On Transparency in Organizations*

Philippe Jehiel†

21st September 2009

Abstract

Non-transparency both in the form of incomplete information disclosure and in the form of coarse feedback disclosure is optimal in virtual all organizational arrangements of interest. Specifically, in moral hazard interactions, some form of non-transparency is always desirable, as soon as the dimensionality of the problem exceeds the dimensionality of the action spaces of the various agents.

1 Introduction

A central question of organizational design is about how to make workers exert more productive effort in contexts in which effort is not directly observable. Such situations with non-observable actions also called moral hazard problems concern team interactions as well as the conditions governing the productive effort of isolated workers.

Workers in organizations are typically involved in many moral hazard problems, and the exact conditions affecting agents' incentives may vary from one moral hazard problem to the next. For example, problems may differ in how costly it is to agents to exert various types of effort, or they may differ in the degree of complementarity of agents' efforts in team interactions, or they may differ in the degree of congruence of interest both between

*I would like to thank seminar participants at Cesifo IO workshop, Paris 1 and the workshop on multidimensional mechanism design in Bonn for insightful comments.

†PSE, 48 boulevard Jourdan, 75014 Paris, France and University College London ; jehiel@enpc.fr

the agents and the organization, or in what is being observed by the employer or in the wage and bonus structures.

In this paper, I wish to explore whether it is in the interest of the organization that the workers or agents be as informed as possible about the specifics of the various moral hazard problems they are engaged in. Specifically, I ask myself whether the organizational objective can be enhanced when the agents are not fully informed of these. I also ask, for team interactions, whether it can be in the interest of the organization to let the agents know only the aggregate distribution of other team members' actions over various team problems as opposed to letting them know these distributions for each problem separately (an information that is typically required if agents are assumed -as in a Nash equilibrium- to play a best-response to the actions of their opponents in the various problems).

I say that some form of non-transparency is desirable in the organization either 1) when it is in the interest of the organization not to let the agents have full information on the characteristics of the moral hazard problem they are in or, in team interactions, 2) when it is in the interest of the organization that the agents know only other team members' distributions of actions in aggregate over various team problems.

The main result of this paper is that when the space of moral hazard problems is rich enough (i.e., it has dimension strictly larger than the dimensionality of each agent's action space), non-transparency is desirable both in the sense that keeping secret some information on the specifics of the problems strictly enhances the organizational objective and the organization would be strictly better off if the agents were only informed of the aggregate distribution of other team members' efforts in aggregate over several (well chosen) team problems. The conclusion need not be the same if the dimensionality of the space of moral hazard problems is lower, as I demonstrate through examples.

Observe that the analysis is not solely concerned with the issue of optimal design in moral hazard interactions. Even if instruments such as bonuses or wages are set suboptimally,[1] the main result of this paper still holds and non-transparency in the two senses defined above is always desirable.

[1] One can think of various constraints (for example legal ones) why these would not be optimally adjusted.

Related literature:

1) Myerson (1986) considers very general mechanisms including the possibility that a central mediator sends recommendations as to which actions the agents should perform. In such contexts, the optimal mechanism requires that the agents be just told what to do and no more (as providing more information could only make the incentive constraints potentially harder to satisfy). The result of this paper that the agents should not be fully informed of the specific of the problems is of a different nature. First, it applies beyond the framework of optimal organizational design and in particular to frameworks in which no mediator can make recommendations as to which actions to perform.[2] Second, even if such mechanisms are feasible, it is not clear from Myerson's analysis that non-full disclosure of the specifics of the problems strictly enhance the organizational objective (it only shows that the outcome of a mechanism in which such information is disclosed can always be replicated by the recommendation mechanism in which the agents are only told what to do).[3] By contrast, in this paper, it is shown that non-full disclosure is strictly optimal whenever the dimensionality of the moral hazard problem exceeds the dimensionality of agents' actions. Third, the second form of non-transparency considered in this paper, i.e. about coarse feedback disclosure in moral hazard team interactions, has no counterpart in Myerson's setup. As will be shown, by providing coarse feedback about other workers' attitudes, one can strictly improve upon Myerson's optimal design whenever the dimensionality of the moral hazard problem is big enough.[4]

2) When formalizing the interaction of workers who would only be coarsely informed of other workers' attitudes (in the various moral hazard problems), I follow the methodology developed in Jehiel (2005) in which it is assumed that subjects adopt the simplest theory consistent with the feedback they receive.[5]

3) There have been many other approaches to transparency in organizations. I mention

[2] It may be argued that in a number of real life organizations, such recommendation schemes would be subjet to ex post manipulation and thus hard to implement in a credible way.

[3] See Rahman (2009) for a recent paper that shows how the use of mediators (à la Myerson) or secret contracts can be beneficial in moral hazard team interactions.

[4] Of course, this can be achieved because the solution concept governing the interaction is no longer the Nash equilibrium in this case but the feedback equilibrium (see subsection 3.2 for formal definitions).

[5] It should be mentioned that the analog of the second question addressed in this paper in the context of moral hazard in team interactions is analyzed in Jehiel (2009) in the context of private value auctions.

here just a few to help locate the present contribution in the literature. In seminal contributions, Holmström (1979-1982) show in static moral hazard problems that it is always best that the principal be as informed as possible, as it allows her to better monitor the agent(s). In subsequent works, (Crémer (1995), Dewatripont et al. (1999a) or Prat (2005) to name just a few), dynamic considerations have been introduced, imposing some limited commitment capabilities on the Principal's side. There, less information for the Principal may help the Principal, as it may alleviate her commitment problems. Note that this line of research is more concerned about changing the information held by the Principal whereas the focus of this paper is on the information held by the agents (as well as the feedback transmitted to them).[6]

4) There is also, of course, a vast number of papers that analyze the effect of changing the information structure in strategic environments. These include, in particular, Milgrom and Weber (1982) in the context of auction with affiliated signals in which it is shown that providing more information to bidders increases revenues (an insight to be contrasted with the finding of this paper that non-full disclosure is always optimal in sufficiently rich moral hazard interactions). It also includes the extensive literature interested in when oligopolists should share their information so as to increase aggregate profits (see Vives (1984), Gal-Or (1985) or Raith (1996) to mention just a few). Again, none of these papers seem to have highlighted the role of the dimensionality of the private information in addressing these questions.

The rest of the paper is structured as follows. A general framework is presented in Section 2. The main questions are formally stated in Section 3, which also contains some preliminary examples. Section 4 contains the main results as well as a discussion of these. Section 5 illustrates the key role played by the dimensionality in deriving the main insights. A brief conclusion appears in Section 6.

[6]In a context related to the literature just mentioned, Dewatripont et al. (1999b) analyze when it can be good for the Principal to give fuzzy missions to the agent given the monitoring technology available to the Principal. Note however that the infomation available to the agent is not being changed in Dewatripont et al. (1999b). More generally, it seems that the literature on moral hazard interactions has not considered the role of the dimensionality, which plays an essential role in the present analysis (Battaglini (2006) is an exception but the questions addressed in his and this paper are completely different).

2 A general framework

I consider moral hazard problems with one or two agents parameterized by $\alpha \in \mathbb{R}^s$. The parameter α is assumed to be distributed according to a smooth (i.e., continuously differentiable) density $p(\alpha)$ that is strictly positive on some open subset of $\mathbb{R}^s$. Extensions to more than two agents raise no difficulties. In every moral hazard problem α, agent $i = 1, 2$ chooses an action a_i in A_i, an open subset of $\mathbb{R}^{n_i}$. Agents choose their actions simultaneously, that is, without observing the actions chosen by the other agent.

While the designer is assumed to know α, I consider various informational assumptions regarding what the agents know about α. In addition, the designer may or may not (depending on the application) be allowed to use instruments $w = (w_1, w_2) \in \mathbb{R}^{\omega_1} \times \mathbb{R}^{\omega_2}$ that affect agents 1 and 2' incentives respectively, and that are based on what the designer can observe (typically actions a_i are not observable to make the problem non-trivial).[7]

In problem α, agent i's expected payoff is $u_i(a_1, a_2; w, \alpha)$. The designer's expected payoff is $\pi(a_1, a_2; w, \alpha)$.

It should be mentioned that in the above formulation, agents' participation constraints are not explicitly taken into account. Yet, when one of the actions in A_i ensures that agent i gets at least what he can get outside the interaction (whatever a_j), then agent i's participation constraint is automatically satisfied. Participation constraints will be further discussed later on after our main results are stated. Mechanisms allowing the use of mediators (à la Myerson (1986)) will also be discussed then.

2.1 Applications

The framework covers lots of classic moral hazard problems. To mention, just a few:

Moral hazard in teams **(à la Holmström, 1982)**

Two risk-neutral agents 1 and 2 in a team simultaneously exert effort a_1 and a_2 say within the range $[\underline{a}, \overline{a}]$. With probability $\widetilde{p}(a_1, a_2; \beta) = a_1 + a_2 + \beta a_1 a_2$ the team is successful giving reward R to the organization where the parameter $\beta \in [\underline{\beta}, \overline{\beta}]$ reflects

[7] In Subsection 4.2 we will also consider the possibility that a mediator makes recommendations to the agents, as in Myerson (1986).

the degree of complementarity between the effort levels chosen by the two agents.[8]

Efforts are not directly observable, only success is. Agents must receive non-negative wages in all events. The instruments available to the designer are the bonuses w_1 and w_2 given to agents 1 and 2 respectively in case of success (they optimally get 0 wage in case of failure). Letting $g_i(a_i)$ denote the cost to agent i of making effort a_i, this moral hazard in team problem falls in the general framework just defined with:

$$\begin{aligned} u_i(a_1, a_2; w, \alpha) &= \widetilde{p}(a_1, a_2; \beta) w_i - g_i(a_i) \\ \pi(a_1, a_2; w, \alpha) &= \widetilde{p}(a_1, a_2; \beta)(R - w_1 - w_2) \end{aligned}$$

Here, the team problem is parameterized by $\alpha = (\beta, R, g_1, g_2)$, the profile of complementarity, reward and cost parameters.

Multi-task and moral hazard **(Holmström and Milgrom, 1991)**

Even though the general framework admits several agents, it may obviously be particularized to one agent moral hazard problems (simply by freezing one of the two agents). Given that no restrictions are being made on the dimensionality of the action space of the agent, the framework covers the important application of multi-tasking. For example, a single agent may consider exerting effort a_x, a_y in two two tasks x and y with a corresponding cost $g(a_x, a_y)$. The expected output $z = h(a_x, a_y)$ is a function of the effort produced in the two dimensions, and the designer only observes some signal $q = r(e_x, e_y) + \varepsilon$ where ε is the realization of a normal distribution with variance σ^2 and mean 0. The designer may use a signal-dependent wage schedule $w(q)$ as instrument. The objective of the designer assumed to be risk neural writes $z - E(w(q))$ and the agent assumed to exhibit constant absolute risk aversion gets an expected utility: $-E \exp[-\rho(w(q) - g(a_x, a_y)]$. The multi-task problem is parameterized by $\alpha = (h, r, \sigma, \rho, g)$.

Models of authority **(Aghion and Tirole, 1997)**

An agent 1 exerts effort a_1 to find out which project to adopt. The principal, agent 2, can exert effort so as to improve upon the choice of the agent. A good project for

[8] We should assume that $2\overline{a} + \overline{\beta}\overline{a} < 1$ so that $p(a_1, a_2; \beta) \in (0, 1)$ for all a_1, a_2 in $[\underline{a}, \overline{a}]$.

the agent gives him a private benefit b and a good project for the principal gives her a private benefit B. The probability that a good project for the agent is also a good project for the principal is γ and the probability that a good project for the principal is a good project for the agent is β. Identifying the effort levels with the probability of finding a good project and letting $g_1(a_1)$ and $g_2(a_2)$ denote the costs of efforts made by the agent and the principal, respectively, the expected utilities of the agent and the principal write:

$$u_1(a_1, a_2; \alpha) = a_2\beta b + (1 - a_2)a_1 b - g_1(a_1)$$
$$u_2(a_1, a_2; \alpha) = a_2 B + (1 - a_2)a_1\gamma b - g_2(a_2)$$

Here the designer's objective coincides with 2's objective $\pi(a_1, a_2; \alpha) = u_2(a_1, a_2; \alpha)$, and the authority problem is parameterized by $\alpha = (\beta, \gamma, b, B, g_1, g_2)$, the profile of congruence, private benefit and cost parameters.[9]

3 Main questions

Within the framework described in Section 2, I ask the following two questions.

Question 1. Can it be beneficial for the designer that at least one agent, say agent 1, be partially rather than fully informed of α?

Question 2. Can it be beneficial for the designer to confuse agent 1 about agent 2's distribution of actions in the various team problems α?

The first question echoes familiar investigations in economic theory. For example, it is similar to a question addressed by Milgrom and Weber (1982) in standard auctions with affiliated information. There, in a context of one-dimensional adverse selection auction models, Milgrom and Weber show that under affiliation, it is optimal for the seller to release as much information as she can to the bidders. To the best of my knowledge, such a question does not seem to have been addressed with much generality in the context of moral hazard problems. Note that when addressing question 1, I simply perform comparative statics varying the information structure of agent 1 (as Milgrom and Weber do in the context of auctions). That is, I do not discuss the issue of how the information

[9] In Aghion-Tirole's model, there are no monetary instruments.

disclosure policy chosen by the designer would be interpreted by the agents so as to refine their estimate of α. Such a view seems appropriate to deal with organizations in which there is enough time to commit in advance (before the realization of α is known) to whatever disclosure policy sounds best.

The second question is slightly less conventional. Yet, it seems relevant for a number of practical organizational designs. Equilibrium conditions are generally more meaningfully thought of as resulting from learning and/or adaptive processes. If agents manage to learn the distribution of actions of other agents in every team problem (simply by analyzing the data available from previous team problems), they can optimally adjust a best-response, and a Nash equilibrium is then viewed as a plausible description of the interaction.

But, the designer can possibly hide somehow the conditions α that prevailed when some action profile (a_1, a_2) was played in the past. In such a case, a Nash equilibrium would be less likely to be played even in the limit as lots of data accumulate. I postulate that faced with such a coarse feedback (to be made explicit below), agents would play an analogy-based expectation equilibrium (as defined in Jehiel, 2005) in which the analogy partition applying to each agent would be chosen by the designer (as in Jehiel, 2009 in the context of private values auctions).

When the answer to either question 1 or 2 is positive, I say that some form of non-transparency is optimal in the organization. While a positive answer to question 1 would require that the agent be kept uninformed of the condition α of the team problem at the time he must choose his action, a positive answer to question 2 would require that the designer somehow hides some information from past experiences in the organization but not the conditions prevailing in the team interaction of current interest. It may be argued that in a number of cases the second type of non-transparency is slightly easier to implement than the first one from a practical viewpoint (it may also be more efficient as well in some cases).

3.1 The solution concepts

In order to address questions 1 and 2, I need to describe how agents 1 and 2 would interact under different information and feedback assumptions.

3.1.1 Full information benchmark

In the benchmark scenario, agents 1 and 2 know α and in equilibrium they know each other's strategies. That is, given the instruments w, agents 1 and 2 play a Nash equilibrium of the complete information game defined by the payoff $u_i(a_1, a_2; w, \alpha)$ received by agent i for every action profile $a = (a_1, a_2)$ and the instrument(s) w.

In order to avoid technical complications, I will assume that u_i is a concave function of a that varies smoothly with w and α. Moreover, I will assume that whatever a_j,w, α, the function $a_i \to u_i(a_1, a_2; w, \alpha)$ is never maximized on the boundary of A_i.

Such assumptions guarantee that an interior pure strategy Nash equilibrium exists, and that for almost all (w, α), Nash equilibria are locally unique and vary smoothly with (w, α) (see MasColell et al. (1995)).

I will denote by $a_i^{NE}(w, \alpha)$ one such equilibrium and I will assume it is the one describing the interaction in our team problem. Thus, in the benchmark scenario, in problem α, the designer sets the instruments $w = w(\alpha)$ (available to him) so as to maximize:

$$\pi(a_1^{NE}(w, \alpha), a_2^{NE}(w, \alpha); w, \alpha).$$

I will be interested in situations in which the solution obtained is typically different from the first-best solution the designer would choose if she could herself decide on a_1, a_2 as well as w. This is typically the case in moral hazard problems with one or two agents (unless agents' preferences are perfectly aligned with those of the designer and/or the designer can observe agents' actions) and also more generally if transfers must be bounded and/or agents are risk averse. Observe that in such cases, it is typically the case that $\frac{\partial}{\partial a_1}\pi(a_1^{NE}(w, \alpha), a_2^{NE}(w, \alpha); w, \alpha) \neq 0$. That is, even adjusting the instruments w optimally, the marginal effect of a_1 in every direction need not be 0.

3.1.2 The coarse information case

To address question 1, I will consider situations in which agent 1 does not know whether $\alpha = \alpha_x$ or α_y while the designer and agent 2 do. In this case, the relevant solution concept is the Nash Bayes equilibrium. The above concavity and smoothness assumptions guarantee again the existence of an interior pure strategy equilibrium and that Nash equilibria (which are locally unique) inherit the smoothness properties of u_i for almost all w and α_x or α_y.

For each w_x, w_y, a Nash Bayes equilibrium is such that player 1 chooses action a_1^{CI} in both α_x, α_y and player 2 chooses actions $a_{2,x}^{CI}$ and $a_{2,y}^{CI}$ in α_x and α_y with

$$\begin{aligned} a_{2,x}^{CI} &\in \arg\max_{a_2} u_2(a_1^{CI}, a_2; w_x, \alpha_x) \\ a_{2,y}^{CI} &\in \arg\max_{a_2} u_2(a_1^{CI}, a_2; w_y, \alpha_y) \\ a_1^{CI} &\in \arg\max_{a_1} p(\alpha_x) u_1(a_1, a_{2,x}^{CI}; w_x, \alpha_x) + p(\alpha_y) u_1(a_1, a_{2,y}^{CI}; w_y, \alpha_y) \end{aligned}$$

Letting $a^{CI}(w)$ denote the Nash Bayes equilibrium prevailing in the team problem, the best choice of instruments w is then obtained by maximizing

$$p(\alpha_x)\pi(a_1^{CI}(w), a_{2,x}^{CI}(w); w_x, \alpha_x) + p(\alpha_y)\pi(a_1^{CI}(w), a_{2,y}^{CI}(w); w_y, \alpha_y).$$

In the analysis, I will assume that if $a_1^{NE}(w_x, \alpha_x) = a_1^{NE}(w_y, \alpha_y)$ in the full information benchmark, then in the game in which agent 1 does not know whether α_x or α_y, the play is described by the complete information equilibrium strategy profile, as well. That is, such a strategy profile is clearly a Nash Bayes equilibrium of the incomplete information game, and the assumption just made ensures that the comparison between the two informational scenarios is meaningful.[10]

A positive answer to question 1 is obtained when one can find α_x and α_y such that

$$\max_w p(\alpha_x)\pi(a_1^{CI}(w), a_{2,x}^{CI}(w); w_x, \alpha_x) + p(\alpha_y)\pi(a_1^{CI}(w), a_{2,y}^{CI}(w); w_y, \alpha_y)$$

is strictly larger than

$$p(\alpha_x)\pi(a_1^{NE}(w_x, \alpha_x), a_2^{NE}(w_x, \alpha_x); w_x, \alpha_x) + p(\alpha_y)\pi(a_1^{NE}(w_y, \alpha_y), a_2^{NE}(w_y, \alpha_y); w_y, \alpha_y)$$

for all w.[11]

[10] Alternatively, stronger conditions on u_i and u_j could be imposed that guarantee the uniqueness of an equilibrium.

[11] For Theorem 1, it is enough to consider information sets consisting of two states.

3.2 The coarse feedback case

To address question 2, I will consider scenarios in which agent 1 is led (by the designer) to confuse the distribution of actions of agent 2 in $\alpha = \alpha_x$ and α_y while still knowing whether $\alpha = \alpha_x$ or α_y. I adopt the approach developed in Jehiel (2005-2009) to model this.

Consider α_x and α_y and fix w_x and w_y. Agents 1 and 2 know whether $\alpha = \alpha_x$ or α_y. Agent 2 is rational as usually modeled. In problem $\alpha = \alpha_z$ he plays a best-response to the actual action $a_{1,z}^{CF}$ of player 1. Agent 1 in problems $\alpha = \alpha_x$, α_y is assumed to play a best-response to the aggregate distribution of player 2's actions over α_x and α_y. That is, calling $a_{2,x}^{CF}$ and $a_{2,y}^{CF}$ the actions of player 2 in α_x and α_y respectively, agent 1 plays a best response to the conjecture that agent 2 chooses $a_{2,x}^{CI}$ with probability $\frac{p(\alpha_x)}{p(\alpha_x)+p(\alpha_y)}$ and $a_{2,y}^{CI}$ with probability $\frac{p(\alpha_y)}{p(\alpha_x)+p(\alpha_y)}$ in each $\alpha = \alpha_x$, α_y. Or to put it more formally, for $z = x, y$

$$a_{1,z}^{CF} \in \arg\max_{a_1} p(\alpha_x)u_1(a_1, a_{2,x}^{CI}; w_z, \alpha_z) + p(\alpha_y)u_1(a_1, a_{2,y}^{CI}; w_z, \alpha_z)$$

Note that unless $u_1(\cdot,\cdot; w_x, \alpha_x) = u_1(\cdot,\cdot; w_y, \alpha_y)$ this is typically different from the situation in which 1 does not know whether $\alpha = \alpha_x$ or α_y, as, for example, it may lead agent 1 to pick different actions in α_x and α_y.

I call such a profile a^{CF} a feedback equilibrium. A feedback equilibrium should be interpreted as a steady state of a learning process involving in each round populations of agents 1 and 2 engaged in problem α_x with $w = w_x$ (in proportion $\frac{p(\alpha_x)}{p(\alpha_x)+p(\alpha_y)}$) and problem α_y with $w = w_y$ (in proportion $\frac{p(\alpha_y)}{p(\alpha_x)+p(\alpha_y)}$). While agents 2 would be told the past empirical distribution of actions a_1 in α_x and α_y separately, agents 1 would only be told the aggregate empirical distribution of agents 2's actions over α_x and α_y. If agents 1 adopt the simplest conjecture about agents 2 based on the feedback they receive and if the distributions of play of agents 1 and 2 stabilize in α_x and α_y, a feedback equilibrium is being played (see Jehiel (2005) for further elaborations on this concept and how it differs from Nash equilibrium in general).

It should be mentioned that, as for the coarse information approach, the concavity and smoothness of u guarantee the existence of feedback equilibria in pure strategies that are locally unique and that vary smoothly (almost everywhere) with w and α_x, α_y.

Letting $a^{CF}(w)$ denote the feedback equilibrium prevailing in the team problem, the best choice of instruments w is then obtained by maximizing:

$$p(\alpha_x)\pi(a_{1,x}^{CF}(w), a_{2,x}^{CI}(w); w_x, \alpha_x) + p(\alpha_y)\pi(a_{1,y}^{CF}(w), a_{2,y}^{CI}(w); w_y, \alpha_y).$$

In the analysis, I will assume that if $a_2^{NE}(w_x, \alpha_x) = a_2^{NE}(w_y, \alpha_y)$ in the full information benchmark, then in the game in which agent 1 is being confused about agent 2's actions in α_x and α_y, the play is described by this same strategy profile. This is again to make the comparison with the full information benchmark meaningful.

A positive answer to question 2 is obtained when one can find α_x and α_y such that[12]

$$\max_w p(\alpha_x)\pi(a_{1,x}^{CF}(w), a_{2,x}^{CI}(w); w_x, \alpha_x) + p(\alpha_y)\pi(a_{1,y}^{CF}(w), a_{2,y}^{CI}(w); w_y, \alpha_y)$$

is strictly larger than

$$p(\alpha_x)\pi(a_1^{NE}(w_x, \alpha_x), a_2^{NE}(w_x, \alpha_x); w_x, \alpha_x) + p(\alpha_y)\pi(a_1^{NE}(w_y, \alpha_y), a_2^{NE}(w_y, \alpha_y); w_y, \alpha_y)$$

for all w.

3.3 Preliminary examples

Before stating the main results, I first provide simple examples in which the answers to questions 1 and 2 are positive.

3.3.1 When coarse information is good

Consider two one-agent moral hazard problems α_x, α_y in which the agent must perform two tasks $a_x, a_y \in [0, 1]$. The cost incurred by the agent is $c(a_x, a_y) = h(a_x) + h(a_y)$ both in α_x and α_y where $h(0) = h'(0) = 0$ and $h(\cdot)$ is assumed to be increasing.

The output exhibits complementarities between the two tasks and it is given by $z = a_x a_y + \varepsilon$ where ε is the realization of some normal distribution centered around 0.

[12] For Theorem 2, it is enough to consider feedback classes involving two states.

Output is not assumed to be observable (at least within a reasonable amount of time). In situation α_x, only $q_x = a_x + \varepsilon_x$ is observed by the principal and similarly in situation α_y, only $q_y = a_y + \varepsilon_y$ is observed by the principal where ε_x and ε_y are the realizations of independent normal distributions centered around 0. Let $p(\alpha_1) = p(\alpha_2) = \frac{1}{2}$.

I assume that wages must be non-negative. The principal's instrument thus boils down to offering bonus schemes $w_x(q_x) \geq 0$ in α_x or $w_y(q_y) \geq 0$ in α_y. The agent and the principal are assumed to be risk neutral. The agent gets a payoff equal to $w - c(a_x, a_y)$ when he earns w and exerts effort $a = (a_x, a_y)$; the principal gets an expected payoff equal to $a_x a_y - w$ under the same circumstances.

It is rather easy to see the advantage of not letting the agent know whether α_x or α_y in this problem. Assume that the agent knows α_x. Then clearly, the agent will pick $a_y = 0$ whatever $w_x(\cdot)$ (this is because a_y does not affect q_x and any $a_y > 0$ would induce strictly positive extra cost). Thus, expected output is 0 in the full information case (and w_x and w_y are optimally set at 0).

By contrast, consider the case in which the agent does not know whether α_x or α_y. It is fairly easy to induce $a_x > 0$ and $a_y > 0$ through the choice of strictly increasing $w_x(\cdot), w_y(\cdot)$ because now the agent chooses (a_x, a_y) so as to maximize:

$$\frac{1}{2}E(w_x(a_x + \varepsilon_x)) + \frac{1}{2}E(w_y(a_y + \varepsilon_y)) - c(a_x, a_y).$$

More precisely, one can establish that the full information benchmark is dominated by the coarse information case whenever $h''(0) < \frac{1}{2}$ (by considering schemes of the form $w(q) = \max(0, \omega q)$ for sufficiently small ω).

In more intuitive terms, not letting the agent know whether α_x or α_y makes it easier to let the agent exert effort on both tasks because he does not know which one will be used as a performance measure to reward him. By contrast, when the agent knows that he will be assessed only on the basis of a_x (which is a consequence of the monitoring technology in α_x) he has no incentive to exert effort on a_y, which when the two tasks are sufficiently complement, is very detrimental to the output. A related intuition appears in a recent paper by Ederer et al. (2008) who consider mixed moral hazard Principal-agent problems in which the agent has superior information.

Of course, the above example should not be interpreted to mean that coarse information is always good. An obvious potential disadvantage of coarse information is that the agent can no longer adjust his effort decision to the exact conditions governing the moral hazard interaction. In general, coarse information has the advantage of easing the incentive constraints (because it aggregates several incentive constraints into a single one, thus easier to satisfy), and it has the disadvantage of making the strategy less sensitive to the environment (the strategy must be measurable with respect to a coarser information partition). The trade-off between these two forces can go either way in general, but as will be seen, in a rich environment space case, one can always find a coarse information structure that strictly enhances the designer's objective as compared with the full information benchmark.

3.3.2 When coarse feedback is good

Consider the following family of moral hazard in team problems. Each agent $i = 1, 2$ must simultaneously exert an effort $a_i \in \mathbb{R}^+$. The outcome of the team interaction is either successful with probability $p(a_1, a_2; \beta)$ or it is not successful. Assume that $p(a_1, a_2; \beta) = a_1 + a_2 + \beta a_1 a_2$. Assume further that wages must be non-negative so that the instruments boil down to picking a bonus $w_i \geq 0$ for agent $i = 1, 2$ in case of success. The cost of effort a_i is $g(a_i) = \frac{1}{2}\gamma\,(a_i)^2$ to agent i. Agents are risk neutral so that agent i's payoff writes: $u_i(a_1, a_2; \alpha) = p(a_1, a_2; \beta)w_i - g(a_i)$.

The output is R in case of success; it is 0 otherwise. The principal is risk neutral and her payoff thus writes: $\pi(a_1, a_2; w, \alpha) = p(a_1, a_2; \beta)(R - w_1 - w_2)$.

Two such problems are considered: α_x in which $\beta_x = 0$ and $\gamma_x = \underline{\gamma}$ and α_y in which $\beta_y > 0$ and $\gamma_y = \overline{\gamma}$ with $\overline{\gamma} > \underline{\gamma}$. Both problems are assumed to be equally likely: $p(\alpha_1) = p(\alpha_2) = \frac{1}{2}$.

I claim that for $\overline{\gamma}$ large enough, providing coarse feedback to agent 1 about agent 2's effort over α_x and α_y (with effects as described in Section 3) is good for the principal.

To see this, consider the full information benchmark and the corresponding optimal $w_{1,z}$, $w_{2,z}$ in α_z for $z = x, y$. Clearly, for $\overline{\gamma}$ large enough, it holds that $a_{2,x} > a_{2,y}$ in Nash equilibrium. Consider such $\overline{\gamma}$.

Consider now the coarse feedback scenario in which agent 1 is led to aggregate agent 2's actions over α_x and α_y (and agent 2 is fully rational). Agent 1's expectation about agent 2's effort is that agent 2 exerts effort $a_{2,x}^{CF}$ with probability $\frac{1}{2}$ and effort $a_{2,y}^{CF}$ with probability $\frac{1}{2}$ (where $a_{2,z}^{CF}$ denotes agent 2 action in α_z).

As compared with the full information benchmark, such a confusion has no effect in agent 1's effort choice in α_x because agent 1 does not care about agent 2's effort in this case ($\beta_x = 0$). In α_y, however, the upward shift of agent 1's expectation moves $a_{1,y}$ upwards (due to strategic complementarity, $\beta_y > 0$). Such a confusion leads to an improvement of the principal's objective because it allows her to obtain the same levels of efforts of both agents with a lower bonus $w_{1,y}$ to agent 1 in α_y.[13]

4 Main Results

In this Section, the main results about transparency are stated. To this end, I first define the notion of genericity employed here. Let $X = \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^{\omega_1+\omega_2} \times \mathbb{R}^s$ denote the domain of the profit functions π. Consider functions $\pi \in C^2(X)$. The set $\overline{\Pi}$ of $C^2(X)$ profit functions is endowed with a Whitney C^2 topology by letting a sequence $\pi_n \in \overline{\Pi}$ converge to π if and only if $\pi_n - \pi$ as well as the Jacobian of $\pi_n - \pi$ and the matrix of second derivative of $\pi_n - \pi$ converge uniformly to zero in the space of continuous functions with euclidean norm. The usual definition of genericity is:

Definition. A set $\Pi \subseteq \overline{\Pi}$ is generic in $\overline{\Pi}$ if it contains a set that is open and dense in $\overline{\Pi}$.

The main results are:

Theorem 1 *Suppose the dimensionality of α is strictly bigger than the dimensionality of a_1, that is, $s > n_1$. Then for generic functions π, some non-full disclosure of α to agent 1 strictly enhances the designer's objective as compared with the full information benchmark.*

Theorem 2 *Suppose the dimensionality of α is strictly bigger than the dimensionality of a_2, that is, $s > n_2$. Then for generic functions π, some non-full disclosure to agent 1*

[13] As in the case of coarse information, the above example should not be interpreted to mean that coarse feedback is always good, as for example in some cases it may be better that agent 1 knows more precisely the effort made by agent 2.

about agent 2's effort strictly enhances the designer's objective as compared with the full information benchmark.

Before getting into the formal arguments, observe that the conditions of Theorems 1 and 2 are naturally met given that the parameter α characterizing the team problem should be thought of as containing at least information on the structure of the marginal cost incurred by each agent i along the various dimensions of his effort a_i (this has dimension no less than n_i), together with say information on the effect of the action profiles on the designer's objective. Thus, the dimension of the team problem s should in general be thought of to be strictly larger than $n_1 + n_2 \geq \max(n_1, n_2)$, thereby ensuring that the conditions of both Theorems 1 and 2 hold.

4.1 The main arguments

Theorem 1 will first be established in the special case in which the principal has no instrument w, there is a single agent who has to choose a one-dimensional action, and the problem varies along two dimensions. It will then be explained how the same type of arguments can be used to extend the result to general multi-agent settings with arbitrary instruments for the designer and arbitrary dimensions n_1, n_2 whenever $s > n_1$. Finally, it will be explained how similar arguments can be made to prove Theorem 2.

Consider a setting with one agent whose action is $a \in \mathbb{R}$, and $\alpha \in \mathbb{R}^2$ parameterizes the agent's payoff function $u(a;\alpha)$. The complete information solution $a(\alpha)$ satisfies: $\frac{\partial}{\partial a}u(a;\alpha) = 0$.

Consider α_0 in the interior of the α-space, and let $\overline{A}(\alpha_0) = \{\alpha \text{ such that } a(\alpha) = a(\alpha_0)\}$. For smooth u and generic α_0 this is a smooth (i.e. locally differentiable) manifold of dimension 1, which (locally around α_0) lies in the interior of the α-space (this is the key place where $s > n_1$ is being used). Let $\alpha_1 \in \overline{A}(\alpha_0)$, $\alpha_1 \neq \alpha_0$ be in the interior of the α-space.

Starting from α_1, consider a direction δ in the α space in which $\alpha_1 + \varepsilon\delta$ is not in $\overline{A}(\alpha_0)$ for ε small enough and such that $\frac{\partial^2 u}{\partial a \partial \alpha^\delta}(a_0;\alpha_1) \neq 0$. (Such a direction exists for generic u. For example, such a direction may be one in which the marginal effect of a on u is modified in proportion to a.)

Consider the problems $\alpha = \alpha_0$ and $\alpha_1 + \varepsilon\delta$ for ε either positive or negative but small (remember α_1 lies in the interior of the α-space). The idea is to compare the aggregate expected effect on the objective π of letting the agent know whether $\alpha = \alpha_0$ or $\alpha_1 + \varepsilon\delta$ and letting the agent ignore whether $\alpha = \alpha_0$ or $\alpha_1 + \varepsilon\delta$.

Clearly for $\varepsilon = 0$, the two cases generate the same aggregate π by definition of $\overline{A}(\alpha_0)$.[14] 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 a coarse information of the above type either for $\varepsilon > 0$ and small or $\varepsilon < 0$ and small dominates the complete information case.

Let $a_0 = a(\alpha_0)$ and $a_1(\varepsilon) = a(\alpha_1 + \varepsilon\delta)$. They satisfy

$$\begin{aligned} \frac{\partial}{\partial a} u(a_0; \alpha_0) &= 0 \\ \frac{\partial}{\partial a} u(a_1(\varepsilon); \alpha_1 + \varepsilon\delta) &= 0 \end{aligned} \tag{1}$$

Let $a^{CI}(\varepsilon)$ denote the action when the agent does not know whether $\alpha = \alpha_0$ or $\alpha_1 + \varepsilon\delta$. It satisfies:

$$p(\alpha_0)\frac{\partial}{\partial a} u(a^{CI}(\varepsilon); \alpha_0) + p(\alpha_1 + \varepsilon\delta)\frac{\partial}{\partial a} u(a^{CI}(\varepsilon); \alpha_1 + \varepsilon\delta) = 0. \tag{2}$$

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

$$p(\alpha_0)[\pi(a_0; \alpha_0) - \pi(a^{CI}(\varepsilon); \alpha_0)] + p(\alpha_1 + \varepsilon\delta)[\pi(a_1(\varepsilon); \alpha_1 + \varepsilon\delta) - \pi(a^{CI}(\varepsilon); \alpha_1 + \varepsilon\delta)].$$

Clearly, if $\Delta(\varepsilon) < 0$, it is strictly better that the agent does not know whether $\alpha = \alpha_0$ or $\alpha_1 + \varepsilon\delta$.

I now expand $\Delta(\varepsilon)$ at the first order in ε. Since $a^{CI}(0) = a_1(0) = a_0$, $\Delta(\varepsilon)$ writes at the first order:

$$p(\alpha_0)\frac{\partial \pi}{\partial a}(a_0; \alpha_0)[a_0 - a^{CI}(\varepsilon)] + p(\alpha_1)\frac{\partial \pi}{\partial a}(a_0; \alpha_1)[a_1(\varepsilon) - a^{CI}(\varepsilon)] + o(\varepsilon)$$

[14] In the general multi-agent extension, the selection hypothesis for Nash Bayes equilibria of games of incomplete information is being used as well.

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

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

$$\begin{aligned} a_1(\varepsilon) - a_0 &= \frac{-\frac{\partial^2 u}{\partial a \partial \alpha^\delta}(\alpha_1)}{\frac{\partial^2 u}{\partial a^2}(\alpha_1)}\varepsilon + o(\varepsilon) \\ a^{CI}(\varepsilon) - a_0 &= \frac{-p(\alpha_1)\frac{\partial^2 u}{\partial a \partial \alpha^\delta}(\alpha_1)}{p(\alpha_0)\frac{\partial^2 u}{\partial a^2}(\alpha_0) + p(\alpha_1)\frac{\partial^2 u}{\partial a^2}(\alpha_1)}\varepsilon + o(\varepsilon) \end{aligned}$$

where $\partial h / \partial \alpha^\delta$ denotes the derivative of h with respect to the direction α^δ and all functions are taken at $a = a_0$.

After multiplying $\Delta(\varepsilon)$ by $\frac{\partial^2 u}{\partial a^2}(\alpha_1)[p(\alpha_0)\frac{\partial^2 u}{\partial a^2}(\alpha_0)+p(\alpha_1)\frac{\partial^2 u}{\partial a^2}(\alpha_1)]$ and dividing by $p(\alpha_0)p(\alpha_1)$ (which are both strictly positive) we get that $\Delta(\varepsilon)$ has the same sign as

$$\left[\frac{\partial \pi}{\partial a}(\alpha_0)\frac{\partial^2 u}{\partial a^2}(\alpha_1) - \frac{\partial \pi}{\partial a}(\alpha_1)\frac{\partial^2 u}{\partial a^2}(\alpha_0)\right]\frac{\partial^2 u}{\partial a \partial \alpha^\delta}(\alpha_1)\varepsilon + o(\varepsilon)$$

Three cases may a priori occur.

1) $\left[\frac{\partial \pi}{\partial a}(\alpha_0)\frac{\partial^2 u}{\partial a^2}(\alpha_1) - \frac{\partial \pi}{\partial a}(\alpha_1)\frac{\partial^2 u}{\partial a^2}(\alpha_0)\right]\frac{\partial^2 u}{\partial a \partial \alpha^\delta}(\alpha_1) < 0$. Then taking $\varepsilon > 0$ and sufficiently small, we can infer from the above that not letting the agent know whether $\alpha = \alpha_0$ or $\alpha_1 + \varepsilon\delta$ strictly dominates the complete information benchmark.

2) Likewise, if $\left[\frac{\partial \pi}{\partial a}(\alpha_0)\frac{\partial^2 u}{\partial a^2}(\alpha_1) - \frac{\partial \pi}{\partial a}(\alpha_1)\frac{\partial^2 u}{\partial a^2}(\alpha_0)\right]\frac{\partial^2 u}{\partial a \partial \alpha^\delta}(\alpha_1) > 0$, then taking $\varepsilon < 0$ and sufficiently small, not letting the agent know whether $\alpha = \alpha_0$ or $\alpha_1 + \varepsilon\delta$ strictly dominates the complete information benchmark (remember than since α_1 is in the interior of the α-space, one can move in any direction from α_1).

3) The only case in which one cannot conclude is when

$$\left[\frac{\partial \pi}{\partial a}(\alpha_0)\frac{\partial^2 u}{\partial a^2}(\alpha_1) - \frac{\partial \pi}{\partial a}(\alpha_1)\frac{\partial^2 u}{\partial a^2}(\alpha_0)\right]\frac{\partial^2 u}{\partial a \partial \alpha^\delta}(\alpha_1) = 0$$

or

$$\frac{\partial \pi}{\partial a}(\alpha_0)\frac{\partial^2 u}{\partial a^2}(\alpha_1) - \frac{\partial \pi}{\partial a}(\alpha_1)\frac{\partial^2 u}{\partial a^2}(\alpha_0) = 0 \tag{3}$$

But, this condition is not satisfied for generic π functions (observe that changing π does not affect the expressions of $a_1(\varepsilon)$, $a^{CI}(\varepsilon)$).

To see this, consider the family of π_λ functions $\pi_\lambda(a;\alpha) = \pi(a;\alpha) + \lambda a \parallel \alpha - \alpha_0 \parallel^2$ where $\lambda \in \mathbb{R}$ and $\parallel \alpha - \alpha_0 \parallel$ denotes the euclidean distance between α and α_1. Obviously, if π satisfies (3), then for $\lambda \neq 0$, π_λ does not satisfy (3) from which one can conclude that the set of π for which (3) does not hold is dense. Moreover, this set is also clearly open given the continuity of the mapping $\pi \longrightarrow \frac{\partial \pi}{\partial a}(\alpha_0)\frac{\partial^2 u}{\partial a^2}(\alpha_1) - \frac{\partial \pi}{\partial a}(\alpha_1)\frac{\partial^2 u}{\partial a^2}(\alpha_0)$ according to the Whitney C^2 topology.

I now sketch how the argument extends to the general case considered in Theorem 1.

1) **Adding instruments w.**

Suppose the designer can now (optimally) choose instrument(s) w still assuming that there is a single agent. To fix ideas, take the above setting and assume that the designer can set $w \in \mathbb{R}$. For any α, there is an optimal w, say $w(\alpha)$. This function is locally a smooth function of α for generic π and u. It is implicitly defined by

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

Define $\overline{\pi}(a;\alpha) = \pi(a, w(\alpha);\alpha)$ and apply the argument developed above when there were no instruments assuming $\overline{\pi}$ is the designer's objective. Clearly, if not letting the agent know whether $\alpha = \alpha_0$ or $\alpha_1 + \varepsilon\delta$ 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 $w(\alpha)$ in problem α).

It remains to show that generically it is not the case that

$$\frac{\partial \overline{\pi}}{\partial a}(\alpha_0)\frac{\partial^2 u}{\partial a^2}(\alpha_1) - \frac{\partial \overline{\pi}}{\partial a}(\alpha_1)\frac{\partial^2 u}{\partial a^2}(\alpha_0) = 0 \tag{4}$$

To see this, consider the family of π_λ functions

$$\pi_\lambda(a;\alpha) = \pi(a;\alpha) + \lambda \parallel \alpha - \alpha_0 \parallel^2 (a - \frac{\partial^2 u/\partial a \partial w}{\partial^2 u/\partial a^2}(a_1, w(\alpha_1);\alpha_1)w)$$

where $\lambda \in \mathbb{R}$. For such a family, $w(\alpha)$ are the same at $\alpha = \alpha_0$ (resp. α_1) whatever λ so that $\frac{\partial \overline{\pi}_\lambda}{\partial a}(\alpha) = \frac{\partial \overline{\pi}}{\partial a} + \lambda \parallel \alpha - \alpha_0 \parallel^2$ for $\alpha = \alpha_0$ and α_1. Thus, if $\overline{\pi}$ satisfies (4), for any $\lambda \neq 0$, $\overline{\pi}_\lambda$ does not, and one can conclude as before.

2) **Having more than one player.**

Roughly, this consists in extending the above differential arguments that were derived from one agent optimization conditions to a system of simultaneous optimization conditions as derived from the Nash equilibrium conditions.

Specifically, consider the case in which there is no instrument w. The FOC for NE (full information) write:

$$\left\{ \begin{array}{c} \frac{\partial u_1}{\partial a_1}(a_1, a_2; \alpha) = 0 \\ \frac{\partial u_2}{\partial a_2}(a_1, a_2; \alpha) = 0 \end{array} \right.$$

which defines implicitly $a_1(\alpha)$ and $a_2(\alpha)$. Given that α has higher dimension than a_1 one can define (for generic u_1 and u_2) a manifold of dimension $s - n_1 \geq 1$ in the α space such that $a_1(\alpha) = a_1(\alpha_0)$, i.e. $\overline{A}(\alpha_0) = \{\alpha \text{ s.t. } a_1(\alpha) = a_1(\alpha_0)\}$.

Consider $\alpha_1 \in \overline{A}(\alpha_0)$ and a direction δ in the α space so that $\alpha_1 + \varepsilon\delta$ is known not to be in $\overline{A}(\alpha_0)$. If agent 1 does not know whether α_0 or $\alpha_1 + \varepsilon\delta$:

$$\left\{ \begin{array}{c} \frac{\partial u_2}{\partial a_2}(a_1^c(\varepsilon), a_{2,0}^c(\varepsilon); \alpha_0) = 0 \\ \frac{\partial u_2}{\partial a_2}(a_1^c(\varepsilon), a_{2,1}^c(\varepsilon); \alpha_1 + \varepsilon\delta) = 0 \\ p(\alpha_0)\frac{\partial u_1}{\partial a_1}(a_1^c(\varepsilon), a_{2,0}^c(\varepsilon); \alpha_0) + p(\alpha_1 + \varepsilon\delta)\frac{\partial u_1}{\partial a_1}(a_1^c(\varepsilon), a_{2,1}^c(\varepsilon); \alpha_1 + \varepsilon\delta) = 0 \end{array} \right.$$

And if there is full information:

$$\left\{ \begin{array}{c} \frac{\partial u_2}{\partial a_2}(a_{1,0}, a_{2,0}; \alpha_0) = 0 \\ \frac{\partial u_2}{\partial a_2}(a_{1,1}(\varepsilon), a_{2,1}(\varepsilon); \alpha_1 + \varepsilon\delta) = 0 \\ \frac{\partial u_1}{\partial a_1}(a_{1,0}, a_{2,0}; \alpha_0) = 0 \\ \frac{\partial u_1}{\partial a_1}(a_{1,1}(\varepsilon), a_{2,1}(\varepsilon); \alpha_1 + \varepsilon\delta) = 0 \end{array} \right.$$

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

$$\begin{aligned}\Delta(\varepsilon) &= p(\alpha_0)[\pi(a_1^c(\varepsilon), a_{2,0}^c(\varepsilon); \alpha_0) - \pi(a_{1,0}, a_{2,0}; \alpha_0)] + \\ & \quad p(\alpha_1 + \varepsilon\delta)[\pi(a_1^c(\varepsilon), a_{2,1}^c(\varepsilon); \alpha_1 + \varepsilon\delta)] - \pi(a_{1,1}(\varepsilon), a_{2,1}(\varepsilon); \alpha_1 + \varepsilon\delta)]\end{aligned}$$

Similarly to the one agent case if $\Delta'(0) \neq 0$, then it implies that not letting agent 1 know whether α_0 or $\alpha_1 + \varepsilon d$ 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 $\pi_\lambda(a_1, a_2; \alpha) = \pi(a_1, a_2; \alpha) + \lambda a_1 \parallel \alpha - \alpha_0 \parallel^2$.

3) **Feedback manipulation (Theorem 2).**

Theorem 2 is proven in the same way now considering $\alpha_1 \in \overline{B}(\alpha_0)$

$$\overline{B}(\alpha_0) = \{\alpha \text{ s.t. } a_2(\alpha) = a_2(\alpha_0)\}$$

$\overline{B}(\alpha_0)$ is generically a smooth manifold of dimension $s - n_2$. Clearly, inducing confusion between α_0 and α_1 does not affect the outcome and one may as before consider the effect of inducing confusion between α_0 and $\alpha_1 + \varepsilon\delta$ yielding generically a strict improvement either for $\varepsilon > 0$ or $\varepsilon < 0$ but small.

4.2 Discussion

1) The above argument for Theorem 1 shows that one can gain by not letting agent 1 know whether $\alpha = \alpha_0$ or $\alpha_1(\alpha_0)$ ($\alpha_1(\alpha_0) = \alpha_1 + \varepsilon\delta$ in the above notation). By considering a positive mass neighborhood of $N(\alpha_0)$ and the corresponding $\alpha_1(\alpha)$ for almost all $\alpha \in N(\alpha_0)$, one can in fact show that the gains of not letting agent 1 know whether $\alpha \in N(\alpha_0)$ or $\alpha_1(\alpha)$ are strictly positive in expected terms. The same comment applies to the manipulation in Theorem 2.

2) Getting back to the trade-off (resulting from coarsening the information partition) between relaxing the incentive constraints (through aggregation) and constraining the strategy (through measurability constraints), Theorem 1 shows that one can always find an information partition such that the former effect dominates the latter. Yet, the argument used to prove this is not to show that the latter effect can be made of second order as

compared with the former effect. In the construction, when agent 1 does not know whether $\alpha = \alpha_0$ or $\alpha_1 + \varepsilon\delta$, both effects are of the same order. The result follows because, it is generically the case that for either $\varepsilon > 0$ or $\varepsilon < 0$ but small the former effect dominates the latter effect.

3) In the above analysis, I have implicitly ignored agents' participation constraints. This is fine to the extent that the participation constraints are not binding.[15] For example, in contexts with limited liability, agents typically receive a positive rent in moral hazard problems and the participation constraints are not binding. In the absence of limited liability constraints though, the designer would typically adjust the instruments w so that agents get their outside option payoff 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, agents were assumed to possess some private information then most "types" of agents would receive positive rent even in the absence of limited liability constraints. Theorems 1 and 2 could then be applied to such settings.

4) In the above framework, I have not allowed for mechanisms in which a mediator could make recommendations to agents as to which actions to choose. If such mechanisms are allowed and full rationality of players is assumed, one can always implement the optimal mechanism by having agents be only informed of what to do (action a_i for agent i) (see Myerson, 1986). From that perspective, what Theorem 1 shows is the stronger property that when the dimension of α is larger than the dimension of agents' actions it *cannot* be optimal to let the agents know α. A natural question one may wish to address is whether it is possible to improve the organizational objective upon the best recommendation mechanism by providing agent 1 with coarse feedback as to how agent 2 behaves. I now illustrate (though a simple example) why one should expect such improvements to be possible.

Specifically, consider the following scenario. Agent 1 is assumed not to be informed of $\alpha \in \Omega$ which contains a manifold of dimension at least $s - n_1$ (remember that in the

[15] If the participation constraints are binding both at $\alpha = \alpha_0$ and $\alpha_1 + \varepsilon\delta$ in the main argument used to prove Theorem 1 when w is set at $w(\alpha)$ in problem α, one has to worry that the agent gets no less than his outside option payoff when the agent does not know whether $\alpha = \alpha_0$ or $\alpha_1 + \varepsilon\delta$, which may require increasing the burden to the designer.

optimal recommendation mechanism the agent is only told what to do and the set of α such that $a_1(\alpha) = a_1(\alpha_0)$ thus contains at least a manifold of dimension $s - n_1$). Moreover, assume that independently of how w is set and how agent 1 behaves the behavior of agent 2 in problem α is given by $a_2(\alpha)$ where $a_2(\alpha)$ is a smooth function of α that is not locally constant (agent 2's behavior can be rationalized by assuming that $a_2 \neq a_2(\alpha)$ in problem α would result in a huge cost for agent 2). To fix ideas, assume also that $n_1 = 1$ and that at the optimal recommendation mechanism, $a_1 \to E_{\alpha \in \Omega} \pi(a_1, a_2(\alpha); w(\alpha), \alpha)$ is an increasing function of a_1 for a_1 close to the action a_1^* performed by agent 1 when this mechanism is in place.

Consider a small neighborhood $N \subset \Omega$ of $\alpha_0 \in \Omega$. For each $\alpha \in N$, one can consider the set $\overline{B}(\alpha) = \{\gamma \in \Omega \text{ such that } a_2(\gamma) = a_2(\alpha)\}$. For generic α, $\overline{B}(\alpha)$ is a manifold of dimension at least $s - n_1 - n_2$ that intersects $\Omega \setminus \overline{B}(\alpha)$. Call $\alpha^*(\alpha)$ an element of $\Omega \setminus \overline{B}(\alpha)$. Because $a_2(\cdot)$ is not locally constant, one can always find a direction δ tangent to Ω at $\alpha^*(\alpha)$ such that $a_2(\cdot)$ does not remain locally constant along that dimension. Besides, for generic functions u_1 one can find $\alpha^*(\alpha, \varepsilon) \in \Omega$ close to $\alpha^*(\alpha)$, such that $a_2(\alpha^*(\alpha, \varepsilon)) \neq a_2(\alpha_0)$ and[16]

$$\frac{\partial}{\partial a_1}\left[p(\alpha)\overline{u}_1(a_1, \alpha) + p(\alpha^*(\alpha, \varepsilon))\overline{u}_1(a_1, \alpha^*(\alpha, \varepsilon))\right] \mid_{a_1^*} >$$
$$\frac{\partial}{\partial a_1}\left[p(\alpha)u_1(a_1, \alpha) + p(\alpha^*(\alpha, \varepsilon))u_1(a_1, \alpha^*(\alpha, \varepsilon))\right] \mid_{a_1^*}$$

where

$$\begin{aligned} u_1(a_1, \gamma) &= u_1(a_1, a_2(\gamma); w(\gamma), \gamma) \\ \overline{u}_1(a_1, \gamma) &= \frac{p(\alpha)u_1(a_1, a_2(\alpha); w(\gamma), \gamma) + p(\alpha^*(\alpha, \varepsilon))u_1(a_1, a_2(\alpha^*(\alpha, \varepsilon)); w(\gamma), \gamma)}{p(\alpha) + p(\alpha^*(\alpha, \varepsilon))} \end{aligned}$$

The idea then is for each $\alpha \in N$ to bundle α and $\alpha^*(\alpha, \varepsilon)$ into one feedback class,[17] and have all other α forming singleton feedback classes. By integrating over all $\alpha \in \Omega$, it

[16] Generically, the two partial derivatives will be different. If the comparison goes in the wrong way, one can always picks a value of γ symmetrically located on the other side of $\alpha^*(\alpha)$ for which the comparison will be right.

[17] I am implicitly assuming that $\alpha^*(\alpha, \varepsilon)$ are all different as α varies in B, which can easily be ensured by playing on the choice of $\alpha^*(\alpha)$ in $\overline{B}(\alpha)$.

is readily verified that in the coarse feedback case agent 1 will choose a level $\overline{a}_1$ larger but close to a_1^*,[18] which is beneficial to the organization given that $E_{\alpha\in\Omega}\pi(a_1, a_2(\alpha); w(\alpha), \alpha)$ is locally increasing in a_1.

5) In the context of Theorem 1, only the information structure of agent 1 was varied (as agent 2 was assumed to have complete information). If one further imposes that the information (about α) should be public among agents 1 and 2, then the same kind of non-transparency result as in Theorem 1 prevails as long as the dimensionality of α is bigger than the sum of the dimensions of both agents' actions, i.e. $s > n_1 + n_2$. The idea is now to work with the manifold

$$\begin{aligned}\overline{C}(\alpha_0) &= \{\alpha \text{ s.t. } a_1(\alpha) = a_1(\alpha_0) \text{ and } a_2(\alpha) = a_2(\alpha_0)\} \\ &= \overline{A}(\alpha_0) \cap \overline{B}(\alpha_0)\end{aligned}$$

which for generic α_0 has dimension $s - (n_1 + n_2)$.

5 Disclosure policy in low dimensional cases

In the previous section we have seen that when the dimension of α is bigger than the dimension of a_1 or of a_2, some form of non-transparency is always good for the organization. While I think this is the most relevant case for practical problems, I now consider cases in which this is not so and in which full transparency may be the best design for the organization.

[18] This is because in the coarse feedback case, we have that a_1 is perceived to give

$$\begin{aligned}&\int_{\alpha\in N} \left[p(\alpha)\overline{u}_1(a_1, \alpha) + p(\alpha^*(\alpha, \varepsilon))\overline{u}_1(a_1, \alpha^*(\alpha, \varepsilon))\right] d\alpha \\ &+ \int_{\gamma\notin N \text{ and } \gamma\neq\alpha^*(\alpha,\varepsilon),\ \alpha\in N} p(\gamma)u_1(a_1, \gamma)d\gamma\end{aligned}$$

whose derivative at $a_1 = a_1^*$ is strictly positive (yet small) by construction (if one replaces u_1 by $\overline{u}_1$ in the above expression, the derivative is nul given that agent 1 should find optimal to play a_1^* in the original recommendation mechanism).

5.1 When complete information disclosure is best

Consider the authority model of Aghion and Tirole (1997):

$$
\begin{aligned}
u_1(a_1, a_2; \alpha) &= a_2\beta b + (1 - a_2)a_1 b - g_1(a_1) \\
u_2(a_1, a_2; \alpha) &= a_2 B + (1 - a_2)a_1\gamma b - g_2(a_2)
\end{aligned}
$$

in which the sole source of heterogeneity is the parameter γ of congruence and in which $b = B$, $\beta = 1$ and $g_1(a) = g_2(a) = \frac{a^2}{2}$.

I show that no matter what $\gamma_1 < \gamma_2 < ... < \gamma_n$ are, the principal (agent 2) is better off when the agent (agent 1) knows which γ is prevailing rather than when he does not know whether $\gamma = \gamma_1$, γ_2... or γ_n.

Routine calculations yield

$$
\begin{aligned}
a_1^{NE}(\gamma) &= \frac{B(1-B)}{1-\gamma B^2} \\
a_2^{NE}(\gamma) &= \frac{B(1-\gamma B)}{1-\gamma B^2}
\end{aligned}
$$

and when the agent does not know whether $\gamma = \gamma_1$.... or γ_n (while the principal does):

$$
\begin{aligned}
a_1^{CI} &= \frac{B(1-B)}{1-E(\gamma)B^2} \\
a_2^{CI}(\gamma_i) &= B[1 - \gamma_i \frac{B(1-B)}{1-E(\gamma)B^2}]
\end{aligned}
$$

where $E(\gamma)$ is the expected value of γ.

Given the convexity of $\gamma \rightarrow \frac{B(1-B)}{1-\gamma B^2}$, it is readily verified that $E(a_1^{NE}(\gamma)) > a_1^{NE}(E(\gamma)) = a_1^{CI}$. Furthermore, as common sense suggests, agent 2's effort decreases with the degree of congruence in the coarse information case $a_2^{CI}(\gamma_1) > a_2^{CI}(\gamma_2).... > a_2^{CI}(\gamma_n)$, and agent 1's effort increases with the degree of congruence γ in the full information case $a_2^{NE}(\gamma_1) < a_2^{NE}(\gamma_2).... < a_2^{NE}(\gamma_n)$.

The difference of agent 2's expected payoff in the coarse information case and the complete information case writes:

$$\sum_i p(\gamma_i) \max_{a_2}[Ba_2 + B(1-a_2)\gamma_i a_1^{NE}(\gamma_i) - g_2(a_2)] -$$
$$\sum_i p(\gamma_i) \max_{a_2}[Ba_2 + B(1-a_2)\gamma_i a_1^{CI} - g_2(a_2)]$$

It is no smaller than $\sum_i p(\gamma_i)B(1-a_2^{CI}(\gamma_i))\gamma_i(a_1^{NE}(\gamma_i) - a_1^{CI})$ (because in the max appearing in the first summation on can always pick $a_2 = a_2^{CI}(\gamma_i)$ when $\gamma = \gamma_i$, i.e. the argument for the corresponding max in the second summation). This is itself no smaller than

$$\sum_i p(\gamma_i)B(1-a_2^{CI}(\gamma_i))\gamma_i(a_1^{NE}(\gamma_i) - E(a_1^{NE}(\gamma))$$

which is strictly positive given that $\gamma \to (1-a_2^{CI}(\gamma))\gamma$ and $\gamma \to a_1^{NE}(\gamma)$ are both increasing with γ.

To summarize,

Proposition 1 *In the optimal delegation problem with quadratic cost of effort, when the sole heterogeneity is on the congruence parameter γ, full disclosure of γ is always the best policy for the principal.*

The intuition for this result is as follows. Not letting the agent know γ leads him to pick his effort level as a best-response to a mixed distribution of Principal's effort. This in turn leads the agent to make less effort than in the complete information case when the congruence parameter γ is bigger and more effort when it is smaller. But, the Principal would prefer the bias to be the other way round given the implication of the congruence parameter, thereby explaining why full information disclosure is preferable in this case.

Several remarks are in order regarding this proposition. First, it is not in contradiction with Theorem 1 above because the setup analyzed here is one in which the dimensionality of α is the same as the dimensionality of the effort of the agent (so that there is no manifold of strictly positive dimension in the α space in which at the Nash equilibrium, the agent performs the same effort level).[19] Second, even though the result was presented

[19]Reproducing the argument for Theorem 1 with $\alpha_1 = \alpha_0$ would yield that $\Delta(\varepsilon)$ is of the same order

in the context of coarse information, in this case it can equivalently be presented in terms of confusing the agent about the effort level of the principal (agent 2) as in Theorem 2. The reason is that agent 1's payoff function does not directly depend on γ. Thus, in the incomplete information case, agent 1 best-responds to the aggregate distribution of a_2 irrespective of γ, and this is the same outcome as in the feedback equilibrium with the corresponding feedback partition.

5.2 When complete feedback disclosure is best

Consider the following moral hazard in team problem in which agent $i = 1, 2$'s payoff is

$$u_i(a_1, a_2; \beta) = (a_1 + a_2 + \beta a_1 a_2)\, w - \frac{(a_i)^2}{2}$$

and the corresponding profit is

$$\pi(a_1, a_2; \beta) = (a_1 + a_2 + \beta a_1 a_2)\,(R - 2w).$$

Assume that the sole degree of heterogeneity is the complementarity parameter $\beta \in \left[\underline{\beta}, \overline{\beta}\right]$. I simplify the analysis by assuming that the bonus w is not an instrument of the designer and that it is set independently of β and satisfies $w < R/2$.[20]

Given the symmetry between agents 1 and 2, I consider symmetric feedback policies for the two agents. Specifically, let $\beta^1 < \beta^2 .... < \beta^n$ and let p^k denote the probability of β^k. Consider both the case of complete feedback disclosure policy (thereby relying on the Nash equilibrium concept with complete information) and the case of coarse feedback disclosure policy in which every agent $i = 1$ or 2 receives feedback only about the aggregate distribution of effort of agent $-i = 2$ or 1 over $\beta^1, ...\beta^n$, and thus in every problem β^k agents choose their effort level as a best-response to this aggregate effort distribution.

Proposition 2 *The coarse feedback disclosure policy always generates less expected profit to the designer than the complete feedback disclosure policy.*

as ε^2, and thus one would not be able to conclude from the argument given there.

[20] Such an assumption would fit if we have in mind that the bonus w is negotiated after a success is being obtained and the two agents have the same bargaining power.

The intuition for this result is as follows. Confusing the agents about which β prevails when the effort level is being made leads agents to make comparatively more effort when the complementarity parameter is low and less effort when it is large. This is bad for the overall profit because the marginal effect of effort is larger when the complementarity parameter is larger, thereby explaining why the complete feedback disclosure policy dominates in this case. The detailed proof of Proposition 2 appears in Appendix. Observe again that this result is not in contradiction with the insight of Theorem 2 given that here the dimensionality of the problem is equal to the dimensionality of the effort level.[21]

6 Conclusion

In this paper, we have shown that non-transparency both in the form of incomplete information disclosure and in the form of coarse feedback disclosure is optimal in virtual all organizational arrangements of interest. Open questions left for future research are about the optimal form of non-transparency in organizations and when it is more effective to rely on coarse information disclosure or coarse feedback disclosure.

[21] I also expect to get the same type of insights in the case in which agents can be incentivized through an optimally adjusted w.

Appendix

Proof of Proposition 2.

Routine calculations yield that in the full disclosure case agents choose $a^{NE}(\beta) = \frac{w}{1-\beta w}$ when the complementarity parameter is β. In the coarse disclosure case, agents choose $a^{CF}(\beta) = w(1 + \beta\overline{a})$ where $\overline{a}^{CF} = E(a^{CF}(\beta))$ denotes the expected value of the effort level in this case. Thus,

$$\overline{a}^{CF} = \frac{w}{1 - E(\beta)w}$$

where $E(\beta)$ denotes the expected value of β. Given the convexity of $\beta \to \frac{w}{1-\beta w}$, it follows by Jensen's inequality that

$$\overline{a}^{CF} < E(a^{NE}(\beta)). \tag{5}$$

The difference of expected profit in the complete disclosure case and in the coarse disclosure case writes:

$$\begin{aligned}
\Delta/(R-2w) &= \sum\nolimits_i p^i(2a^{NE}(\beta^i) + \beta^i(a^{NE}(\beta^i))^2 - 2a^{CF}(\beta^i) + \beta^i(a^{CF}(\beta^i))^2) \\
&= 2\sum\nolimits_i p^i(a^{NE}(\beta^i) - a^{CF}(\beta^i)) + \sum\nolimits_i p^i\beta^i((a^{NE}(\beta^i))^2 - (a^{CF}(\beta^i))^2)
\end{aligned}$$

We have that $\sum_i p^i(a^{NE}(\beta^i) - a^{CF}(\beta^i)) > 0$ by (5).

Moreover, let $i^* = \arg\min_i \left\{i \text{ such that } a^{NE}(\beta^i) \geq \overline{a}^{CF}\right\}$. Given the monotonicity of $i \to a^{NE}(\beta^i)$, we have that for $i \geq i^*$, $a^{NE}(\beta^i) \geq a^{CF}(\beta^i)$ and for $i < i^*$, $a^{NE}(\beta^i) \leq a^{CF}(\beta^i)$. Writing $(a^{NE}(\beta^i))^2 - (a^{CF}(\beta^i))^2$ as $(a^{NE}(\beta^i) + a^{CF}(\beta^i))(a^{NE}(\beta^i) - a^{CF}(\beta^i))$, making use of the monotonicity of $i \to \beta^i\left(a^{NE}(\beta^i) + a^{CF}(\beta^i)\right)$, and of the change of sign of $a^{NE}(\beta^i) - a^{CF}(\beta^i)$ at $i = i^*$, we get that for all i

$$\beta^i\left((a^{NE}(\beta^i))^2 - (a^{CF}(\beta^i))^2\right) \geq \beta^{i^*}(a^{NE}(\beta^{i^*}) + a^{CF}(\beta^{i^*}))(a^{NE}(\beta^i) - a^{CF}(\beta^i))$$

In turn, this implies that

$$\begin{aligned}
&\sum\nolimits_i p^i\beta^i((a^{NE}(\beta^i))^2 - (a^{CF}(\beta^i))^2) \\
\geq\ &\beta^{i^*}(a^{NE}(\beta^{i^*}) + a^{CF}(\beta^{i^*}))\sum\nolimits_i p^i(a^{NE}(\beta^i) - a^{CF}(\beta^i))
\end{aligned}$$

which again is strictly positive by (5). **Q. E. D.**

References

[1] Aghion, P. and J. Tirole (1997): 'Formal and Real Authority in Organizations,' *Journal of Political Economy*, **105**, 1-29.

[2] Battaglini, M. (2006): 'Joint Production in Teams,' *Journal of Economic Theory*, **130**, 138-167.

[3] Crémer, J. (1995): 'Arm's Length Relationships' The *Quarterly Journal of Economics*, **110**, 275-295.

[4] Dewatripont M., I. Jewitt and J. Tirole (1999a): 'The Economics of Career Concerns, Part I: Comparing Information Structures,' The *Review of Economic Studies*, **66**, Special Issue, 183-198.

[5] Dewatripont M., I. Jewitt and J. Tirole (1999b): 'The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government' The *Review of Economic Studies*, **66**, Special Issue, 199-217.

[6] Elderer, F., R. Holden and M. Meyer (2008): 'Gaming and Strategic Ambiguity in Incentive Provision', mimeo.

[7] Gal-Or, E. (1985): 'Information Sharing in Oligopoly,' *Econometrica* **53**, 329-343.

[8] Holmström, B (1979): 'Moral Hazard and Observability' The *Bell Journal of Economics*, **10**, 74-91.

[9] Holmström, B. (1982): 'Moral Hazard in Teams' The *Bell Journal of Economics*, **13**, 324-340

[10] Holmström, B. and P. Milgrom (1991): 'Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design' *Journal of Law, Economics, and Organization*, 7: 24.

[11] Jehiel, P. (2005): 'Analogy-based Expectation Equilibrium', *Journal of Economic Theory*, **123**, 81-104.

[12] Jehiel, P. (2009): 'Manipulative Auction Design', mimeo.

[13] Mas-Colell, A., M. D. Whinston, and J. R. Green (1995): *Microeconomic Theory* Oxford University Press.

[14] Milgrom, P. and R. Weber (1982): 'A Theory of Auctions and Competitive Bidding,' *Econometrica*, **50**, 1089-1122.

[15] Myerson, R. (1986): 'Multistage Games with Communication,' *Econometrica*, **54**, 323-358.

[16] Prat, A. (2005): 'The Wrong Kind of Transparency,' The *American Economic Review*, **95**, 862-877

[17] Raith, M. (1996): 'A General Model of Information Sharing in Oligopoly,' *Journal of Economic Theory* **71**, 260-288.

[18] Vives, X. (1984): 'Duopoloy Information Equilibrium: Cournot and Bertrand,' *Journal of Economic Theory* **34**, 71-94.