ_1 or θ_2 , he/she picks action

$$a^{CI} = \arg \max_a [p(\theta_1)u(a; \theta_1) + p(\theta_2)u(a; \theta_2)].$$

Assume further that $a(\theta_2) > a(\theta_1)$. The concavity of $a \rightarrow u(a; \theta)$ for all θ ensures that

$$a(\theta_1) < a^{CI} < a(\theta_2)$$

3.2.1. Relative concavity of u . Clearly, if $u(\cdot; \theta)$ is much more concave at θ_2 than at θ_1 —in the sense that $\left| \frac{\partial^2 u}{\partial a^2}(\cdot; \theta_2) \right|$ is sufficiently (and uniformly) bigger than $\left| \frac{\partial^2 u}{\partial a^2}(\cdot; \theta_1) \right|$ —then a deviation at θ_2 from $a(\theta_2)$ would be too costly (as compared with a deviation from $a(\theta_1)$ at θ_1) and thus a^{CI} must be close to $a(\theta_2)$. To the extent that $\pi(\cdot; \theta)$ is increasing in a , the Principal strictly benefits from the Agent not knowing whether $\theta = \theta_1$ or θ_2 in such a case.

3.2.2. Relative sensitivity of π . For a given $u(\cdot; \theta)$, if the principal cares much more about the Agent's effort a at θ_1 than at θ_2 —in the sense that $\left| \frac{\partial \pi}{\partial a}(\cdot; \theta_1) \right|$ is sufficiently (and uniformly) bigger than $\left| \frac{\partial \pi}{\partial a}(\cdot; \theta_2) \right|$ —then the increase of a (from $a(\theta_1)$ to a^{CI}) at θ_1 will more than compensate (in terms of Principal's pay-off) the decrease of a (from $a(\theta_2)$ to a^{CI}) at θ_2 so that the Principal strictly prefers that the Agent be uninformed as to whether $\theta = \theta_1$ or θ_2 .

The two basic effects of relative concavity of u and relative sensitivity of π will play a key role in the rest of the article.

7. A third effect in favour of non-transparency one might think of is related to the concavity of the Principal's pay-off function π with respect to the agent's action a , but this turns out not to play any role in the derivation of the non-transparency result, as will be explained next.

3.3. *An illustrative example*

Consider the following moral hazard scenario. An Agent must choose an effort level $a \in [0, 1]$ to increase the chance of success of a task. Effort a gives a private benefit $b \cdot a$ (e.g. proportional to a) to the Agent. The cost of exerting effort a is $c(a) = \frac{\lambda}{2} a^2$ where λ can be interpreted as a parameter defining the degree of decreasing returns to scale in the activity. Depending on the effort a , the probability of success for the organization is γa where $\gamma < 1$. Success is not directly observed by the Principal (because there are many activities in the organization and it is not obvious to pin down whether an isolated activity is successful or not). What the Principal observes is a signal σ which can take two values $\sigma = 1$ or 0 . The observation of σ provides a noisy signal as to whether the activity is successful. Specifically, assume that $\Pr(\sigma = 1 | success) = \beta > \frac{1}{2}$. The Agent receives a bonus of 1 in case $\sigma = 1$ is observed. The Principal receives an extra benefit R in case of success. Both the Principal and the Agent are assumed to be risk neutral.

The state that describes the interaction can be parameterized by $\theta = (b, \lambda, \beta, \gamma, R)$. In a given state θ with effort level a , the expected utility of the Agent is:

$$[\gamma a \beta + (1 - \gamma a)(1 - \beta)] + ba - \frac{\lambda}{2} a^2$$

where the first (resp. second) term in the bracket represents the probability that the Principal observes $\sigma = 1$ and the task is successful (resp. non-successful) and thus the overall term in bracket represents the expected bonus received by the Agent when exerting effort a .

The corresponding expected utility of the Principal is:

$$\gamma a [R - (2\beta - 1)] - (1 - \beta).$$

For future reference, one can rewrite the Agent and Principal's expected pay-offs as (P)

$$u(a; \theta) = xa - \frac{y}{2} a^2 + K$$

$$\pi(a; \theta) = za + K'$$

where $x = \gamma\beta - \gamma(1 - \beta) + b$, $y = \lambda$, $z = \gamma[R - (2\beta - 1)]$, $K = 1 - \beta$ and $K' = -(1 - \beta)$, and redefine θ to be (x, y, z, K, K') .

In this class of problems, if the Agent is uncertain as to which state he is facing, he/she picks action⁸

$$a = \frac{E(x)}{E(y)}$$

where the expectation is computed according to the belief of the Agent regarding the state. The expected pay-off of the principal in the corresponding range of states amounts to

$$\frac{E(x)E(z)}{E(y)} + K'$$

In principle, these expressions allow us to compare the effect on the Principal's expected pay-off of any disclosure policy.

8. I am assuming that in the range of relevant θ , one has $0 < \frac{E(x)}{E(y)} < 1$.

Just to illustrate how much can be gained with non-transparency, assume that there are two states equally likely in which only β and γ vary (specifically, assume $b=0$, $\lambda=R=1$). The monitoring technology takes two possible forms: either the signal σ received by the Principal is totally uninformative about success so that $\beta=\frac{1}{2}$, or the signal σ is perfectly informative so that $\beta=1$. Moreover, when the signal is uninformative about success, the sensitivity of success to effort is highest with $\gamma=1$, and when the signal σ is perfectly informative about success, the sensitivity of success to effort is lower and $\gamma=\underline{\gamma}<1$. In such a case, non-transparency gives an expected pay-off to the Principal proportional to $(1+\underline{\gamma})\underline{\gamma}$, whereas transparency gives an expected pay-off proportional (with the same coefficient) to $(\underline{\gamma})^2$. The gain of non-transparency in relative terms is then $\frac{1-\underline{\gamma}}{2\underline{\gamma}}$ and it can be made arbitrarily large as $\underline{\gamma}$ is made smaller and smaller. Here non-transparency is good because it allows the Principal to incentivize the Agent to exert more effort in the state in which effort leads to more success.

Observe that in the two-state scenario just considered if γ and $\gamma(2\beta-1)$ were positively (instead of negatively) correlated, then transparency over these two states would be better than non-transparency. Yet, as I show later, in cases like the one considered here in which the same action $a(\theta)$ would be chosen at several states were the Agent to be fully informed of the state, one can always find a set of states so that if the Agent is not told which state in the set occurred, the Principal is strictly better off than if the Agent is fully informed of the state.

It should be mentioned that the model studied by Rayo and Segal (2010) can be cast in the class (P) by letting $y=1$, $K=K'=0$ and interpreting a as the probability of acceptance of the project by the receiver, z as the value to the Sender of the project and x as the value to the Receiver of the project. Although Rayo and Segal assume acceptance is decided after the Receiver observes his/her outside option (which is assumed to be distributed according to a uniform distribution), one gets the same acceptance probability as the one arising from problem (P) as specified with $y=1$, thereby explaining the equivalence of the two formulations.

An important observation made by Rayo and Segal is that if two states (x_1, z_1) and (x_1, z_2) are unordered in the sense that $x_2 > x_1$ implies that $z_2 < z_1$, then it is preferable not to let the Agent know whether the state is (x_1, z_1) or (x_1, z_2) . This observation can be related to the above identification of the role of the relative sensitivity of π with respect to the Agent's effort. Whenever $x_2 > x_1$, it must be that the action at state (x_2, z_2) is larger than at state (x_1, z_1) , since $a_i = x_i$ for $i=1, 2$. Since $\frac{\partial \pi}{\partial a} = z$, the condition $z_2 < z_1$ precisely means that the Principal cares more about effort in the state in which the full information effort would be smaller. This explains, through the lens of the basic considerations proposed in Section 3.2, Rayo and Segal's observation. I will come back to Rayo and Segal later in the article including when I discuss the effect of allowing for monetary payments and discussing the optimal way of remaining silent in the context of problems in class (P).

4. NON-TRANSPARENCY

In this section, I establish quite generally that if the same action a were to be chosen in at least two distinct states θ_1, θ_2 in the benchmark scenario in which the Agent would be fully informed of the state, then for generic objective functions π , the Principal can achieve a strictly higher expected pay-off by not disclosing fully θ . Thus, some form of non-transparency is desirable. As a corollary, I obtain that full transparency is not the optimal disclosure policy whenever the dimension of the state space exceeds the dimension of the action space because in such cases there must exist distinct states that induce the same choice of action in the complete information benchmark.

4.1. The basic non-transparency result

To establish my first non-transparency result, let me denote by $a(\theta) = \arg \max_a u(a; \theta)$. I will assume that for all θ , there is always a direction δ of θ such that

$$\frac{\partial}{\partial \delta} \nabla_a u(a(\theta); \theta) \neq 0$$

where $\nabla_a u$ refers to the gradient of u with respect to a . When u satisfies this assumption, I say that u is non-satiated. Such an assumption typically implies that maxima $a(\theta)$ are never locally constant (it is in fact equivalent to this). It should typically be thought of as a fairly weak assumption as such a direction of θ may, for example, correspond to a reduction in the marginal cost of producing effort in some of the tasks to be fulfilled by the Agent. This assumption was trivially met in the moral hazard scenario considered in Section 3.

I will also need to define formally what I mean by generic objective functions π . I provide below a set theoretic definition of genericity (so as to deal with the space of objective functions in great generality) but note that a measure theoretic definition would work equally well if I were, for example, to consider that objective functions can only take polynomial forms, say. Formally, let $X = \mathbb{R}^n \times \mathbb{R}^m$ denote the domain of the objective functions π . Consider functions that are twice continuously differentiable $\pi \in C^2(X)$. The set $\bar{\Pi}$ of $\pi \in C^2(X)$ is endowed with a Whitney C^2 topology by letting a sequence $\pi_k \in \bar{\Pi}$ converge to π if and only if $\pi_k - \pi$ as well as the Jacobian of $\pi_k - \pi$ and the matrix of second derivative of $\pi_k - \pi$ converge uniformly to zero in the space of continuous functions over the relevant range of a, θ (assumed to be bounded, see above). The notion of genericity is defined as follows:

Definition. A set $\Pi \subseteq \bar{\Pi}$ is generic in $\bar{\Pi}$ if it contains a set that is open and dense in $\bar{\Pi}$ for the Whitney C^2 topology.

The first main result is:

Proposition 1. *Assume that there exist two distinct states θ_1, θ_2 in the interior of the θ -space such that the same action would be chosen in the complete information benchmark, i.e. $a(\theta_1) = a(\theta_2)$. Assume that the utility function u is non-satiated. There exists a generic set Π such that for all objective functions $\pi \in \Pi$, the Principal can do strictly better than disclosing fully the state θ to the Agent.*

The general case covered by Proposition 1 is shown in the Appendix. I now provide a detailed intuition for the case in which the action space is one-dimensional $a \in \mathbb{R}$. The complete information solution $a(\theta)$ satisfies: $\frac{\partial}{\partial a} u(a(\theta); \theta) = 0$. Let θ_1 and $\theta_2 \neq \theta_1$ be such that $a(\theta_2) = a(\theta_1)$, as allowed by Proposition 1, and consider a direction δ such that $\frac{\partial^2 u}{\partial a \partial \delta}(a(\theta_1); \theta_2) \neq 0$, as allowed by the non-satiation of u .

Consider the states $\theta = \theta_1$ and $\theta_2 + \varepsilon \delta$ for ε either positive or negative but small (remember θ_2 lies in the interior of the θ -space). The central part of the argument consists in comparing the aggregate expected pay-off to the Principal when the Agent knows whether $\theta = \theta_1$ or $\theta_2 + \varepsilon \delta$ and the expected pay-off to the Principal when the Agent ignores whether $\theta = \theta_1$ or $\theta_2 + \varepsilon \delta$. From there, it will be relatively straightforward to conclude that the Principal can strictly gain in expected terms as compared with the full information benchmark.

Clearly for $\varepsilon = 0$, the two informational scenarios generate the same aggregate expected value of π . This is because the same action being optimal for the Agent at θ_1 and θ_2 ($a(\theta_1) = a(\theta_2)$), the uncertainty as to whether $\theta = \theta_1$ or θ_2 still leads the Agent to find $a(\theta_1) = a(\theta_2)$ optimal. But, for $\varepsilon \neq 0$, the two solutions will not in general lead to the same aggregate effect on π . I will now

compute the first-order effect in ε of this difference and show that it is generically different from 0, thereby allowing me to conclude that coarse information of the above type either for $\varepsilon > 0$ and small or $\varepsilon < 0$ and small dominates the complete information benchmark.

Let $a_1 = a(\theta_1)$ and $a_2(\varepsilon) = a(\theta_2 + \varepsilon\delta)$. They satisfy

$$\frac{\partial}{\partial a} u(a_1; \theta_1) = 0 \quad (1)$$

$$\frac{\partial}{\partial a} u(a_2(\varepsilon); \theta_2 + \varepsilon\delta) = 0$$

Let $a^{CI}(\varepsilon)$ denote the action when the Agent does not know whether $\theta = \theta_1$ or $\theta_2 + \varepsilon\delta$. It satisfies:

$$p(\theta_1) \frac{\partial}{\partial a} u(a^{CI}(\varepsilon); \theta_1) + p(\theta_2 + \varepsilon\delta) \frac{\partial}{\partial a} u(a^{CI}(\varepsilon); \theta_2 + \varepsilon\delta) = 0.$$

I wish to sign $\Delta(\varepsilon)$ defined as

$$p(\theta_1)[\pi(a_1; \theta_1) - \pi(a^{CI}(\varepsilon); \theta_1)] + p(\theta_2 + \varepsilon\delta)[\pi(a_2(\varepsilon); \theta_2 + \varepsilon\delta) - \pi(a^{CI}(\varepsilon); \theta_2 + \varepsilon\delta)].$$

Clearly, if $\Delta(\varepsilon) < 0$, it is strictly better that the Agent does not know whether $\theta = \theta_1$ or $\theta_2 + \varepsilon\delta$.

I now expand $\Delta(\varepsilon)$ at the first order in ε . Since $a^{CI}(0) = a_2(0) = a_1$, $\Delta(\varepsilon)$ can be written (at the first order in ε) as:

$$p(\theta_1) \frac{\partial \pi}{\partial a}(a_1; \theta_1)[a_1 - a^{CI}(\varepsilon)] + p(\theta_2) \frac{\partial \pi}{\partial a}(a_1; \theta_2)[a_2(\varepsilon) - a^{CI}(\varepsilon)] + o(\varepsilon)$$

where $o(\varepsilon)$ denotes a function such that $\frac{o(\varepsilon)}{\varepsilon}$ goes to 0 as ε goes to 0.

Moreover from (1) (and using that $\frac{\partial^2 u}{\partial a^2} < 0$ is different from 0), one has that:

$$a_2(\varepsilon) - a_1 = \frac{-\frac{\partial^2 u}{\partial a \partial \delta}(\theta_2)}{\frac{\partial^2 u}{\partial a^2}(\theta_2)} \varepsilon + o(\varepsilon)$$

$$a^{CI}(\varepsilon) - a_0 = \frac{-p(\theta_2) \frac{\partial^2 u}{\partial a \partial \delta}(\theta_2)}{p(\theta_1) \frac{\partial^2 u}{\partial a^2}(\theta_1) + p(\theta_2) \frac{\partial^2 u}{\partial a^2}(\theta_2)} \varepsilon + o(\varepsilon)$$

where $\partial h / \partial \delta$ denotes the derivative of (an arbitrary function denoted) h in the direction δ and all functions are taken at $a = a_1$.

After multiplying $\Delta(\varepsilon)$ by $\frac{\partial^2 u}{\partial a^2}(\theta_2)[p(\theta_1) \frac{\partial^2 u}{\partial a^2}(\theta_1) + p(\theta_2) \frac{\partial^2 u}{\partial a^2}(\theta_2)]$ and dividing by $p(\theta_1)p(\theta_2)$ (which are both strictly positive) I get that $\Delta(\varepsilon)$ has the same sign as

$$\left[\frac{\partial \pi}{\partial a}(\theta_1) \frac{\partial^2 u}{\partial a^2}(\theta_2) - \frac{\partial \pi}{\partial a}(\theta_2) \frac{\partial^2 u}{\partial a^2}(\theta_1) \right] \frac{\partial^2 u}{\partial a \partial \delta}(\theta_2) \varepsilon + o(\varepsilon)$$

Three cases may *a priori* occur.

(i) $\left[\frac{\partial \pi}{\partial a}(\theta_1) \frac{\partial^2 u}{\partial a^2}(\theta_2) - \frac{\partial \pi}{\partial a}(\theta_2) \frac{\partial^2 u}{\partial a^2}(\theta_1) \right] \frac{\partial^2 u}{\partial a \partial \delta}(\theta_2) < 0$. Then taking $\varepsilon > 0$ and sufficiently small, I can infer from the above that not letting the Agent know whether $\theta = \theta_1$ or $\theta_2 + \varepsilon\delta$ strictly dominates the complete information benchmark.

(ii) Likewise, if $\left[\frac{\partial \pi}{\partial a}(\theta_1) \frac{\partial^2 u}{\partial a^2}(\theta_2) - \frac{\partial \pi}{\partial a}(\theta_2) \frac{\partial^2 u}{\partial a^2}(\theta_1) \right] \frac{\partial^2 u}{\partial a \partial \delta}(\theta_2) > 0$, then taking $\varepsilon < 0$ and sufficiently small, not letting the Agent know whether $\theta = \theta_1$ or $\theta_2 + \varepsilon \delta$ strictly dominates the complete information benchmark (remember that since θ_2 is in the interior of the θ -space, one can move in any direction from θ_2).

(iii) The only case in which one cannot conclude is when

$$\left[\frac{\partial \pi}{\partial a}(\theta_1) \frac{\partial^2 u}{\partial a^2}(\theta_2) - \frac{\partial \pi}{\partial a}(\theta_2) \frac{\partial^2 u}{\partial a^2}(\theta_1) \right] \frac{\partial^2 u}{\partial a \partial \delta}(\theta_2) = 0$$

or

$$\frac{\partial \pi}{\partial a}(\theta_1) \frac{\partial^2 u}{\partial a^2}(\theta_2) - \frac{\partial \pi}{\partial a}(\theta_2) \frac{\partial^2 u}{\partial a^2}(\theta_1) = 0 \tag{2}$$

But, this condition is not satisfied for generic π functions.

To see this formally, consider the family of π_λ functions

$$\pi_\lambda(a; \theta) = \pi(a; \theta) + \lambda a \|\theta - \theta_1\|^2$$

where $\lambda \in \mathbb{R}$ and $\|\theta - \theta_1\|$ denotes the Euclidean distance between θ and θ_1 . Obviously, if π satisfies (2), then for $\lambda \neq 0$, π_λ does not satisfy (2). This is because changing π does not affect the expressions of $a_2(\varepsilon)$, $a^{CI}(\varepsilon)$. From this, one can conclude that the set of π for which (2) does not hold is dense. Moreover, this set is also open given the continuity of the mapping $\pi \rightarrow \frac{\partial \pi}{\partial a}(\theta_1) \frac{\partial^2 u}{\partial a^2}(\theta_2) - \frac{\partial \pi}{\partial a}(\theta_2) \frac{\partial^2 u}{\partial a^2}(\theta_1)$ according to the Whitney C^2 topology.⁹

The rest of the argument, in particular, showing how one can induce a strict gain in expectation (integrating over all possible realizations of θ) relies on a simple continuity argument. If not letting the Agent know whether $\theta = \theta_1$ or $\theta_2 + \varepsilon \delta$ is strictly preferable to full transparency (as just considered) then there must exist (small enough) neighbourhoods B_1 , B_2 of θ_1 and $\theta_2 + \varepsilon \delta$, respectively, such that not letting the Agent know $\theta \in B_1 \cup B_2$ induces an expected (strictly) positive gain in terms of the Principal's objective as compared with the full information benchmark. Observe that this gain can be achieved even in the scenario in which the Principal is constrained either to disclose the state θ or to remain silent (as considered later in the simple class (P) introduced at the end of Section 3).

4.2. Discussion

In this subsection, I first interpret the two-state construction shown above in light of the basic considerations made in Section 3. I then show the implication of Proposition 1 for the case in which the dimension of the state space is higher than the dimension of the action space.

4.2.1 Why is full transparency dominated? Suppose in the two-state scenario considered immediately after Proposition 1 that the action of the Agent is strictly larger at $\theta_2 + \varepsilon \delta$ than at θ_1 , that is, $a(\theta_2 + \varepsilon \delta) > a(\theta_1)$.

If the Principal is more sensitive to effort at θ_1 than at θ_2 in the sense that $\frac{\partial \pi}{\partial a}(\theta_1) > \frac{\partial \pi}{\partial a}(\theta_2)$ while the Agent's utility function has the same concavity at the two states ($\frac{\partial^2 u}{\partial a^2}(\theta_1) = \frac{\partial^2 u}{\partial a^2}(\theta_2)$), then not letting the Agent know whether $\theta = \theta_1$ or $\theta_2 + \varepsilon \delta$ is good for the Principal because it

9. Clearly, if one were to consider polynomial functions π , then (2) would not hold for a measure 1 set of parameter values.

allows the transfer of some effort made in state $\theta_2 + \varepsilon\delta$ to state θ_1 and the Principal is more sensitive to extra effort in state θ_1 .

Alternatively, if the Principal is equally sensitive to the Agent's effort in states θ_1 and θ_2 ($\frac{\partial \pi}{\partial a}(\theta_1) = \frac{\partial \pi}{\partial a}(\theta_2) > 0$) and the Agent's utility function is more concave at θ_1 than at θ_2 in the sense that $|\frac{\partial^2 u}{\partial a^2}(\theta_1)| > |\frac{\partial^2 u}{\partial a^2}(\theta_2)|$ then not letting the Agent know whether $\theta = \theta_1$ or $\theta_2 + \varepsilon\delta$ is good for the Principal. This is because deviation from the optimal effort is more costly at $\theta_2 + \varepsilon\delta$ than at θ_1 and thus keeping the Agent in the dark will lead to an effort level (at both θ_1 and $\theta_2 + \varepsilon\delta$) closer to that at $\theta_2 + \varepsilon\delta$, which is favourable to the Principal (because the effort at $\theta_2 + \varepsilon\delta$ in the complete information case is assumed to be larger than that at θ_1).

The above two cases for non-transparency are, of course, closely related to the basic observations made in Section 3. Incorporating both effects of the relative sensitivity of π to a on the one hand and the relative concavity of u on the other reveals that not letting the Agent know whether $\theta = \theta_1$ or $\theta_2 + \varepsilon\delta$ dominates full transparency whenever

$$-\frac{\partial \pi / \partial a}{\partial^2 u / \partial a^2}(\theta_1) > -\frac{\partial \pi / \partial a}{\partial^2 u / \partial a^2}(\theta_2) \quad (3)$$

which provides an exact expression of how the two effects should be aggregated.

Observe that the key observation behind Proposition 1 is that either (3) is satisfied, and not letting the Agent know whether $\theta = \theta_1$ or $\theta_2 + \varepsilon\delta$ is good for the Principal, or else one should generically have that $-\frac{\partial \pi / \partial a}{\partial^2 u / \partial a^2}(\theta_1) < -\frac{\partial \pi / \partial a}{\partial^2 u / \partial a^2}(\theta_2)$ and then not letting the Agent know whether $\theta = \theta_1$ or $\theta_2 - \varepsilon\delta$ is good for the Principal.

Two remarks may be worth making here. First, it should be noted that the argument developed above makes no use of the concavity of the objective function π with respect to a . One might have conjectured that the concavity of π would play a role in the argument given that a coarser information of the Agent reduces the variability of the action with the state and this reduced variability would seem to be favourable to the Principal whenever π is concave in a . Yet, as Proposition 1 implies, the desirability of some form of non-transparency holds no matter whether π is concave or not, in particular even in cases in which π would be extremely convex.

The reason for this seemingly surprising result is that by considering the bundling of the states $\theta = \theta_1$ or $\theta_2 + \varepsilon\delta$ which are such that $a(\theta_1) = a(\theta_2)$, the concavity of π has an effect of order ε^2 on the comparison as to whether the Principal prefers that the Agent be uninformed about the state (by contrast, the previous highlighted effects are of order ε). As a result, the concavity of π plays no role as to whether some form of non-transparency may be desirable under the assumptions made in Proposition 1. Another way to phrase this is that the choice of information structure made above is such that the effect of the concavity of π is frozen and only the relative concavity of u and the relative sensitivity of π with respect to a play a role. Of course, the concavity of π with respect to a will play a role as to what exact form of information disclosure is best for the Principal. But no matter what π looks like, some form of non-transparency will be desirable, as long as there exist at least two distinct states that would induce the same choice of action in the complete information benchmark.¹⁰

It should also be stressed that if in the above construction, I had considered $\theta_2 = \theta_1$, then the first-order effects as expressed above would cancel out (this can be seen by plugging $\theta_2 = \theta_1$ in

10. In relation to the latter point, it can be said that if I had considered two distinct states θ_1, θ_2 with $a(\theta_1) \neq a(\theta_2)$, then the concavity of π would have played a non-negligible role in the assessment as to whether the Principal would be better off when the agent does not know the state $\theta \in \{\theta_1, \theta_2\}$. More precisely, it is readily verified that, for sufficiently convex π , the Principal would have been strictly better off in the full information benchmark as compared with the situation in which the agent does not know whether $\theta = \theta_1$ or θ_2 .

the above expressions). Whether the Principal is better off when the Agent knows whether $\theta = \theta_1$ or $\theta_1 + \varepsilon\delta$ would involve a comparison of order ε^2 and it would thus not be possible to conclude that some form of non-transparency is desirable.¹¹

Comment. As can be inferred from the above, in the case in which the dimension of the action space is 1, the relative density $p(\theta)$ around $\theta = \theta_1$ and θ_2 plays no role as to whether the Principal is better off when the Agent does not know whether $\theta = \theta_1$ or $\theta_2 + \varepsilon\delta$. However, when the action space has a dimension larger than 1, this relative density may play a role. Inspecting the proof of Proposition 1 in the Appendix reveals this.

4.2.2. When the state space has more dimensions than the action space. A simple corollary of Proposition 1 is obtained when the state $\theta \in \mathbb{R}^m$ varies over more dimensions than the action $a \in \mathbb{R}^n$, *i.e.* $m > n$. Indeed in such a case, there are typically many different states θ that induce the same action in the full information benchmark. Specifically, given the smoothness of u , for almost any θ_1 ,

$$\bar{\Theta}(\theta_1) = \{\theta \text{ such that } a(\theta) = a(\theta_1)\}$$

defines a manifold of dimension $m - n$ (see, for example, Milnor (1965), chapter 2). Thus one can find $\theta_2 \neq \theta_1$ such that $a(\theta_2) = a(\theta_1)$. As a consequence of Proposition 1, I can state:

Theorem 1. *Suppose the dimensionality of the state θ is strictly bigger than the dimensionality of the action a of the Agent, that is, $m > n$, and suppose the utility of the Agent is non-satiated. Then there exists a generic set Π such that for all objective functions $\pi \in \Pi$, some non-full disclosure of θ to the Agent is strictly beneficial to the Principal as compared with the full information benchmark.*

I think that the scope of application of Theorem 1 is significant given that in many applications it would seem reasonable to believe that the state varies over more dimensions than the action. In the leading example of Section 3, the state included as many as five dimensions whereas the action space was one-dimensional, the action taking the form of an effort level. Thus, the condition of Theorem 1 is clearly satisfied in this class of problems. Moving into the territory of multi-tasking in which the dimension of the action space would be greater, one should have in mind that the type of signals that the Principal observes would then be of greater richness and also that the state would include descriptions of how complementary the tasks of the Agent are to the working of the organization (about which the Principal could get noisy signals), thereby suggesting that $m > n$ is still a good assumption in applications of this sort.

In some applications, it may be hard to not let the Agent know about some aspects of the state. For example, if the state includes a description of the private costs incurred by the Agent when exerting a certain type of effort, it would seem natural to assume that the Agent knows about this no matter what disclosure policy is chosen by the Principal. With this constraint in mind, one would reinterpret Theorem 1 as follows: some form of non-transparency is desirable, as long as the number of dimensions of the state that can be hidden from the Agent is strictly larger than the number of dimensions of the action to be chosen by the Agent. I believe such a dimensional gap between what can be hidden from the Agent about the state and the action of the Agent is quite

11. Considering the bundling of θ_1 or $\theta_1 - \varepsilon\delta$ as compared with the fully transparent benchmark would now have an effect of the same sign as the bundling of θ_1 or $\theta_1 + \varepsilon\delta$, and as a matter of fact, it may well be then that full transparency dominates as illustrated in the quadratic example of Crawford–Sobel.

common in applications, and the leading example of the article provides a natural illustration of this.

Comment. To illustrate why the non-transparency conclusion of Theorem 1 does not hold whenever the action space and the state space have the same dimension, one can consider a specific case of the above moral hazard example and assume that the only parameter that varies is the sensitivity of success to effort as measured by γ .¹² The desirability of transparency in this case is economically intuitive. If the only heterogeneity lies in the sensitivity γ of success to effort, the Principal would like the highest effort to be made when success is easier to achieve, which is most efficiently obtained when the Agent knows more about the state.

5. EXTENSIONS

In this section, I establish the robustness of the non-transparency insight derived in Section 4 in two directions. First, I consider the case of several agents. Secondly, I allow for the possibility of using optimally designed monetary instruments. In both extensions, I observe that the non-transparency insight holds, as long as the dimension of what can be hidden from the agent(s) exceeds the dimension of the action(s). Thanks to these extensions, I believe the scope of application of the main non-transparency insight of this article is even larger.

5.1. Multi-agent interactions

Compared to the model described in Section 2, I allow for the presence of several agents $i \in I$. Each agent i must now simultaneously choose an action $a_i \in A_i$ where A_i is an open subset of \mathbb{R}^{n_i} . The state is still denoted by θ and it belongs to \mathbb{R}^m . When the profile of actions is $a = (a_i)_i$ and the state is θ , the utility derived by agent i is denoted by $u_i(a; \theta)$, and the pay-off to the Principal is denoted by $\pi(a; \theta)$. All utility functions are assumed to be smooth (*i.e.* twice continuously differentiable) functions of the relevant variables, and for each agent i , u_i is assumed to be concave in a_i .

In the fully transparent benchmark, for each state θ , the play of agents is described by a Nash equilibrium of the corresponding normal form game. I denote by $a^{NE}(\theta) = (a_i^{NE}(\theta))_{i \in I}$ the action profile that would be played in state θ . I will assume that in some open neighbourhood of states, $a^{NE}(\theta)$ varies smoothly (*i.e.* in a twice continuously differentiable fashion) with θ , and in the rest of this Section I am assuming the state space is included in such a neighbourhood.¹³

As in Section 4.1, I will assume in the following Proposition that there are two distinct states θ_1 and θ_2 such that $a_i^{NE}(\theta_1) = a_i^{NE}(\theta_2)$ for some agent i . Consider the Bayesian game in which agent i does not know whether $\theta = \theta_1$ or θ_2 , whereas agents other than i know the state. One Bayes Nash equilibrium of this Bayesian game would result in the same distribution of action profiles as in the complete information benchmark. Also, thanks to the smoothness assumptions, if agent i is uninformed as to whether $\theta = \theta_1$ or $\theta_2 + \varepsilon \delta$ for some direction δ and some scalar ε , there will exist a Bayes Nash equilibrium of the corresponding Bayesian game that varies smoothly with ε and such that it converges to the complete information Nash equilibrium a^{NE} as ε goes to 0. When considering Bayesian games, I make the assumption that the play follows such a Bayes Nash equilibrium.

12. It is readily verified that full transparency dominates any form of non-transparency using Jensen's inequality (and the convexity of $x \rightarrow x^2$).

13. For every θ , an equilibrium in pure strategies exists thanks to our concavity assumption. From the Nash equilibrium correspondence, I am considering here a selection that varies smoothly with θ (which can be done thanks to our smoothness assumptions at least locally using again the techniques of topology from the differentiable viewpoint, Milnor 1965).

Proposition 2. *Assume that there exist two distinct states θ_1, θ_2 in the interior of the θ -space such that the same action would be chosen by at least one agent i in the complete information benchmark, i.e. $a_i^{NE}(\theta_1) = a_i^{NE}(\theta_2)$, and that the utility function u_i of agent i is non-satiated.¹⁴ There exists a generic set Π such that for all objective functions $\pi \in \Pi$, the Principal can do strictly better than disclosing fully the state θ to the agents.*

A first observation is that it is not possible to infer Proposition 2 from Proposition 1 simply by specifying exogenously how agents $j \neq i$ chose their actions $a_j(\theta)$ as a function of the state θ . The reason why such an approach would fail is that the actions $a_j(\theta)$ would typically depend on the information structure of agent i in a Bayes Nash equilibrium and thus $a_j(\theta)$ cannot be held fixed (while changing the information structure of agent i).

Despite this, the argument to prove Proposition 2 is very similar to that used to prove Proposition 1. Consider the incomplete information setting in which agents other than i know the state, and agent i does not know whether $\theta = \theta_1$ or $\theta_2 + \varepsilon\delta$ where δ is a direction such that $\frac{\partial}{\partial \delta} \nabla_{a_i} u_i(a_i^{NE}(\theta_2); \theta_2) \neq 0$. The key step consists in showing that the difference of the Principal's expected pay-off in this incomplete information setting and in the complete information benchmark is of the same order as ε for generic π functions, and thus either $\varepsilon > 0$ or $\varepsilon < 0$ but small would ensure the strict superiority of the incomplete information scenario. The argument is detailed in the Appendix.

As in the one agent case, if the dimension m of the state θ is strictly bigger than the dimension n_i of the action a_i of at least one agent i , then the conditions for Proposition 2 will be satisfied for this agent, and thus some form of non-transparency will be desirable.

In the arguments just developed, I have allowed the Principal to use a different disclosure policy for the various agents. If the disclosure policy must be the same for all agents, then the same conclusion as that in Proposition 2 can be reached, as long as there exist two distinct states θ_1, θ_2 such that the same action profile would be chosen at the two states in the complete information benchmark, i.e. $a^{NE}(\theta_1) = a^{NE}(\theta_2)$. As a corollary, I can infer that some form of non-transparency is desirable in the public information disclosure case, as long as the dimension m of θ is strictly larger than the sum $\sum_i n_i$ of dimensions of a_i over all agents $i \in I$.

5.2. Monetary instruments

Monetary instruments can be used in several ways. First, one can think of the Principal who possesses private information on θ as being able to sell his/her information to the Agent. The difference with the previous analysis is that now the Principal is not constrained to disclose freely his/her information if he/she chooses to be transparent. She can disclose this in exchange for side-payments. Secondly, in the tradition of contract theory, monetary instruments can be used to better align the incentives of the Agent with the interest of the Principal making the side-payments contingent on what the Principal observes *ex post*. In both cases, it is of interest to review what happens to the main non-transparency result derived above.

5.2.1. The sale of information. A simple way to model that information can be sold is to assume that some bargaining between the Principal and the Agent takes place at the *ex ante* stage before the realization of θ is known where bargaining bears on which disclosure policy the Principal will commit to in exchange for side-payments.

14. Non-satiation is defined here with respect to a_i , i.e. $\forall \theta, \exists \delta$ such that $\frac{\partial}{\partial \delta} \nabla_{a_i} u_i(a_i^{NE}(\theta); \theta) \neq 0$.

It is not difficult to see that the non-transparency result as described in Proposition 1 or Theorem 1 continues to hold in such a setting no matter what the exact details of the bargaining protocol are.

To see this most simply, assume that the utilities of the Principal and the Agent are quasi-linear in money (where the dependences of u and π with respect to a and θ are still assumed to be of the most general form). Then, bargaining necessarily results in the maximization of $u + \pi$, and it is not difficult to infer from Proposition 1 that, as long as the utility function u of the Agent is non-satiated and there are two distinct states θ_1, θ_2 such that the same action would be chosen in the complete information benchmark, *i.e.* $a(\theta_1) = a(\theta_2)$, then full transparency is not the optimal disclosure policy for generic objective functions π (because genericity of $\pi + u$ as confined to the set of preferences that are quasi-linear in money immediately translates into genericity of π viewed as a function of (a, θ)). The same conclusion would also hold true in the case of non-quasi-linear utilities even though the argument would now have to make the bargaining protocol more precise, as it may now have an impact on which disclosure policy would be adopted.

It is worth contrasting the above argument with the result obtained by Eso and Szentes (2007) who show that it is optimal for the seller of a good to disclose as much as he/she can to potential buyers in a context of private value auctions with quasi-linear preferences. Assuming away the private information of the buyers, it is readily verified that in the context studied by Eso and Szentes, the welfare-maximizing outcome can be achieved whenever the seller discloses all what he/she knows, thereby explaining Eso–Szentes' insight in terms of the optimal disclosure policy. Yet, in general moral hazard problems, even assuming utilities are quasi-linear in money, there is no reason why full disclosure would result in the maximization of total welfare, and thus the full transparency insight of Eso and Szentes does not extend to such contexts.

5.2.2. Side-payments as incentive instruments. A full description of the state θ should in general include a specification of what the Principal observes *ex post* and how this is affected by the action a of the Agent. To simplify, assume that the Principal can only observe a finite number s of signals $\sigma = 1, \dots, s$. Then the monetary instrument available to the Principal can be described as a vector $w = (w_\sigma)_\sigma \in \mathbb{R}_+^s$ specifying the side-payment assumed to be non-negative $w_\sigma \in \mathbb{R}_+$ the Agent would receive from the Principal in case the Principal observes σ *ex post*. In state θ , action a would result in a probability distribution over σ that I denote by $q_\theta(\cdot | a)$, and I assume that for all θ, σ , and a , $q_\theta(\sigma | a) > 0$.¹⁵ Agent's expected pay-off is $u(a; w, \theta)$ when w prevails, the state is θ , and the Agent chooses a . The corresponding expected pay-off to the Principal is denoted by $\pi(a; w, \theta)$, and it is assumed to be decreasing in w . All functions are assumed to vary smoothly (*i.e.* in a twice continuously differentiable fashion) with a, w, θ . Besides u is assumed to be a concave function of a for all w and θ . Moreover, I assume that the function $a \rightarrow u(a; w, \theta)$ is always maximized in a bounded subset of A ,¹⁶ and I denote the maximand by $a(w, \theta)$. Finally, I denote by $w(\theta)$ the bonus scheme that maximizes $\pi(a(w, \theta); w, \theta)$.

I rephrase the transparency question for the case with monetary transfers as follows:

Question. Can it be beneficial for the Principal that the Agent be partially rather than fully informed of (θ, w) ?

15. This full support assumption typically implies that the first-best cannot be achieved in the full information benchmark.

16. Such assumptions will typically guarantee that one can work with first-order conditions to deal with the maximization problems of agents. To the extent that interior pure strategy equilibria exist and that one focuses on these, the insights developed below would apply.

Compared to the case without monetary transfers, transparency is defined as a situation in which the Agent knows the state θ as well as the exact incentive scheme as defined by w that prevails in θ . Observe that if the Principal were to choose a stochastic incentive scheme (a distribution over w) and have the Agent know the state θ (but not the realization of w), this would be considered as a form of non-transparency in the context of the above question (even if I conjecture that a non-transparency result solely in terms of partial disclosure of the state θ is likely to hold even in cases in which stochastic contracts are optimal).

The notion of genericity is the same as that introduced in Section 4 except that the domain of π is now $X = \mathbb{R}^n \times \mathbb{R}^s \times \mathbb{R}^m$ given that π depends now also on the monetary instruments $w \in \mathbb{R}^s$. The non-transparency result can be stated as follows.

Proposition 3. *Assume that there exist two distinct states θ_1, θ_2 in the interior of the θ -space such that the same action would be chosen in the complete information benchmark, i.e. $a(w(\theta_1), \theta_1) = a(w(\theta_2), \theta_2)$, and that the utility function u is non-satiated. There exists a generic set Π such that for all objective functions $\pi \in \Pi$, the Principal can do strictly better than disclosing fully the state θ to the Agent.*

The idea of the proof is very similar to that of Proposition 1. For each θ , define $\tilde{u}(a; \theta) = u(a, w(\theta); \theta)$, and $\tilde{\pi}(a; \theta) = \pi(a, w(\theta); \theta)$. The genericity of π translates into the genericity of $\tilde{\pi}$, and one can apply Proposition 1 to derive Proposition 3. The economic intuition for this result is that even if the monetary instrument w is optimally adjusted to the state θ , $u(a, w; \theta)$ is not perfectly aligned with $\pi(a, w; \theta)$ and thus the Principal can gain by using a well-chosen non-transparent disclosure policy.¹⁷

Of course, similarly to Theorem 1, I also observe that as soon as the dimension of the state space m is larger than the dimension of the action space n , i.e. $m > n$, one must have two distinct states θ_1 and θ_2 such that $a(w(\theta_1), \theta_1) = a(w(\theta_2), \theta_2)$ and thus some form of non-transparency is desirable.

5.2.3. Discussion. (i) The general issue as to whether transparency is desirable can be formulated in terms of the following trade-off. If an Agent receives incomplete information about the state θ , it makes it easier to satisfy the incentive constraints of the Agent through an appropriate use of the monetary instruments. This is because incomplete information allows the Principal to aggregate the various Agent's incentive constraints (for each state) into one incentive constraint that is easier to satisfy. But, incomplete information forces the Agent not to be able to adjust his/her action to the state θ , which is sometimes harmful to the Principal. In light of this, what Proposition 3 establishes is that one can always find information structures in which the first effect dominates the second, as soon as there are at least two distinct states in which the action would be the same in the case in which the Agent knows the state.¹⁸

(ii) The above result about the desirability of some form of non-transparency in the presence of monetary instruments seems to be at odds with the conclusion obtained by Rayo and Segal (2010) who show in their set-up in which the state space has dimension 2 and the action

17. Based on this intuition, I would expect the result of Proposition 3 to continue to hold even if the observation made by the Principal were to take values on the continuum, but the formal argument would require dealing with first-order conditions in the space of bonus functions which would be technically more challenging (as compared with dealing with local optima over finitely many bonuses for each possible observation).

18. Observe that in the derivation of this result, it is not the case that the first effect is made of an order greater than the second (the two effects appear to be of comparable magnitude, but for well-chosen information structures, the former dominates the latter).

space has dimension 1 that when monetary instruments are allowed, the best disclosure policy is full transparency. They get at this conclusion by observing that any outcome with partial disclosure (and possibly monetary incentives) can be replicated with full disclosure and well-chosen monetary instruments. Yet, the conclusion that full disclosure of θ is optimal can only hold for special (non-generic) specifications of the Principal's objective function π , as shown in Proposition 3. Specifically, perturbing the Principal's objective with respect to a and θ alone while maintaining the quasi-linearity assumption regarding money in Rayo–Segal's model would lead the Principal to find full transparency suboptimal,¹⁹ even if he/she can incentivize the Agent through well-chosen side-payments.

(iii) In the above analysis, I have implicitly ignored agents' participation constraints. Consider now imposing that an agent should get at least his/her outside option pay-off. Clearly, nothing changes if participation constraints are not binding.²⁰ For example, in contexts with limited liability, an agent typically receives a positive rent in moral hazard problems, and the participation constraints are not binding. In the absence of limited liability constraints though, the Principal would typically adjust the instruments w so that the Agent gets his/her outside option pay-off in pure moral hazard problems (see Holmström (1979–1982) or Holmström–Milgrom (1991) in the context of risk-averse agents without limited liability constraints). It should be noted, however, that if in addition to the moral hazard problem, the Agent was assumed to possess some private information then most "types" of the Agent would receive positive rent even in the absence of limited liability constraints. Although dealing with such environments requires further analysis, I conjecture that the above non-transparency result would still hold in this case.

(iv) In the context of optimal mechanisms, the most general form of mechanisms allows for the presence of a mediator who could make recommendations to the Agent as to which action to choose. When such mechanisms are available, one can always implement the optimal mechanism by having the Agent be only informed of what to do (which action a to choose) (Myerson 1982). When several states induce the same action as considered in the above Proposition, Myerson's result implies that it is enough to let the Agent know what to do and that there is no gain in letting him know more. From that perspective, what the above non-transparency result shows is the stronger property that under the conditions of the above Proposition, the Principal can strictly improve upon what he/she can achieve in the fully transparent case (which Myerson's result does not imply).

(v) Gjesdal (1982) was the first to note that stochastic contracts may sometimes help the Principal achieve larger expected pay-offs. His/Her result though is of a very different nature than the one derived above as it does not rely on variations of the state.²¹

19. Inspecting the Appendix reveals that the perturbations of π used to prove Proposition 3 allow me to keep $\frac{\partial \pi}{\partial w}$ unchanged so that the genericity required is about how π depends on a and θ (rather than w).

20. If the participation constraints are binding both at $\theta = \theta_1$ and $\theta_2 + \varepsilon\delta$ in the main argument used to prove the above Proposition when w is set at $w(\theta)$ in state θ , one has to worry that the Agent gets no less than his/her outside option pay-off when the Agent does not know whether $\theta = \theta_1$ or $\theta_2 + \varepsilon\delta$, which may require increasing the burden to the Principal.

21. Roughly, Gjesdal's argument can be illustrated as follows. The reference to the state θ will be omitted from the Agent and the Principal's utilities, since Gjesdal considers only one state. When the Agent chooses a and the Principal chooses w , I let $u(a, w)$ denote the utility of the Agent and I assume that the pay-off of the designer takes the linear form: $\pi(a, w) = a - w$.

That is, the Principal wishes to induce a high effort level a from the agent, and he/she can affect the choice of a using monetary instrument w at some cost assumed to be linear in w .

Gjesdal's main observation is that if $\frac{\partial u}{\partial a}(a, w)$ is convex in w , then a stochastic contract increases the expected utility of the Principal. To see this most simply, assume that instead of using a deterministic w , the Principal chooses a stochastic contract that assigns probability $\frac{1}{2}$ to $w_1 = w - \varepsilon$ and probability $\frac{1}{2}$ to $w_2 = w + \varepsilon$. For any given ε , the first-order condition

6. ON THE ART OF STAYING SILENT

The above analysis has shown that full transparency is suboptimal whenever the same action would be chosen by the Agent at two distinct states, which in particular implies that full transparency can be improved upon whenever the state space has a dimensionality greater than the action space.

In this section, I particularize the problem and I assume that the Principal can either disclose fully the state or remain silent. I characterize the optimal way for the Principal to remain silent in the class (P) of problems introduced at the end of Section 3.

Specifically, assume that

$$u(a; \theta) = xa - y \frac{a^2}{2} + K$$

$$\pi(a; \theta) = za + K'$$

Since K and K' play no role in the comparison of the various disclosure policies, from now on, I can assume without loss of generality that $K = K' = 0$ and I identify θ with (x, y, z) .

If the Agent is not told anything whenever $\theta \in \Theta^*$, the Agent picks action $a = \frac{E(x|\theta \in \Theta^*)}{E(y|\theta \in \Theta^*)}$ in the case the Principal remains silent. Conditional on $\theta \in \Theta^*$, the Principal thus gets in expectation $E(z|\theta \in \Theta^*) \frac{E(x|\theta \in \Theta^*)}{E(y|\theta \in \Theta^*)}$ and compared to the fully transparent benchmark in which the state θ is always disclosed, the Principal gets an extra (possibly negative) expected pay-off of

$$H(\Theta^*) = \Pr(\theta \in \Theta^*) [E(z|\theta \in \Theta^*) \cdot \frac{E(x|\theta \in \Theta^*)}{E(y|\theta \in \Theta^*)} - E(z \cdot \frac{x}{y} | \theta \in \Theta^*)]$$

The best silence set is the set Θ^* that maximizes $H(\Theta^*)$.

Considering the marginal effect of the addition (or subtraction) of some small neighbourhood of $\theta = (x, y, z)$ to (or from) Θ^* yields the following characterization result:

Proposition 4. *The optimal set of states Θ^* over which the Principal should remain silent is of the form:*

$$\Theta^* = \{\theta = (x, y, z) \text{ such that } (\frac{z}{y} - \frac{z^e}{y^e}) \cdot (\frac{x}{y} - \frac{x^e}{y^e}) \leq 0\}$$

that determines the Agent's action is:

$$\frac{1}{2} \frac{\partial u}{\partial a}(a(\varepsilon), w + \varepsilon) + \frac{1}{2} \frac{\partial u}{\partial a}(a(\varepsilon), w - \varepsilon) = 0 \quad (4)$$

Total differentiation of (4) with respect to ε and simple Taylor expansions yield:

$$\frac{da}{d\varepsilon}(\varepsilon) = - \frac{\frac{\partial^3 u}{\partial a \partial w^2}(a(0), w)}{\frac{\partial^2 u}{\partial a^2}(a(0), w)} \varepsilon + o(\varepsilon)$$

where $\frac{o(\varepsilon)}{\varepsilon} \rightarrow 0$ as $\varepsilon \rightarrow 0$. Thus, $\frac{da}{d\varepsilon}(\varepsilon) > 0$ for ε small enough whenever $\frac{\partial^3 u}{\partial a \partial w^2}(a(0), w) > 0$ (given that $\frac{\partial^2 u}{\partial a^2}(a(0), w) < 0$). This implies that when $\frac{\partial u}{\partial a}(a, w)$ is convex in w , the proposed stochastic contract does strictly better than the original deterministic contract, since the expected monetary transfer is the same in the two cases and the effort level $a(\varepsilon)$ is (slightly) bigger in the stochastic case than in the deterministic case.

Thus, when $\frac{\partial u}{\partial a}(a(0), w)$ is convex in w , any deterministic contract can be improved upon by using a stochastic contract that is a (sufficiently small) mean preserving spread of the original deterministic contract.

Of course, when $\frac{\partial u}{\partial a}(a(0), w)$ is not convex in w , it may be that no stochastic contract dominates the best deterministic contract. So Gjesdal's insight is not that in general optimal contracts are stochastic, but rather than under some convexity conditions they may be stochastic. This is, of course, very different from our non-transparency insight stipulating that when θ has dimension bigger than a , it is always best not to let the agent know perfectly (θ, w) .

where

$$x^e = E(x | \theta \in \Theta^*)$$

$$y^e = E(y | \theta \in \Theta^*)$$

$$z^e = E(z | \theta \in \Theta^*)$$

To see the shape of Θ^* , it is instructive to particularize Proposition 4 in different ways.

6.1. Optimal silence set in the (x, z) -space

First, consider the case in which y is set at 1. Θ^* is then of the form $\{\theta = (x, z) \text{ such that } (z - z^e) \cdot (x - x^e) \leq 0\}$. That is, the Principal should remain silent either when $x < x^e$ and $z > z^e$ or when $x > x^e$ and $z < z^e$. The corresponding area is depicted in grey in Figure 1.

The intuition for the shape of Θ^* in this case is that the bundling of states $\Theta^{NW} = \{\theta = (x, z) \text{ such that } x < x^e \text{ and } z > z^e\}$ with states $\Theta^{SE} = \{\theta = (x, z) \text{ such that } x > x^e \text{ and } z < z^e\}$ is good for the Principal because it allows his/her to transfer the comparatively high effort made in Θ^{SE} (this comparison can be inferred from x being larger in Θ^{SE} than in Θ^{NW}) to Θ^{NW} in which the Principal cares comparatively more about effort than in Θ^{SE} (this comparison can be inferred from z being larger in Θ^{NW} than in Θ^{SE}). So here it is the relative sensitivity to effort of the Principal's objective π in the various states that explains the shape of the optimal silence set.

I have already noted that when $y = 1$, the above specification of pay-offs results in the same reduced form specification as the one considered by Rayo and Segal. The shape of the optimal silence set in this case is somehow reminiscent of Rayo and Segal's observation that the bundling of two states (x_1, z_1) and (x_2, z_2) into one information set is profitable whenever the two states are unordered (*i.e.* if $x_1 < x_2$ then $z_1 > z_2$). Yet, the question investigated here is not the bundling of two states but rather over which set of states (which are many more than two) the Principal should remain silent, and Proposition 4 establishes a precise sense in which x and z should be negatively correlated in the optimal silence set.

From the viewpoint of the moral hazard scenario considered in Section 3, assume the heterogeneity lies only in the quality of the monitoring as measured by β as well as the sensitivity

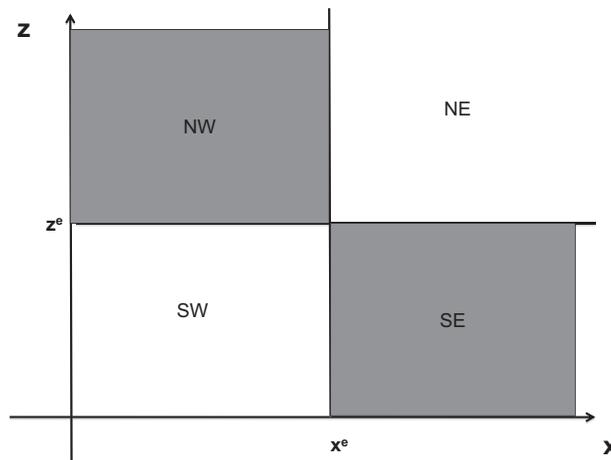


FIGURE 1

Optimal silence set in the (x, z) -space.

of success to effort as measured by γ . Phrased in the (x, y, z) -space, this is a situation in which y can be held fixed. The naive approach in this application would suggest remaining silent whenever the monitoring technology is of bad quality as suggested in the Introduction. In the (x, z) -space, this would amount to remaining silent whenever $x - z$ is above some threshold. As can be seen from Proposition 4 or Figure 1, the optimal silence set is not of this form because if the Principal were to follow such a (naive) disclosure policy, the Agent would infer from the Principal remaining silent that the monitoring technology must be poor and the absence of communication on the state would then not be profitable. By contrast, what emerges as the optimal silence set suggests a form of negative correlation between the monitoring technology and the productivity of effort as can be inferred from Figure 1 and how x and z relate to β and γ (see the end of Section 3).

6.2. *Optimal silence set in the (x, y) -space*

I next consider the case in which z is set at 1. Θ^* is then of the form $\{\theta = (x, y) \text{ such that } (\frac{1}{y} - \frac{1}{y^e}) \cdot (\frac{x}{y} - \frac{x^e}{y^e}) \leq 0\}$. That is, the Principal should remain silent either when $y < y^e$ and $\frac{x}{y} < \frac{x^e}{y^e}$ or when $y > y^e$ and $\frac{x}{y} > \frac{x^e}{y^e}$. The corresponding area in the $(\frac{x}{y}, y)$ space is depicted in grey in Figure 2.

The intuition for the shape of Θ^* in this case is that the bundling of states $\Theta^{SW} = \{\theta = (x, y) \text{ such that } y < y^e \text{ and } \frac{x}{y} < \frac{x^e}{y^e}\}$ with states $\Theta^{NE} = \{\theta = (x, y) \text{ such that } y > y^e \text{ and } \frac{x}{y} > \frac{x^e}{y^e}\}$ is good for the Principal because it bundles high effort states in Θ^{NE} with low effort states in Θ^{SW} ($\frac{x}{y}$ is larger in Θ^{NE} than in Θ^{SW}) and the Agent's utility is more concave at states in Θ^{NE} than at states in Θ^{SW} (since y is larger in Θ^{NE} than in Θ^{SW}). So here it is the relative concavity in effort of the Agent's utility at the various states that explains the shape of the optimal silence set.

When the heterogeneity lies only in (x, y) then some form of positive correlation between $\frac{x}{y}$ and y is desirable for the Principal in the precise sense depicted in Figure 2. For example, in the context of the moral hazard example introduced in Section 3, if the heterogeneity lies only in the private benefit b and the decreasing returns to effort in the task as measured by λ , then the optimal silence requires a form of positive correlation between b and λ .

6.2.1. Discussion. The above investigation can be looked at through the lens of how tasks should be described in organizations so that agents make the best use of their time in the

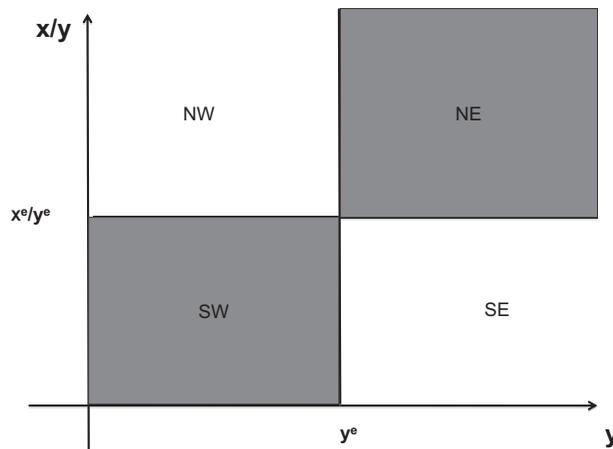


FIGURE 2
Optimal silence set in the (x, z) -space.

various tasks. Indeed agents are typically involved in lots of tasks in organizations and how the tasks are described will typically affect how agents allocate their time and effort across tasks possibly including some training phases before the tasks must be fulfilled. Such a problem fits in well with the above investigation if one has in mind that the cost of time in a given task must include the opportunity cost of not being able to use extra time in other tasks. The present section has considered when it is desirable not to describe finely the task to the agent (the analog of remaining silent), and the above derived insights can be used to shed light on this and also on the breadth of task assignment.

For example, suppose an employee of an organization must do some applied development as well as pure research as well as some training of other employees. Within the class of situations considered in this section, these various tasks may correspond to different realizations of the state θ , but to fix ideas assume the main lines of differentiation of tasks lie in how good the monitoring technology is as measured by β and how productive the effort is as measured by γ . Proposition 4 would suggest to not specify the task whenever the monitoring technology is worst and productivity is highest (e.g. as is the case with pure research) or when the monitoring technology is very good and productivity very low (e.g. as is the case with training) while describing the task more finely in other cases (e.g. as here with applied development). The absence of clear definition of the task when faced with (pure) research jobs (which are typically bundled with training/teaching elements) seems in line with this finding.²²

7. CONCLUSION

In this article, I have shown that some form of non-transparency is desirable, as long as in the complete information benchmark, the same actions would be chosen by the agent(s) for at least two distinct states. Based on this result, I would suggest that some form of non-transparency is optimal quite generally, given that the state is likely to vary over more dimensions than the action in most applications. I have also characterized the best way to remain silent for a class of problems, suggesting how the insights obtained could be used to shed light on when tasks should not be described finely in organizations. Clearly, more work should be devoted to understanding the best disclosure policy in more general settings, in particular, how it interacts with the best use of monetary instruments.

APPENDIX

Proof of Proposition 1. I consider the case in which the action a may have arbitrary many dimensions. Let θ_1 and θ_2 be two states such that in the complete information case $a(\theta_1) = a(\theta_2) = a_1$ and consider a direction δ and ε small enough such that $a(\theta_2 + \varepsilon\delta) \neq a(\theta_2)$.

Let $a_2(\varepsilon)$ and $a^{CI}(\varepsilon)$ be defined by

$$\begin{cases} \nabla_a u(a_1; \theta_1) = 0 \\ \nabla_a u(a_2(\varepsilon); \theta_2 + \varepsilon\delta) = 0 \\ p(\theta_1)\nabla_a u(a^{CI}(\varepsilon); \theta_1) + p(\theta_2 + \varepsilon\delta)\nabla_a u(a^{CI}(\varepsilon); \theta_2 + \varepsilon\delta) = 0 \end{cases} \quad (\text{A.1})$$

Let $J_u(\theta)$ denote the Jacobian of u at state θ . Total differentiation of (A.1) yields

$$\begin{aligned} \frac{da_2}{d\varepsilon}(\varepsilon = 0) &= -(J_u(\theta_2))^{-1} \cdot \frac{\partial}{\partial \delta} \nabla_a u(\theta_2) \\ \frac{da^{CI}}{d\varepsilon}(\varepsilon = 0) &= -p(\theta_2)[p(\theta_1)J_u(\theta_1) + p(\theta_2)J_u(\theta_2)]^{-1} \cdot \frac{\partial}{\partial \delta} \nabla_a u(\theta_2) \end{aligned}$$

22. I am indebted to the editor for suggesting this interpretation.

Let $\Delta(\varepsilon)$ denote the expected organizational gain of not letting the Agent know $\theta = \theta_1$ or $\theta_2 + \varepsilon\delta$ relative to the full information case. We have:

$$\frac{\Delta(\varepsilon)}{\varepsilon} = -p(\theta_1)\nabla_a\pi(a_1; \theta_1) \cdot \frac{da^{CI}}{d\varepsilon}(\varepsilon=0) + p(\theta_2)\nabla_a\pi(a_1; \theta_2) \cdot \left[\frac{da_2}{d\varepsilon}(\varepsilon=0) - \frac{da^{CI}}{d\varepsilon}(\varepsilon=0) \right] + o(\varepsilon)$$

where $o(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$.

One can conclude as in the case in which a is one-dimensional by noting that for generic π functions

$$-p(\theta_1)\nabla_a\pi(a_1; \theta_1) \cdot \frac{da^{CI}}{d\varepsilon}(\varepsilon=0) + p(\theta_2)\nabla_a\pi(a_1; \theta_2) \cdot \left[\frac{da_2}{d\varepsilon}(\varepsilon=0) - \frac{da^{CI}}{d\varepsilon}(\varepsilon=0) \right] \neq 0 \quad (\text{A.2})$$

which can be seen by considering the family of Principal's objective

$$\pi_\lambda(a; \theta) = \pi(a; \theta) + \lambda \|\theta - \theta_2\| a \cdot \frac{da^{CI}}{d\varepsilon}(\varepsilon=0)$$

If $\pi(a; \theta)$ does not satisfy (A.2), $\pi_\lambda(a; \theta)$ will for small non-zero values of λ .

So we may now assume that not letting the Agent know whether $\theta = \theta_1$ or $\theta_2 + \varepsilon\delta$ strictly dominates the full information case. Consider a small η and the effect of not letting the Agent know the realization of θ in $B_\eta(\theta_1) \cup B_\eta(\theta_2 + \varepsilon\delta)$ as compared with the full information benchmark where $B_\eta(\theta)$ denotes the set of states which are at (euclidean) distance no more than η from θ . Continuity of π and u implies that conditional on $\theta \in B_\eta(\theta_1) \cup B_\eta(\theta_2 + \varepsilon\delta)$, this difference converges to $\Delta(\varepsilon)/(p(\theta_1) + p(\theta_2 + \varepsilon\delta))$ as η converges to 0 where $\Delta(\varepsilon)$ is the function described above. It follows that there must exist η small enough so that the Principal is strictly better off not letting the Agent know $\theta \in B_\eta(\theta_1) \cup B_\eta(\theta_2 + \varepsilon\delta)$ as compared with the full information benchmark. \parallel

Proof of Proposition 2. To present the argument in a more reader-friendly way, I restrict myself to the case in which the action space has dimension 1 and there are two agents i and j .

For any θ , $a^{NE}(\theta)$ must solve:

$$\begin{cases} \frac{\partial u_i}{\partial a_i}(a_i, a_j; \theta) = 0 \\ \frac{\partial u_j}{\partial a_j}(a_i, a_j; \theta) = 0 \end{cases}$$

Consider θ_1 and θ_2 such that $a_i^{NE}(\theta_1) = a_i^{NE}(\theta_2)$ and consider a direction δ of θ such that for ε small enough $a^{NE}(\theta_2 + \varepsilon\delta) \neq a^{NE}(\theta_1)$.

If agent i does not know whether $\theta = \theta_1$ or $\theta_2 + \varepsilon\delta$, Nash equilibrium (NE) actions $a_i^C(\varepsilon)$, $a_{j,1}^C(\varepsilon)$ and $a_{j,2}^C(\varepsilon)$ are given by:

$$\begin{cases} \frac{\partial u_i}{\partial a_j}(a_i^C(\varepsilon), a_{j,1}^C(\varepsilon); \theta_1) = 0 \\ \frac{\partial u_j}{\partial a_j}(a_i^C(\varepsilon), a_{j,2}^C(\varepsilon); \theta_2 + \varepsilon\delta) = 0 \\ p(\theta_1) \frac{\partial u_i}{\partial a_i}(a_i^C(\varepsilon), a_{j,1}^C(\varepsilon); \theta_1) + p(\theta_2 + \varepsilon\delta) \frac{\partial u_i}{\partial a_i}(a_i^C(\varepsilon), a_{j,2}^C(\varepsilon); \theta_2 + \varepsilon\delta) = 0 \end{cases}$$

And if there is full information, NE actions $a_{i,1}$, $a_{i,2}(\varepsilon)$, $a_{j,1}$ and $a_{j,2}(\varepsilon)$ are given by:

$$\begin{cases} \frac{\partial u_j}{\partial a_j}(a_{i,1}, a_{j,1}; \theta_1) = 0 \\ \frac{\partial u_j}{\partial a_j}(a_{i,2}(\varepsilon), a_{j,2}(\varepsilon); \theta_2 + \varepsilon\delta) = 0 \\ \frac{\partial u_i}{\partial a_i}(a_{i,1}, a_{j,1}; \theta_1) = 0 \\ \frac{\partial u_i}{\partial a_i}(a_{i,2}(\varepsilon), a_{j,2}(\varepsilon); \theta_2 + \varepsilon\delta) = 0 \end{cases}$$

I expand at order 1 in ε (the diff. of π in coarse vs full info)

$$\begin{aligned} \Delta(\varepsilon) &= p(\theta_1)[\pi(a_i^C(\varepsilon), a_{j,1}^C(\varepsilon); \theta_1) - \pi(a_{i,1}, a_{j,1}; \theta_1)] + \\ &\quad p(\theta_2 + \varepsilon\delta)[\pi(a_i^C(\varepsilon), a_{j,2}^C(\varepsilon); \theta_2 + \varepsilon\delta)] - \pi(a_{i,2}(\varepsilon), a_{j,2}(\varepsilon); \theta_2 + \varepsilon\delta) \end{aligned}$$

Similarly to the one agent case if $\Delta'(0) \neq 0$, then it implies that not letting agent i know whether $\theta = \theta_1$ or $\theta_2 + \varepsilon\delta$ with $\varepsilon > 0$ or $\varepsilon < 0$ but small strictly improves over the full information benchmark and $\Delta'(0) = 0$ can be shown to be non-generic by considering perturbations of the form $\pi_\lambda(a_i, a_j; \theta) = \pi(a_i, a_j; \theta) + \lambda a_i \|\theta - \theta_1\|^2$ where $\|\cdot\|$ denotes the Euclidean distance. \parallel

Proof of Proposition 3. To make the argument more reader-friendly, assume the action of the Agent a varies over one dimension and let w_σ denote the bonus obtained by the Agent when the Principal observes σ *ex post*. For any θ , there is an optimal $w = (w_\sigma)_\theta$, say $w(\theta)$. It is defined as

$$w(\theta) = \operatorname{argmax}_w \pi(a(w), w; \theta)$$

$$a(w) = \operatorname{argmax}_a u(a, w; \theta)$$

Thus, keeping θ constant on has: $\frac{\partial u}{\partial a}(a, w; \theta) = 0$, which after complete differentiation w.r.t w_σ yields $\frac{\partial^2 u}{\partial a^2} \frac{\partial a}{\partial w_\sigma} + \frac{\partial^2 u}{\partial a \partial w_\sigma} = 0$. The first-order condition on the designer's programme writes $\frac{\partial \pi}{\partial a} \frac{\partial a}{\partial w_\sigma} + \frac{\partial \pi}{\partial w_\sigma} = 0$, which combined with the previous condition yields:

$$-\frac{\partial \pi}{\partial a} \frac{\frac{\partial^2 u}{\partial a \partial w_\sigma}}{\frac{\partial^2 u}{\partial a^2}} + \frac{\partial \pi}{\partial w_\sigma} = 0$$

Define $\tilde{\pi}(a; \theta) = \pi(a, w(\theta); \theta)$ and $\tilde{u}(a; \theta) = u(a, w(\theta); \theta)$. Apply the argument developed around Proposition 1 assuming $\tilde{\pi}$ is the designer's objective and \tilde{u} is the Agent's utility function. Clearly, if not letting the Agent know whether $\theta = \theta_1$ or $\theta_2 + \varepsilon \delta$ strictly dominates the complete information benchmark for this case, then in the case when the designer can choose w , it also strictly dominates (because the designer always has the option to set w to be $w(\theta)$ in state θ).

It remains to show that generically it is not the case that (see (2) above)

$$\frac{\partial \tilde{\pi}}{\partial a}(\theta_1) \frac{\partial^2 \tilde{u}}{\partial a^2}(\theta_2) - \frac{\partial \tilde{\pi}}{\partial a}(\theta_2) \frac{\partial^2 \tilde{u}}{\partial a^2}(\theta_1) = 0 \quad (\text{A.3})$$

To see this, consider the family of π_λ functions

$$\pi_\lambda(a, w; \theta) = \pi(a, w; \theta) + \lambda \|\theta - \theta_1\|^2 \left(a + \sum_\sigma \frac{\partial^2 u / \partial a \partial w_\sigma}{\partial^2 u / \partial a^2}(a_2, w(\theta_2); \theta_2) \cdot w_\sigma \right)$$

where $\lambda \in \mathbb{R}$. For such a family, $w(\theta)$ are the same at $\theta = \theta_1$ (resp. θ_2) whatever λ so that $\frac{\partial \pi_\lambda}{\partial a}(\theta) = \frac{\partial \pi}{\partial a} + \lambda \|\theta - \theta_1\|^2$ for $\theta = \theta_1$ and θ_2 . Thus, if $\tilde{\pi}$ satisfies (A.3), $\tilde{\pi}_\lambda$ does not for any sufficiently small $\lambda \neq 0$, and one can conclude as in the case of Proposition 1. \parallel

Proof of Proposition 4. The proof takes the following route. Consider a candidate set Θ^* over which the Principal is supposed to optimally stay silent. Let $p(\Theta^*)$ be the probability that θ falls in Θ^* . Let $\theta^* = (x^*, y^*, z^*)$ be a state in Θ and assume the probability that $\theta = \theta^*$ is p^* which is assumed to be arbitrarily small as compared with $p(\Theta^*)$.

Assume first that θ^* lies outside Θ^* . Letting $H(S)$ be the expected organizational gain over full transparency when the Principal remains silent over states $\theta \in S$, routine calculations show that $H(\Theta^* \cup \{\theta^*\}) - H(\Theta^*)$ writes at the first order in p^* :

$$-\left[\frac{z^*}{y^*} - \frac{E(z | \theta \in \Theta^*)}{E(y | \theta \in \Theta^*)} \right] \left[\frac{x^*}{y^*} - \frac{E(x | \theta \in \Theta^*)}{E(y | \theta \in \Theta^*)} \right] p^* \quad (\text{A.4})$$

Clearly if (A.4) is strictly positive, the Principal is strictly better off being silent over $\Theta^* \cup \{\theta^*\}$ than over Θ^* and thus Θ^* is not optimal (in the case of smooth densities, replacing $\{\theta^*\}$ by a small neighbourhood of θ^* would do using continuity arguments). This implies that for all $\theta^* \notin \Theta^*$, one should have:

$$\left[\frac{z^*}{y^*} - \frac{E(z | \theta \in \Theta^*)}{E(y | \theta \in \Theta^*)} \right] \left[\frac{x^*}{y^*} - \frac{E(x | \theta \in \Theta^*)}{E(y | \theta \in \Theta^*)} \right] \geq 0$$

Assume next that $\theta^* \in \Theta^*$. Then the same calculation yields that $H(\Theta^*) - H(\Theta^* \setminus \{\theta^*\})$ has the same sign as

$$-\left[\frac{z^*}{y^*} - \frac{E(z | \theta \in \Theta^*)}{E(y | \theta \in \Theta^*)} \right] \left[\frac{x^*}{y^*} - \frac{E(x | \theta \in \Theta^*)}{E(y | \theta \in \Theta^*)} \right]$$

and thus the optimality of Θ^* implies that for every $\theta^* \in \Theta^*$, one should have:

$$\left[\frac{z^*}{y^*} - \frac{E(z | \theta \in \Theta^*)}{E(y | \theta \in \Theta^*)} \right] \left[\frac{x^*}{y^*} - \frac{E(x | \theta \in \Theta^*)}{E(y | \theta \in \Theta^*)} \right] \leq 0$$

It follows that the optimal silence set Θ^* has the form shown in Proposition 4. \parallel

Acknowledgments. Thanks to seminar participants at Cesifo IO workshop, Paris 1, the workshop on multidimensional mechanism design in Bonn, Warwick theory workshop, StonyBrook game theory conference, Queen Mary University workshop, the conference on gaming in Bonn, Sciences Po, and in particular Cloda Jenkins, Ariane Lambert, Laurent Lamy, David Martimort, Paul Milgrom, Andy Newman, Larry Samuelson, and Wataru Tamura for insightful comments. I also thank the European Research Council for financial support, and finally the editor (Marco Ottaviani) and the referees for helpful comments that helped me reshape the article.

REFERENCES

- CRAWFORD, V. and SOBEL J. (1982), "Strategic information transmission", *Econometrica*, **50**, 1431–1451.
 EDERER, F., HOLDEN, R. and MEYER, M. (2008), "Gaming and Strategic Ambiguity in Incentive Provision" (Mimeo).

- ESO, P. and SZENTES, B. (2007), "Optimal Information Disclosure in Auctions", *Review of Economic Studies*, **74**, 705–731.
- GJESDAL, F. (1982), "Information and Incentives: The Agency Information Problem", *Review of Economic Studies*, **49**, 373–390.
- HOLMSTRÖM, B. (1979), "Moral Hazard and Observability", *The Bell Journal of Economics*, **10**, 74–91.
- HOLMSTRÖM, B. (1982), "Moral Hazard in Teams", *The Bell Journal of Economics*, **13**, 324–340.
- HOLMSTRÖM, B. and MILGROM, P. (1991), "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design", *Journal of Law, Economics, and Organization*, **7**, 24–52.
- KAMENICA, E. and GENTZKOW, M. (2011), "Bayesian Persuasion", *American Economic Review*, **101**, 2590–2615.
- MAS-COLELL, A., WHINSTON, M. D. and GREEN, J. R. (1995), *Microeconomic Theory* (New York, Oxford: Oxford University Press).
- MILGROM, P. and WEBER, R. (1982), "A Theory of Auctions and Competitive Bidding", *Econometrica*, **50**, 1089–1122.
- MILNOR, J. W. (1965), *Topology from the Differentiable Viewpoint* (Princeton NJ: Princeton University Press).
- MYERSON, R. (1982), "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems", *Journal of Mathematical Economics*, **10**, 6781.
- OTTAVIANI, M. and PRAT, A. (2001), "The Value of Public Information in Monopoly", *Econometrica*, **69**, 1673–1683.
- RAYO, L. and SEGAL, I. (2010), "Optimal Information Disclosure", *Journal of Political Economy*, **118**, 949–987.
- SOBEL, J. (1993), "Information Control in the Principal-Agent Problem", *International Economic Review*, **34**, 259–269.
- TAMURA, W. (2012), "A Theory of Multidimensional Information Disclosure" (Mimeo).