# The Wrong Kind of Transparency

Andrea Prat*
London School of Economics†

December 20, 2001

**Abstract**

In a model of career concerns for experts, when is a principal hurt from observing more information about her agent? This paper introduces a distinction between information on the consequence of the agent's action and information directly on the agent's action. It is the latter kind that can hurt the principal by engendering conformism, which worsens *both* discipline and sorting. The paper identifies a necessary and sufficient condition on the agent signal structure under which transparency on action is detrimental to the principal. The paper also shows the existence of complementarities between transparency on action and transparency on consequence. The results are used to interpret some existing disclosure policies in political and corporate governance.

## 1 Introduction

There is a widespread perception, especially among economists, that transparency is a beneficial element in agency relationships. The more the principal knows about what the agent does, the easier it is for her to evaluate the agent's performance. More precise information, in turn, permits to put in place an effective screening and incentive mechanism. This perception has been confirmed by theoretical results (e.g. Holmstrom [13]) and is consistent with empirical evidence (e.g. Besley and Burgess[1] in politics and Dyck and Zingales [8] in corporate governance).

This would lead one to conclude that transparency ought to be the governing principle in agency relations. Whenever it is technologically feasible, the principal should observe everything that the agent does. However, in practice there are systematic deviations from transparency. In politics, the principle of open government has made great inroads in the last decades but there are still important areas in which public decision-making is, by law, protected by secrecy. In the

*I thank Roman Inderst, Alessandro Lizzeri, Stephen Morris, and Marco Ottaviani for useful discussions.
†STICERD, Houghton Street, London WC2A 2AE, England. Email: a.prat@lse.ac.uk. Homepage: econ.lse.ac.uk/staff/prat.

United States, the "executive privilege" allows the president to withhold information from the Congress, the courts, and the public (Rozell [22]). While the executive privilege cannot be used arbitrarily and fell in disrepute during the Watergate scandal, the Supreme Court recognized its validity (US *vs.* Nixon, 1974). In the European Union, the most powerful legislative body, the Council, has a policy of holding meetings behind closed doors and not publishing the minutes (Calleo [4]).[1]

In corporate governance, violations to the transparency principle are so widespread that some legal scholars argue that secrecy is the norm rather than the exception in the relation between stakeholders and managers (Stevensons [24, p. 6]): "Corporations – even the largest among them – have always been treated by the legal system as 'private' institutions. When questions about the availability of corporate information have arisen, the inquiry has typically begun from the premise that corporations, like individuals, are entitled to keep secret all information they are able to secure physically unless some particular reason for disclosure [...] could be adduced in support of a contrary rule. So deeply embedded in our world view is this principle that it is not at all uncommon to hear serious discussions of a corporate 'right to privacy'."

What are the reasons behind this observed lack of transparency? One obvious candidate explanation is that information revealed to the principal would also be revealed to a third party, and that has negative consequences for the principal. In the political arena, voters may choose to ignore information pertaining to national security to prevent hostile countries from learning them as well. In the corporate world, shareholders may wish to keep non-patentable information secret rather than risk that the competitors learn it.[2]

The "third-party rationale" for keeping information secret presumably entails a tradeoff between damage from information leaks and weaker incentives for the agent. This paper is instead concerned with a rationale for secrecy that stems purely from incentive considerations. The conjecture is that in some circumstances revealing more information about the agent makes the agent's interest less aligned with the principal's interest. Three questions should be asked: (a) Can the principal ever benefit from committing not to observe certain kinds of information about the agent's behavior? (b) If so, for what kind of information and in what circumstances is con-

[1] Over thirty countries have passed Freedom of Information Acts, which establish the principle that a citizen should be able to access any public document. There are, however, important types of informations that are exempt from this requirement (Frankel [11]). In Section 10 we shall examine more carefully the rationale behind some exemptions.

[2] Another possible explanation is that transparency is indeed the optimal policy but the existing institutions are suboptimal. The status quo persists perhaps because of the entrenched interests of the current agents or because of the inertia of complex institutional arrangements. If this explanation is correct, it is then essential to make the case in favor of transparency as strong as possible by studying the validity and the scope of potential objections to the transparency principle. The present paper may be seen as a contribution in that sense.

cealment beneficial? (c) How does transparency on one kind of information relate to transparency on other kinds of information?

In the classical moral hazard problem, question (a) has a negative answer. More information can never hurt the principal. At worst, it is superfluous (Holmstrom [13]). To have a positive answer, one should move to a world in which payments cannot depend on performance and contracts are short-term, that is, Holmstrom's [12] career concerns model. The principal, or a market, does not know the type of the agent and uses the agent's performance in the current period to infer his type and predict his performance in future periods. In turn, this inference process affects the career prospects of the agent and determines his behavior in the current period. There are various ways of modeling career concerns according to what is taken as the type of the agent (more about them is said in the Related Literature section). In this paper, we assume that the type corresponds to the agent's ability of understanding the state of the world, that is, we use a model of career concerns for experts (Scharfstein and Stein [23], Prendergast and Stole [21], Ottaviani and Sorensen [18]).

There are two periods: the current period and the future period. In the current period, an agent (expert) is in charge of taking an action on behalf of the principal. The optimal action depends on the state of the world. The agent receives a signal about the state of the world, whose precision depends on the agent's type. The action, together with the state of the world, determines a consequence for the principal. At the end of the current period, the principal forms a posterior about the agent's type, based on information available, and she decides whether to keep the current agent or replace him with another, randomly drawn, agent. In the future period, the agent who is in charge faces a similar decision problem. The wage of the agent cannot be made contingent on the agent's current performance. The agent maximizes the probability of keeping his job. The principal cares about the consequence in the current period (*discipline* component) and the consequence in the second period, which in turn depends on the ability of the principal to screen agents by type (*sorting* component).

With this simple model, we tackle the three questions above. It is easy to see by means of an example that the answer to question (a) is positive. More importantly, we show that more transparency hurts the principal by reducing the two components of her utility – discipline and sorting – at the same time. Transparency pushes the agent toward conformism, that is, it induces him to make less use of his private signal when making decisions. Conformism makes the agent's decision in the current period less appropriate to the state of the world (negative discipline effect) and lowers the ability of the principal to tell the agent's type because performance is less informative (negative sorting effect).

The answer to question (b) hinges on the distinction between information on the consequence of the agent's action and information directly on the agent's action. The possibility that more transparency generates conformism regards the latter kind of information. First note that, even if the principal knows the consequence of the agent's action perfectly, she still stands to gain from knowing the action. Suppose, for instance, that the principal is a firm owner, the agent is the firm manager, the action is the firm strategy, the state of the world is the business environment, and the consequence is profit. The only direct element in the principal's utility is current and future profit. Yet, there are multiple combinations of environnment and strategy that lead to the same profit. Knowing which of the combinations occurred helps the principal understand the type of the agent.

Direct information on the agent's action thus has a potential positive sorting effect. This effect, however, is based on the assumption that the agent's behavior is constant. Instead, if the principal moves from a policy of keeping the action concealed to one of full transparency, the agent understands that his incentive structure is now changed because he is now judged directly on the action he takes. A crucial observation is that, in a generic model, the possible realizations of the agent's signal can be ranked in order of *virtuosity*, that is, according to the posterior on the agent's type given the realization of the signal. If in equilibrium the agent's action is informative of his signal, then also all the possible actions can be ranked in order of virtuosity. The posterior on the agent's type depends on the consequence but also on the virtuosity of the action. This creates a contradiction. If the virtuosity component is too strong, the only possible equilibrium is one in which actions cannot be ranked in order of virtuosity, i.e. an uninformative equilibrium. The agent disregards his private signal and acts in a purely conformist way. If this is the case, the principal is clearly better off committing to keep the action concealed.

This leads to a necessary and sufficient condition under which revealing the agent's action leads to conformism. The condition has to do with the relative virtuosity of signals. If some signal is much more virtuous than the others, then the chain of negative effects describe above takes place and there are only conformist equilibria. In mathematical terms, the condition is expressed as a bound on the relative informativeness of the different realizations of the agent's signal.

This condition implies that the more advantageous it is for the principal to commit to concealment ex ante, the more advantageous it is for her to renege on her commitment ex post and observe the agent's actionfor sorting purposes. This Catch-22 result points to the importance of developing credible institutions to guarantee the agent's privacy.

Lastly, the paper answers point (c) by showing that there is a complementarity between transparency on action and transparency on consequence. To each (exogenously given) probability that the consequence is observed corresponds an optimal probability that the action is observed.

We show that The optimal probaility that action is observed is nondecreasing in the probability that the consequence is observed. The implication that transparency on action goes hand in hand with transparency on consequence is of practical relevance. Indeed, we discuss its explanatory power with regards to the provisions of transparency policies in several countries.

The plan of the paper is as follows. Section 2 introduces the career concern game and shows how it can be interpreted as the reduced form of two economic situations. Section 3 provides a simple example that shows when conformism arises. Sections 4 and 5, which form the core of the paper, characterize the set of perfect Bayesian equilibria with concealed action and revealed action and identify the necessary and sufficient condition under which concealed action is optimal. Section 6 discusses the relation between the ex ante desire to conceal the action and the ex post incentive to renege on the promise and observe the agent's action. Section 7 studies the complementarity between action observation and consequence observation. Section 10 concludes by using the results of the paper to interpret some existing institutional arangements in political and corporate governance.

## 1.1 Related literature

There are a few works that relate to the optimality of information revelation in agency problems. In "classical" moral hazard principal-agent problems, the question has been resolved by Holmström [13]. Observing an additional signal can never hurt the principal and it is strictly beneficial if and only if the principal does not already observe a signal that is a sufficient statistic for the additional signal.

On the contrary, in career concerns there are already examples in which more information about the agent's behavior hurts the principal. There are three main approaches to model career concerns, depending on whether the agent's type is seen as ability to exert effort, congruence of preference with the principal, or ability to observe a signal about the state of the world. For the first approach, the question of comparing information structures is studied in a general way by Dewatripont, Jewitt and Tirole [7]. They first present two examples in which a more precise signal about the agent's performance reduces discipline. They then find general sufficient conditions, similar in spirit to Holmström, under which an additional signal increases effort.

For the second approach, the question of comparing information structures is briefly discussed by Morris [15, p 18-19]. There, an agent observes a signal about the state of the world and makes a report to the principal. The principal makes a decision after hearing the agent's report. Then, the state of the world is revealed. A market then forms a posterior on the basis of the agent's report and on the observation of the state. Morris compares this situation with the situation in which the market observes neither the signal nor the state (because principals are short-lived).

He shows that observing the state and the signal improves sorting and may improve or worsens the current period decision: while the bad type's decision is more aligned with the principal's preference, the good type may be induced to take an extreme action to separate himself from the bad type (the "political correctness" effect).

The third approach – the expert agent model (Scharfstein and Stein [23], Prendergast and Stole [21], Ottaviani and Sorensen [17] [18]) – is the one that is used here. As far as I know, there is no paper in this stream attempting a comparison of information structures. In those papers, it is typically assumed that the principal (or the market) observes the agent's action. In Prendergast and Stole [21], the agent's action – the investment decision – is publicly observed. In Ottaviani and Sorensen [18], the agent's "action" is the message that the expert sends to the evaluator and it is, by definition, observed.[3]

Prendergast [20] analyzes an agency problem in which the agent exerts effort to observe a variable which is of interest to the principal. The principal too receives a signal about the variable and the agent receives a signal about the signal that the principal received. This is not a career concern model and the principal can offer payments conditional on the agent's report. Prendergast shows that the agent uses his information on the principal's signal to biases his report toward the principal's signal. Misreporting on the part of the agent causes a loss of efficiency.

Fingleton and Raith [10] study career concerns for delegated bargaining when the type of bargainers determine their ability of understanding the opponent's valuation. They ask whether bargaining behind closed doors is better or worse from the viewpoint of the principal. If bargaining occurs secretly, the principal is not able to observe offers but only acceptances. Thus, also their paper questions the optimality of transparency in expert models. However – besides the fact that their model is developed in a context of bargaining – their distinction between acceptance and offer does not correspond to our distinction between consequence and action (if the offer is accepted, everything is observed).

Crémer [6] shows that, in a two-period agency model where renegotiation is possible, the principal may be hurt by a decrease in the cost of observing the agent's performance. This is because improving the ex post information of the principal makes a commitment not to renegotiate less credible.

[3]In Ottaviani and Sorensen [18], it is immaterial to think about the agent's decision as message transmitted to the principal or an action taken on behalf of the principal. This is not true anymore in the present model when the action is not observed.

# 2 Model

We write the agency problem in a detail-free reduced form. This form corresponds to two economic situations, one in which the bargaining power is on the principal side, the other in which it is more on the agent side. We first present the reduced form, and we then examine how it can be derived from the two "expanded forms".

## 2.1 Reduced form

There are a principal and an agent. The agent's type $\theta \in \{g, b\}$ is unknown to both players. The probability that $\theta = g$ is $\gamma \in (0,1)$ and it is common knowledge. The state of the world is $x \in \{0,1\}$ with $\Pr(x = 1) = p \in (0,1)$. The random variables $x$ and $\theta$ are mutually independent. The agent selects an *action* $a \in \{0,1\}$. The *consequence* $u(a,x)$ is 1 if $a = x$ and 0 otherwise.

The principal ignores the state of the world. The agent receives a private signal $y \in \{0,1\}$ that depends on the state of the world and on his type. Let

$$q_{y\theta} = \Pr(y = 1|x, \theta).$$

We assume that $0 < q_{1g} < q_{1b} < q_{0b} < q_{0g} < 1$. To make the problem interesting, we also assume that

$$q_{1b}p - q_{0b}(1-p) \in (0,1). \tag{1}$$

This guarantees that, for any $\gamma$, the signal $y$ is decision-relevant.

The mixed strategy of the agent is a pair $\alpha = (\alpha_0, \alpha_1) \in [0,1]^2$, which represents the probability that the agent plays $a = 1$ given the two possible realizations of the signal.

We consider two cases: *concealed* action and *revealed* action. In the first case, the principal observes only the consequence $u$. In the second case, she observes also the action $a$ (and therefore she can infer $x$). The principal's posterior probability that the agent's type is $g$ is $\pi(I)$, where $I$ is the information available to the principal. With concealed action, the posterior is

$$\tilde{\pi}(u) = \Pr(\theta = g|u) = \frac{\gamma \Pr(a = x|\alpha, x, \theta = g)\Pr(x)}{\Pr(a = x|\alpha, x)\Pr(x)}.$$

In term of primitives,

$$\tilde{\pi}(u=1) = \frac{\gamma\left(p\left(\alpha_1 q_{1g} + \alpha_0(1-q_{1g})\right) + (1-p)\left((1-\alpha_1)q_{0g} + (1-\alpha_0)(1-q_{0g})\right)\right)}{\sum_{\theta\in\{b,g\}}\Pr(\theta)\left(p\left(\alpha_1 q_{1\theta} + \alpha_0(1-q_{1\theta})\right) + (1-p)\left((1-\alpha_1)q_{0\theta} + (1-\alpha_0)(1-q_{0\theta})\right)\right)} \tag{2}$$

$$\tilde{\pi}(u=0) = \frac{\gamma\left(p\left((1-\alpha_1)q_{1g} + (1-\alpha_0)(1-q_{1g})\right) + (1-p)\left(\alpha_1 q_{0g} + \alpha_0(1-q_{0g})\right)\right)}{\sum_{\theta\in\{b,g\}}\Pr(\theta)\left(p\left((1-\alpha_1)q_{1\theta} + (1-\alpha_0)(1-q_{1\theta})\right) + (1-p)\left(\alpha_1 q_{0\theta} + \alpha_0(1-q_{0\theta})\right)\right)} \tag{3}$$

With revealed action, the agent's posterior, assuming that $a$, is played in equilibrium with positive probability, is

$$\pi(a,x) = \Pr(\theta = g|a,x) = \frac{\gamma \Pr(a,x|\theta = g)\Pr(x)}{\Pr(a,x)\Pr(x)}.$$

In terms of primitives,

$$\begin{aligned}\pi(1,x) &= \frac{(\alpha_1 q_{xg} + \alpha_0(1-q_{xg}))\gamma}{(\alpha_1 q_{xg} + \alpha_0(1-q_{xg}))\gamma + (\alpha_1 q_{xb} + \alpha_0(1-q_{xb}))(1-\gamma)};\\ \pi(0,x) &= \frac{((1-\alpha_1)q_{xg} + (1-\alpha_0)(1-q_{xg}))\gamma}{((1-\alpha_1)q_{xg} + (1-\alpha_0)(1-q_{xg}))\gamma + ((1-\alpha_1)q_{xb} + (1-\alpha_0)(1-q_{xb}))(1-\gamma)}.\end{aligned}$$

If action $a$ is not played in equilibrium, perfect Bayesian equilibrium imposes no restriction on $\pi(a,x)$.

The payoff to the agent is simply the posterior $\pi(I)$. The payoff to the principal depends on the consequence and on the posterior: $u(a,x) + v(\pi(I))$, where $v$ is a convex function of $\pi$. Given any equilibrium strategy $\alpha^*$, the ex ante expected payoff of the agent must be $\gamma$, while the ex ante expected payoff of the principal is $w(\alpha^*) = E_{a,x}(u(a,x) + v(\pi(I))|\alpha^*)$. As the agent's expected payoff does not depend on $\alpha^*$, the expected payoff of the principal can also be taken as total welfare. Let $W_{revealed}$ be the highest $w(\alpha^*)$ that can be achieved when the action is revealed, and $W_{concealed}$ the corresponding value when the action is concealed. The main question that we shall ask is whether $W_{revealed} > W_{concealed}$.

Attention should be drawn to two assumptions. First, assuming that the agent maximizes the posterior $\pi(I)$, rather than an arbitrary function of the posterior $\pi(I)$, is not without loss of generality (see Ottaviani and Sorensen [18] for a discussion of this point). As we shall see, the assumption is arbitrary in Expanded Form I but it is somewhat more natural in Form II. The assumption is made by most papers in career concerns because it makes the analysis simpler. Second, the agent does not know his own type (again, Ottaviani and Sorensen [18] discuss this point). If the agent knew his own type, he could use his action choice as a costly signal of how confident he is of his own information. This additional signalling component would make results harder to interpret.[4]

## 2.2 Expanded form I: competing agents.

This form is suited to represent a political game, in which agents are competing parties or candidates and the principal is the electorate (see Persson and Tabellini CITE for a discussion of

[4] Also, if the agent knew $\theta$, his type – in the sense of his private information – would be given by $(\theta, y)$. As the type has now four realizations but the action is still binary, we would be imposing an artificial restriction on the agent's message space. We should, hence, modify the game by allowing the agent to transmit another binary message together with his action choice.

retrospective voting models). In this two-period model, there are two agents and one principal. One agent, the *incumbent*, is available in the first period. The second, the *challenger*, appears at the end of the first period. The type of the incumbent is $\theta \in \{g, b\}$, where the probability that $\theta = g$ is $\gamma$. The principal ignores the type of the agent. Following most of the literature on career concerns, we assume that also the agent ignores his own type. The type of the challenger is $\theta_c \in \{g, b\}$, where the probability that $\theta = g$ is $\gamma_c$. While $\gamma$ is known, $\gamma_c$ is itself a stochastic variable with distribution $f$, which is revealed at the end of the first period.

In the first period, the incumbent is in charge of a binary decision $a \in \{0, 1\}$. The state of the world is $x \in \{0, 1\}$. The agent observes a signal $y \in \{0, 1\}$ according to the conditional probability $q$ described above. The assumption in (1) guarantees that, even in the worst-case scenario (when it is learnt that the agent is for sure a bad type), the signal $y$ is decision-relevant. Without this assumption, it may be the case that second-period efficient decision making requires choosing the same action independently of the signal. The consequence $u$ is 1 if the action matches the state and zero otherwise.

At the end of the first period the challenger appears and $\gamma_c$ is learnt. The principal observes the consequence, and possibly the action as well. She then chooses whether to keep the incumbent or replace him with the challenger.

In the second period, the agent that has been retained faces a decision problem that is similar to the first period. He selects action $\hat{a} \in \{0, 1\}$ to match state $\hat{x} \in \{0, 1\}$, where the probability that $\hat{x} = 1$ is still $p$. The agent receives $\hat{y}$ a signal about $\hat{x}$ that is distributed according to $q_{y\theta}$ described above. The consequence $\hat{u}$ is 1 if the action matches the state and zero otherwise.

The payoff to the principal is $u + \delta\hat{u}$, where $\delta \in (0, \infty)$, which captures both the discount rate and the relative importance of the two periods. A $\delta > 1$ occurs when the second period is more important than the first. The payoff to each agent is 1 if he is hired for the second period and zero otherwise (the benefit that the incumbent receives in the first period is normalized to zero). Clearly, this model describes a world of very incomplete contracts. An agent who is hired gets a fixed rent that the principal cannot control. In particular, the principal cannot offer transfers that are conditional on observed performance.

We assume that the "prior" on the challenger's type, $\gamma_c$, is uniformally distributed on the unit interval – that is $f_c$ is a uniform distribution with support $(0, 1)$. This restriction guarantees that the payoff of the incumbent is linear in the posterior.

To summarize, the timing is as follows:

1. The incumbent observes signal $y$ and selects action $a$.

2. The consequence $u$ is realized. The challenger appears: his prior $\gamma_c$ is realized and observed by all. In the concealed action case, the principal observes $u$. In the revealed action case,

the principal observes $a$ and $u$. The principal forms a posterior $\pi$ on the incumbent's type and chooses between the incumbent and the challenger.

3. The agent that has been retained observes signal $\hat{y}$ and selects action $\hat{a}$.

4. The consequence $\hat{u}$ is realized.

We start by analyzing the two last stages, which are straightforward. In the second period, the agent that is retained has no career concerns and he is indifferent with regards to the action he takes. Thus, any strategy is a continuation equilibrium. In line with the rest of the literature on career concerns, we assume that an indifferent agent acts in the interest of the principal. Given (1), it is easy to see that, independently of his belief on his own type, the agent selects $\hat{a} = \hat{y}$. Let $\hat{\gamma}$ be the probability that the agent that is retained for the second period is good, as computed by the principal at the beginning of the second period ($\hat{\gamma} = \pi$ if the incumbent is confirmed, $\hat{\gamma} = \gamma_c$ if the challenger is hired). The second-period expected utility of the principal is

$$\begin{aligned} E(\hat{u}|\hat{\gamma}) &= \Pr(\hat{y} = \hat{x}|\hat{\gamma}) = \hat{\gamma}\left((1-p)(1-q_{0g}) + pq_{1g}\right) + (1-\hat{\gamma})\left((1-p)(1-q_{0b}) + pq_{1b}\right) \\ &= (1-p)(1-q_{0b}) + pq_{1b} + \hat{\gamma}\left((1-p)(q_{0b} - q_{0g}) + p(q_{1g} - q_{1b})\right) \qquad (4) \\ &= \bar{Q} + Q\hat{\gamma}. \end{aligned}$$

Thus, $E(\hat{u}|\hat{\gamma})$ is linear and increasing in $\hat{\gamma}$. This means that the principal chooses to retain the agent with the higher probability of being a good type. Therefore, $\hat{\gamma} = \max\{\pi, \gamma_c\}$.

This determines how the players' payoff functions depend on the posterior $\pi$. The expected payoff for the incumbent is his probability of being retained in the second period. Before $\gamma_c$ is realized, his payoff is then linear in $\pi$, which corresponds to the assumption in the reduced form. The expected payoff of the principal, when she knows $\pi$ but she still ignores $\gamma_c$, is a linear function of $\max\{\pi, \gamma_c\}$. Therefore, it is a convex function of $\pi$.

### 2.3 Expanded form II: competing principals

This form could be taken as a simple representation of a market for highly qualified labor. The principals are firms who compete to hire an agent with a unique talent. Note that if the three firms were identical, it would not matter from an efficiency point of view which firm hires the agent. Hence, sorting would play no role. Thus, to make the problem interesting we need to assume an asymmetry among firms.

There are three principals ($A$, $B$, $C$) and one agent. Again, there are two periods. In the first period there is only principal $A$. As before $a$, $x$, $y$, and $u$ denote first-period variables. In the second period, the three principals compete to hire the agent. A principal who does not hire the agent gets a payoff of zero. Principals $B$ and $C$ are "small". They do not incur fixed costs and their payoff is 1 fi the consequence matches the state and zero otherwise. Principal $A$ is a large principal. In order to become active in the first period, she has to pay an upfront cost $f \in (0,1)$. If the action matches the state she gets 2. Otherwise she gets zero.

Timing is as follows:

1. The first-period state $x$ is realized. The agent works for Principal $A$. He observes $y$ and chooses $a$.

2. The consequence $u$ is observed by everyone. In the revealed action case, also $a$ is observed. Each principal makes a wage offer to the agent.

3. The agent chooses one of the three principals. The second-period state $x$ is realized. The agent observes $\hat{y}$ and chooses $\hat{a}$. The consequence for the principal who hired the agent is $\hat{u} = 1$ if $\hat{a} = \hat{x}$ and zero otherwise. If the principal is $A$, she receives $2\hat{u} - f$. If the principal is $B$ or $C$, she receives $\hat{u}$.

As before, we focus attention on equilibria in which, whenever indifferent, the agent chooses his action in order to maximize the payoff of the principal who hired him. In the second period, the expected payoff of the principal who hires the agent is, similarly to (4), a linear function of the posterior of the agent $\pi$: $\bar{Q} + Q\pi$. In the bidding game at stage 2, Principal $A$ is willing to pay up to $2\left(\bar{Q} + Q\pi\right) - f$, while the other two principals are willing to pay up to $\bar{Q} + Q\pi$. Excluding dominated strategies, the equilibrium bid is $\bar{Q} + Q\pi$. Principal $A$ hires the agent if and only if

$$\pi \geq \frac{f - \bar{Q}}{Q}.$$

The expected payoff of $A$ given $\pi$ is $\max\left(\bar{Q} + Q\pi - f, 0\right)$. Thus, her expected payoff is convex in the agent's posterior. The agent's payoff is instead just the equilibrium bid $\bar{Q} + Q\pi$, and it is therefore linear in the posterior.

## 2.4 Preliminaries

We introduce two notions that will be used extensively in the rest of the paper. We say that the agent signal is *decision-relevant* if

$$\begin{aligned} \Pr\left(x = 1 | y = 1\right) &\geq \frac{1}{2}; \\ \Pr\left(x = 0 | y = 0\right) &\geq \frac{1}{2}. \end{aligned}$$

The concept of decision-relevant signal corresponds to the following mental experiment. Suppose the principal observes $y$ directly and chooses $a$ herself. Would the decision rule $a = y$ be optimal?

With some re-working, the two inequalities that define decision-relevance can be rewritten in terms of the primitives of the model:

$$\frac{q_{0g}\gamma + q_{0b}(1-\gamma)}{q_{1g}\gamma + q_{1b}(1-\gamma)} \leq \frac{p}{1-p} \leq \frac{1 - q_{0g}\gamma - q_{0b}(1-\gamma)}{1 - q_{1g}\gamma - q_{1b}(1-\gamma)}. \tag{5}$$

Decision relevance puts a lower and an upper bound to the prior $p$. If one of the two states is extremely likely to occur and/or the signal $y$ is weak, the decision-maker should disregard the signal and just match the action to the more likely state.

The second notion corresponds to another mental experiment. Suppose the principal could observe the agent signal $y$ directly. Which of the two realizations of the signal is better news about the agent type? This corresponds to comparing $\Pr(\theta = 1|y = 1)$ with $\Pr(\theta = 1|y = 0)$.

We exclude the nongeneric case in which the two probabilities are identical. In such a situation, the posterior about the agent must be equal to the prior and the signalling game is uninsteresting. $\Pr(\theta = 1|y = 1) > \Pr(\theta = 1|y = 0)$ we say that $y = 1$ is the *virtuous realization* of the agent signal (or, with a slight abuse, the virtuous signal). If $\Pr(\theta = 1|y = 1) < \Pr(\theta = 1|y = 0)$, we say that $y = 0$ is the virtuous realization. The following result relates virtuosity to the primitives:

**Proposition 1** *The virtuous signal is $y = 1$ if and only if*

$$\frac{q_{0b} - q_{0g}}{q_{1g} - q_{1b}} < \frac{p}{1-p}.$$

**Proof.** Note that:

$$\begin{aligned}
\Pr(\theta|y) &= \frac{\Pr(y|g)\Pr(g)}{\Pr(y|g)\Pr(g) + \Pr(y|b)\Pr(b)} \\
&= \frac{(\Pr(y|1,g)\Pr(1) + \Pr(y|0,g)\Pr(0))\Pr(g)}{(\Pr(y|1,g)\Pr(1) + \Pr(y|0,g)\Pr(0))\Pr(g) + (\Pr(y|1,b)\Pr(1) + \Pr(y|0,b)\Pr(0))\Pr(b)}
\end{aligned}$$

$$\begin{aligned}
\Pr(\theta &= 1|y = 1) = \frac{(q_{1g}p + q_{0g}(1-p))\gamma}{(q_{1g}p + q_{0g}(1-p))\gamma + (q_{1b}p + q_{0b}(1-p))(1-\gamma)} \\
\Pr(\theta &= 1|y = 0) = \frac{((1-q_{1g})p + (1-q_{0g})(1-p))\gamma}{((1-q_{1g})p + (1-q_{0g})(1-p))\gamma + ((1-q_{1b})p + (1-q_{0b})(1-p))(1-\gamma)}
\end{aligned}$$

Then, $\pi(y = 1)$ is greater than $\pi(y = 0)$ if and only if

$$\frac{(q_{1g}p + q_{0g}(1-p))\gamma}{(q_{1g}p + q_{0g}(1-p))\gamma + (q_{1b}p + q_{0b}(1-p))(1-\gamma)} > \gamma$$

or

$$\begin{aligned}
q_{1g}p + q_{0g}(1-p) &> (q_{1g}p + q_{0g}(1-p))\gamma + (q_{1b}p + q_{0b}(1-p))(1-\gamma) \\
q_{1g}p + q_{0g}(1-p) &> q_{1b}p + q_{0b}(1-p) \\
(q_{1g} - q_{1b})p &> (q_{0b} - q_{0g})(1-p) \\
\frac{q_{0b} - q_{0g}}{q_{1g} - q_{1b}} &< \frac{p}{1-p}
\end{aligned}$$

■

If the two states of the world are equiprobable, Proposition 1 requires that

$$q_{1g} - q_{1b} > (1 - q_{0g}) - (1 - q_{0b}).$$

That is, the difference between the probability that the good type gets the right signal and the probabilty that the bad type gets the right signal must be greater if $x = 1$ than if $x = 0$. Then, observing $y = 1$ raises the agent's posterior above $\gamma$ while observing $y = 0$ decreases it.

If the two states have different probability, then the inequality is:

$$p(q_{1g} - q_{1b}) > (1-p)((1 - q_{0g}) - (1 - q_{0b})).$$

## 3 An Example

Suppose that $\gamma = \frac{1}{2}$, $p = \frac{1}{2}$, $q_{0b} = q_{1b} = \frac{1}{2}$, $q_{0g} = \frac{1}{2}$, and $q_{1g} = 1$. A bad agent receives a purely noisy signal. A good agent observes the state $x = 1$ with certainty and gets noise if the state is $x = 0$. This could correspond to a situation in which $x = 0$ is "business as usual" and $x = 1$ is a changed world. The agent's ability matters when the world changes. A good agent is able to spot a change but a bad one is not. It is easy to check that the agent signal is decision relevant and that $y = 1$ is the virtuous realization.

This asymmetry between signals creates a problem. The signal $y = 0$ is bad news for the ability of the agent. More likely than not the agent is one that cannot spot changes in the world. If the agent reports his signal truthfully, the principal will use it against him. This leads to conformity: the agent has an incentive to tell the principal that the world has changed whether or not he actually thinks so.

We now verify (in an informal way) that in this example the only equilibrium behavior is the principal tries to observe the action is pure conformity.

Suppose that there exists an informative equilibrium. Intuitively, we expect that if $y = 1$ the agent chooses $a = 1$ ($\alpha_1 = 1$) and if $y = 0$ he either chooses $a = 0$ or he randomizes ($\alpha_0 \in [0, 1)$). We will focus on this kind of informative equilibria and show that they cannot exist. To be entirely accurate, we should show that no informative equilibrium can exist, but that will be left for the formal results.

The principal's belief $\pi(a,x)$ in such an informative equilibrium is:

$$\begin{array}{ll} \pi(1,1) = \frac{2}{3+\alpha_0}, & \pi(1,0) = \frac{1}{2} \\ \pi(0,1) = 0 & \pi(0,0) = \frac{1}{2} \end{array}$$

The belief when $a=1$ dominates the one when $a=0$, in the sense that for any realization of $x$, $\pi(1,x) \geq \pi(0,x)$. Hence, for any $\alpha_0$ it must be that

$$E(\pi(1,x)|y=0) > E(\pi(0,x)|y=0),$$

which means that, in any informative equilibrium, the agent who observes $y=0$ has a strict incentive to report conform to $a=1$ – a contradiction. The only possible equilibrium is then a pooling equilibrium in which no information is revealed. It is easy to check the existence of such equilibria and that the principal is indifferent among them (because $x=1$ and $x=0$ are equiprobable). In a pooling equilibrium both the discipline and the screening effect are absent. The principal gets an expected utility of $\frac{1}{2}$ in the first term and $\frac{5}{8}$ in the second (in the second term agents do use their signal to make a choice – see below for how to arrive to $\frac{5}{8}$).

If instead the principal commits not to observe $a$, there exists an informative equilibrium in which the agent follows his signal. He plays $a=1$ if and only $y=1$. In that case the posterior $\tilde{\pi}(u)$ is

$$\begin{aligned} \tilde{\pi}(1) &= \frac{\Pr(g)\left(\Pr(y=1|g,x=1)+\Pr(y=0|g,x=0)\right)}{\Pr(y=1|x=1)+\Pr(y=0|x=0)} = \frac{\frac{1}{2}\left(1+\frac{1}{2}\right)}{\frac{3}{4}+\frac{1}{2}} = \frac{3}{5}; \\ \tilde{\pi}(0) &= \frac{\Pr(g)\left(\Pr(y=0|g,x=1)+\Pr(y=1|g,x=0)\right)}{\Pr(y=0|x=1)+\Pr(y=1|x=0)} = \frac{\frac{1}{2}\left(0+\frac{1}{2}\right)}{\frac{3}{4}+\frac{1}{2}} = \frac{2}{5}. \end{aligned}$$

Thus, the agent chooses $a$ to maximize the expected value of $u$ which implies $a=y$. The probability that the principal gets utility 1 in the first period is

$$\begin{aligned} \Pr(u=1) &= \frac{1}{4}\Pr(u=1|g,x=0)+\frac{1}{4}\Pr(u=1|g,x=1)+\frac{1}{4}\Pr(u=1|b,x=0)+\frac{1}{4}\Pr(u=1|b,x=0) \\ &= \frac{1}{4}\frac{1}{2}+\frac{1}{4}1+\frac{1}{4}\frac{1}{2}+\frac{1}{4}\frac{1}{2}=\frac{5}{8}. \end{aligned}$$

The principal kicks out the agent if and only if $u=0$, thus with probability $\frac{3}{8}$. If the agent is kicked out, the second term is like the first and in the expected utility is again $\frac{5}{8}$. If the agent is kept, it means that the posterior is $\frac{3}{5}$ and the probability of a high utility is

$$\begin{aligned} \Pr(u=1) &= \frac{3}{10}\Pr(u=1|g,x=0)+\frac{3}{10}\Pr(u=1|g,x=1)+\frac{2}{10}\Pr(u=1|b,x=0)+\frac{2}{10}\Pr(u=1|b,x= \\ &= \frac{3}{10}\frac{1}{2}+\frac{3}{10}1+\frac{2}{10}\frac{1}{2}+\frac{2}{10}\frac{1}{2}=\frac{13}{20}. \end{aligned}$$

To sum up, if the principal commits to observe only $u$ (and assuming the separating equilibrium arises), she gets a double benefit. The discipline effect increases the expected utility in the first period of $\frac{1}{10}$. The screening effect increases the expected payoff in the second period of $\frac{1}{40}$.

# 4 Concealed Action

The principal can only observe the consequence of the agent's action. The posterior is then $\pi(u) = \Pr(\theta = g|u)$. The agent observes his signal $y$ and maximizes $E_x[\pi(u(a,x))|y]$.

The main result is that the existence of a separating equilibrium.depends on whether the agent signal is decision relevant:

**Proposition 2** *If the agent signal is decision-relevant, there exists a separating equilibrium. If it is not decision relevant there exists no informative equilibrium.*

**Proof.** For the first part, consider a separating equilibrium in which $a = y$. The posterior is

$$\begin{aligned}\tilde{\pi}(u=1) &= \frac{\gamma(pq_{1g} + (1-p)(1-q_{0g}))}{\gamma(pq_{1g} + (1-p)(1-q_{0g})) + (1-\gamma)(pq_{1b} + (1-p)(1-q_{0b}))};\\ \tilde{\pi}(u=0) &= \frac{\gamma((1-p)q_{1g} + p(1-q_{0g}))}{\gamma((1-p)q_{1g} + p(1-q_{0g})) + (1-\gamma)((1-p)q_{1b} + p(1-q_{0b}))}.\end{aligned}$$

The decision relevance inequalities (5) can be rewritten as

$$\begin{aligned}\gamma(pq_{1g} + (1-p)(1-q_{0g})) &\geq (1-\gamma)(pq_{1b} + (1-p)(1-q_{0b}));\\ (1-\gamma)((1-p)q_{1b} + p(1-q_{0b})) &\leq \gamma((1-p)q_{1g} + p(1-q_{0g})),\end{aligned}$$

which shows that $\tilde{\pi}(u=1) \geq \gamma \geq \tilde{\pi}(u=0)$. In a separating equilibrium in which $a = y$, the agent maximizes the probability that $u = 1$. Given the agent's goal, there exists a separating equilibrium if and only if

$$\begin{aligned}\Pr(u=1|a=1, y=1) &\geq \Pr(u=1|a=0, y=1);\\ \Pr(u=1|a=0, y=0) &\geq \Pr(u=1|a=1, y=0).\end{aligned}$$

But these inequalities correspond, once again, to decision relevance. Thus, we have shown that if the agent signal is decision-relevant, there exists a separating equilibrium in which $a = y$.

For the second part, assume that one of the decision-relevance inequalities in (5) is violated and suppose that there exists an informative equilibrium. The posterior is as in (2) and (3):

$$\begin{aligned}\tilde{\pi}(u=1) &= \frac{\gamma(p(\alpha_1 q_{1g} + \alpha_0(1-q_{1g})) + (1-p)((1-\alpha_1)q_{0g} + (1-\alpha_0)(1-q_{0g})))}{\sum_{\theta\in\{b,g\}} \Pr(\theta)(p(\alpha_1 q_{1\theta} + \alpha_0(1-q_{1\theta})) + (1-p)((1-\alpha_1)q_{0\theta} + (1-\alpha_0)(1-q_{0\theta})))}\\ &= \gamma\frac{(p\alpha_1 + (1-p)(1-\alpha_1)) + (p(\alpha_0-\alpha_1)(1-q_{1g}) + (1-p)(\alpha_1-\alpha_0)(1-q_{0g}))}{p\alpha_1 + (1-p)(1-\alpha_1) + \sum_{\theta\in\{b,g\}} \Pr(\theta)(p(\alpha_0-\alpha_1)(1-q_{1\theta}) + (1-p)(\alpha_1-\alpha_0)(1-q_{0\theta}))}\\ &= \gamma\frac{(p\alpha_1 + (1-p)(1-\alpha_1)) + (\alpha_0-\alpha_1)(p(1-q_{1g}) - (1-p)(1-q_{0g}))}{p\alpha_1 + (1-p)(1-\alpha_1) + (\alpha_0-\alpha_1)\sum_{\theta\in\{b,g\}} \Pr(\theta)(p(1-q_{1\theta}) - (1-p)(1-q_{0\theta}))}.\end{aligned}$$

Note that

$$(\alpha_0 - \alpha_1)(p(1-q_{1g}) - (1-p)(1-q_{0g})) \neq (\alpha_0 - \alpha_1)(p(1-q_{1b}) - (1-p)(1-q_{0b})),$$

because in an informative equilibrium $\alpha_0 \neq \alpha_1$ and we have excluded the nongeneric case

$$\frac{q_{0b} - q_{0g}}{q_{1g} - q_{1b}} = \frac{p}{1-p}.$$

This shows that $\tilde{\pi}(u=1) \neq \gamma$. Either $\tilde{\pi}(u=1) > \tilde{\pi}(u=0)$ in which case the agent maximizes the probability that $u = 1$ or $\tilde{\pi}(u=1) < \tilde{\pi}(u=0)$ in which case the agent minimizes the probability that $u = 1$. In both cases, the fact that the agent signal is not decision-relevant means that the agent chooses the same action independently of $y$ — a contradiction. ■

The analysis of the concealed action case is straightforward. If the agent signal is decision relevant, there exists a separating equilibrium in which the agent follows his signal and the principal put a higher posterior on an agent who obtains $u = 1$ than on one who fails. This equilibrium does not exist if the signal is not decision-relevant. If the posterior were higher for an agent who succeeds, then the agent would maximize the probability that $u = 1$. As $y$ is useless, the agent would always choose the same action, which creates a contradiction.

Thus, if the signal is not decision-relevant, the only possible equilibria are uninformative. The best equilibrium from the viewpoint of the principal is the one in which the agent plays the action corresponding to the most likely state. If the signal is decision-relevant, there may be other equilibria besides the separating equilibrium just described: uninformative, perverse separating, semi-separating. But the separating equilibrium above is clearly the best from the viewpoint of the principal. We summarize the discussion of this section:

**Proposition 3** *In the concealed action case, if the agent signal $y$ is decision-relevant, the best equilibrium for the principal is a separating equilibrium in which $a = y$, while if the the signal is not decision-relevant, the best equilibrium is an uninformative equilibrium in which the agent plays the action corresponding to the more likely state.*

# 5 Revealed Action

We now consider the game in which the principal observes $a$ as well. A perfect Bayesian equilibrium of this game is a mixed strategy for the agent

$$\alpha_y = \Pr(a = 1|y)$$

and a belief for the principal

$$\pi(a,x) = \Pr(\theta = g|a,x) = \frac{\Pr(g,a,x)}{\Pr(g,a,x) + \Pr(b,a,x)}.$$

If the mixed strategy of the agent who observed $y$ puts positive probability on action $a^*$, then it must be that

$$a^* \in \arg\max_a \Pr(x=1|y)\pi(a, x=1) + \Pr(x=0|y)\pi(a, x=0).$$

If action $a$ is taken in equilibrium with positive probability, then the corresponding belief must be consistent with the agent's action. This implies:

$$\begin{aligned}
\pi(1,x) &= \frac{(\alpha_1 q_{xg} + \alpha_0(1-q_{xg}))\gamma}{(\alpha_1 q_{xg} + \alpha_0(1-q_{xg}))\gamma + (\alpha_1 q_{xb} + \alpha_0(1-q_{xb}))(1-\gamma)}; \\
\pi(0,x) &= \frac{((1-\alpha_1) q_{xg} + (1-\alpha_0)(1-q_{xg}))\gamma}{((1-\alpha_1) q_{xg} + (1-\alpha_0)(1-q_{xg}))\gamma + ((1-\alpha_1) q_{xb} + (1-\alpha_0)(1-q_{xb}))(1-\gamma)}.
\end{aligned}$$

Given $y$, the agent maximizes:

$$E(\pi(a,x)|a,y) = \pi(a,1)\Pr(x=1|y) + \pi(a,0)\Pr(x=0|y)$$

We use the following terminology. An *informative equilibrium* is one in which $\alpha_0 \neq \alpha_1$, while a *pooling equilibrium* is when $\alpha_0 = \alpha_1$. A *separating equilibrium* is a pure-strategy informative equilibrium (either $\alpha_0 = 1$ and $\alpha_1 = 0$, or $\alpha_0 = 0$ and $\alpha_1 = 1$). A *semi-separating equilibrium* is an informative equilibrium in which at least one agent plays a mixed strategy.

Pooling equilibria always exist for any value of $\alpha_0 = \alpha_1$. The agent disregards his own signal and the principal, knowing that the agent's action is uninformative, has invariant beliefs: $\pi(0,0) = \pi(0,1) = \pi(1,0) = \pi(1,1)$. The question is whether there are also informative equilibria.

### 5.1 Semi-separating equilibria

The first thing to show is that there cannot exist an informative equilibrium in which both the agent with $y=0$ and the one with $y=1$ use mixed strategies.

**Proposition 4** *There cannot exist a semi-separating equilibrium in which both agents play mixed strategies*

**Proof.** The conditions for an informative equilibrium in which both agents use mixed strategies are:

$$\alpha_0 \in (0,1), \alpha_1 \in (0,1), \alpha_0 \neq \alpha_1,$$

and

$$E(\pi(0,x)|y) = E(\pi(1,x)|y) \qquad \text{for } y = 0,1.$$

or

$$\Pr(x=0|y)\,\pi(1,0)+\Pr(x=1|y)\,\pi(1,1)=\Pr(x=0|y)\,\pi(0,0)+\Pr(x=1|y)\,\pi(0,1)$$

The last condition rewrites as

$$\Pr(x=0|y=1)\,(\pi(0,0)-\pi(1,0)) = \Pr(x=1|y=1)\,(\pi(1,1)-\pi(0,1)), \quad (6)$$

$$\Pr(x=0|y=0)\,(\pi(0,0)-\pi(1,0)) = \Pr(x=1|y=0)\,(\pi(1,1)-\pi(0,1)). \quad (7)$$

We are looking for a contradiction. In an informative equilibrium it cannot be that both $\pi(0,0)=\pi(1,0)$ and $\pi(1,1)=\pi(0,1)$. If $(\pi(0,0)-\pi(1,0))(\pi(1,1)-\pi(0,1))\leq 0$, then we have a contradiction. Thus, we assume that $(\pi(0,0)-\pi(1,0))(\pi(1,1)-\pi(0,1))>0$.

Subtracting (7) from (6),

$$(\Pr(x=0|y=1)-\Pr(x=0|y=0))\,(\pi(0,0)-\pi(1,0))=(\Pr(x=1|y=1)-\Pr(x=1|y=0))\,(\pi(1,1)-\pi$$

But by assumption signals are informative on $x$:

$$\Pr(x=0|y=1)-\Pr(x=0|y=0) < 0;$$

$$\Pr(x=1|y=1)-\Pr(x=1|y=0) > 0.$$

We then have a contradiction. ■

If there existed an informative equilibrium in which both $\alpha_0$ and $\alpha_1$ are interior, the agent would always be indifferent between playing 0 or 1. But this can be true only if signals are uninformative or posteriors are flat – a contradiction. This kind of result is common to many signalling games.

Furthermore, there cannot exist a semi-separating equilibrium unless there exists a separating equilibrium:[5]

**Proposition 5** *There exists a semi-separating equilibrium only if there exists a separating equilibrium.*

**Proof.** We have seen above that there cannot exist an informative equilibrium in which both agents use mixed strategies. Consider a semiseparating equilibrium and assume without loss of generality that $\alpha_1=1$. For this type of equilibrium, let $\Pi(a,x,\alpha_0)=\Pr(g|a,x,\alpha_0)$. That is, $\Pi$ is the posterior probability $\pi(a,x)$ that the type is $g$ given $a$ and $x$ in an semi-separating equilibrium in which an agent with $y=0$ plays according to $\alpha_0$. For simplicity, let $\Pi(a,x)=\Pi(a,x,0)$. It is easy to check that $\Pi(a,x,\alpha_0)$ has a unique value.

[5] A result in a similar vein is found in Ottaviani and Sorensen [17, Lemma 1].

Then, a necessary condition the existence of a semiseparating equilibrium is that, for some $\alpha_0 < 1$, $E(\pi(1,x,\alpha_0)|0) - E(\pi(0,x,\alpha_0)|0) \leq 0$. That is,

$$\Pr(x=1|y=0)(\pi(1,1,\alpha_0) - \pi(0,1,\alpha_0)) + \Pr(x=0|y=0)(\pi(1,0,\alpha_0) - \pi(0,0,\alpha_0)) \leq 0$$

We have

$$\begin{aligned}
& \pi(1,x,\alpha_0) \\
= & \frac{\Pr(a=1|g,x)\Pr(g)}{\Pr(a|x)} \\
= & \frac{(\Pr(y=1|g,x) + \alpha\Pr(y=0|g,x))\Pr(g)}{(\Pr(y=1|x) + \alpha\Pr(y=0|x))} \\
= & \frac{\frac{\Pr(y=1|g,x)\Pr(g)}{\Pr(y=1|x)}\Pr(y=1|x) + \alpha\frac{\Pr(y=0|g,x)\Pr(g)}{\Pr(y=0|x)}\Pr(y=0|x)}{(\Pr(y=1|x) + \alpha\Pr(y=0|x))} \\
= & \frac{\pi(1,x)\Pr(y=1|x) + \alpha\pi(0,x)\Pr(y=0|x)}{(\Pr(y=1|x) + \alpha\Pr(y=0|x))} \\
= & \frac{\Pr(y=1|x)}{(\Pr(y=1|x) + \alpha\Pr(y=0|x))}\pi(1,x) + \frac{\alpha\Pr(y=0|x)}{(\Pr(y=1|x) + \alpha\Pr(y=0|x))}\pi(0,x)
\end{aligned}$$

When $\alpha_0 = 0$, $\pi(1,x,\alpha_0) = \pi(1,x)$. When $\alpha_0 = 1$, $\pi(1,x,\alpha_0) = \gamma$. Also,

$$\pi(0,x,\alpha_0) = \frac{\Pr(a=1|g,x)\Pr(g)}{\Pr(a|x)} = \frac{\Pr(y=0|g,x)\Pr(g)}{\Pr(y=0|x)} = \pi(0,x).$$

Then,

$$\begin{aligned}
& \pi(1,x,\alpha_0) - \pi(0,x,\alpha_0) \\
= & \frac{\Pr(y=1|x)}{(\Pr(y=1|x) + \alpha\Pr(y=0|x))}\pi(1,x) + \frac{\alpha\Pr(y=0|x)}{(\Pr(y=1|x) + \alpha\Pr(y=0|x))}\pi(0,x) - \pi(0,x) \\
= & \frac{\Pr(y=1|x)}{(\Pr(y=1|x) + \alpha\Pr(y=0|x))}\pi(1,x) - \left(1 - \frac{\alpha\Pr(y=0|x)}{(\Pr(y=1|x) + \alpha\Pr(y=0|x))}\right)\pi(0,x) \\
= & \frac{\Pr(y=1|x)}{(\Pr(y=1|x) + \alpha\Pr(y=0|x))}(\pi(1,x) - \pi(0,x))
\end{aligned}$$

implying,

$$\begin{aligned}
& \Pr(1|0)(\pi(1,1,\alpha_0) - \pi(0,1,\alpha_0)) + \Pr(0|0)(\pi(1,0,\alpha_0) - \pi(0,0,\alpha_0)) \\
= & \Pr(1|y=0)\frac{\Pr(y=1|1)}{\Pr(y=1|1) + \alpha\Pr(y=0|1)}(\pi(1,1) - \pi(0,1)) \\
& + \Pr(0|y=0)\frac{\Pr(y=1|0)}{\Pr(y=1|0) + \alpha\Pr(y=0|)}(\pi(1,0) - \pi(0,0))
\end{aligned}$$

Note that

$$
\begin{aligned}
\pi(1,0) &\leq \pi(0,0) \\
\frac{\Pr(y=1|g,0)\Pr(g)}{\Pr(y=1|0)} &\leq \frac{\Pr(y=0|g,0)\Pr(g)}{\Pr(y=0|0)} \\
\frac{\Pr(y=1|g,0)}{\Pr(y=1|0)} &\leq \frac{1-\Pr(y=1|g,0)}{1-\Pr(y=1|0)} \\
\Pr(y &= 1|g,0)(1-\Pr(y=1|0)) \leq (1-\Pr(y=1|g,0))\Pr(y=1|0) \\
\Pr(y &= 1|g,0) - \Pr(y=1|g,0)\Pr(y=1|0) \leq \Pr(y=1|0) - \Pr(y=1|g,0)\Pr(y=1|0) \\
\Pr(y &= 1|g,0) \leq \Pr(y=1|0) \\
& \text{always true}
\end{aligned}
$$

A semi(separating) equilibrium exists only if $E(\pi(1,x,\alpha)|0) \leq E(\pi(0,x,\alpha)|0)$ for some $\alpha$. That is:

$$
\begin{aligned}
& \Pr(1|y=0)\frac{\Pr(y=1|1)}{\Pr(y=1|1)+\alpha\Pr(y=0|1)}(\pi(1,1)-\pi(0,1)) \\
\leq\; & \Pr(0|y=0)\frac{\Pr(y=1|0)}{\Pr(y=1|0)+\alpha\Pr(y=0|)}(\pi(0,0)-\pi(1,0))
\end{aligned}
$$

for some $\alpha$, which rewrites as

$$
\begin{aligned}
\frac{\frac{\Pr(y=1|1)}{\Pr(y=1|1)+\alpha\Pr(y=0|1)}}{\frac{\Pr(y=1|0)}{\Pr(y=1|0)+\alpha\Pr(y=0|0)}} &\leq \frac{\Pr(0|y=0)(\pi(0,0)-\pi(1,0))}{\Pr(1|y=0)(\pi(1,1)-\pi(0,1))} \qquad \text{for some } \alpha \\
\min_\alpha \frac{\frac{\Pr(y=1|1)}{\Pr(y=1|1)+\alpha\Pr(y=0|1)}}{\frac{\Pr(y=1|0)}{\Pr(y=1|0)+\alpha\Pr(y=0|0)}} &\leq \frac{\Pr(0|y=0)(\pi(0,0)-\pi(1,0))}{\Pr(1|y=0)(\pi(1,1)-\pi(0,1))}
\end{aligned}
$$

Note that

$$
\begin{aligned}
\arg\min_\alpha \frac{\frac{\Pr(y=1|1)}{\Pr(y=1|1)+\alpha\Pr(y=0|1)}}{\frac{\Pr(y=1|0)}{\Pr(y=1|0)+\alpha\Pr(y=0|0)}} &= \arg\min_\alpha \frac{\Pr(y=1|0)+\alpha\Pr(y=0|0)}{\Pr(y=1|1)+\alpha\Pr(y=0|1)} \\
&= \arg\min_\alpha \frac{1-\Pr(y=0|0)+\alpha\Pr(y=0|0)}{1-\Pr(y=0|1)+\alpha\Pr(y=0|1)} \\
&= \arg\min_\alpha \frac{1-(1-\alpha)\Pr(y=0|0)}{1-(1-\alpha)\Pr(y=0|1)} \\
&= \begin{cases} 0 & \text{if} \quad \Pr(y=0|0) > \Pr(y=0|1) \\ 1 & \text{if} \quad \Pr(y=0|0) < \Pr(y=0|1) \end{cases} \\
&= 0
\end{aligned}
$$

The condition for the nonexistence of a semiseparating equilibrium is the same as the condition for the nonexistence of a separating equilibrium, namely,

$$
\Pr(1|y=0)(\pi(1,1)-\pi(0,1)) \leq \Pr(0|y=0)(\pi(0,0)-\pi(1,0))
$$

■

Proposition 5 is proven by showing that the condition for a separating equilibrium implies that for a semi-separating equilibrium. To see this, suppose without loss of generality that the agent who observes $y = 1$ plays $a = 1$. The other agent plays $a = 1$ with probability $\alpha_0$. A change in $\alpha_0$ does not affect the posterior given that $a = 0$, $\pi(0, x)$, but it affects the posterior given that $a = 1$, $\pi(1, x)$. Now when $a = 1$ the fact that also $x = 1$ is less of a good signal. Hence, $\pi(1, 1)$ goes down and $\pi(1, 0)$ goes up. In the limit, as $\alpha_0 \to 1$, the two coincides. Thus, so far the effect of an increase in $\alpha_0$ on the incentive for the agent with $y = 0$ to separate is ambiguous. However, in the proof we show that $E(\pi(1, x, \alpha_0) | y = 0)$ is lowest when $\alpha_0 = 0$. This means that the incentive for the agent with $y = 0$ to separate is highest in a separating equilibrium. If the agent does not want to separate when $\alpha_0 = 0$, then he does not want to separate in any semi-separating equilibrium.

The proposition does not imply that there do not exist semi-separating equilibria. It is possible to find games in which there exist both a separating equilibrium and a semi-separating equilibrium.

## 5.2 Separating equilibria

Now that we know that the condition for the existence of an informative equilibrium is the same as the condition for the existence of a separating equilibrium, it is feasible to find it.

**Proposition 6** *There exists an informative equilibrium if and only if*

$$\frac{p}{1-p}\frac{\Pr(b)q_{0b} + \Pr(g)q_{0g}}{\Pr(g)q_{1g} + \Pr(b)q_{1b}} \leq \frac{q_{0b} - q_{0g}}{q_{1g} - q_{1b}} \leq \frac{p}{1-p}\frac{1 - \Pr(b)q_{0b} - \Pr(g)q_{0g}}{1 - \Pr(g)q_{1g} - \Pr(b)q_{1b}}. \tag{8}$$

**Proof.** Necessary and sufficient conditions for the existence of a separating equilibrium are:

$$\begin{aligned}
\Pr(1|y = 0)(\pi(1, 1) - \pi(0, 1)) &\geq \Pr(0|y = 0)(\pi(0, 0) - \pi(1, 0));\\
\Pr(1|y = 1)(\pi(1, 1) - \pi(0, 1)) &\leq \Pr(0|y = 1)(\pi(0, 0) - \pi(1, 0)).
\end{aligned}$$

$$\begin{aligned}
\pi(1, x) - \pi(0, x) &= \left(\frac{\Pr(y = 1|g, x)}{\Pr(y = 1|x)} - \frac{\Pr(y = 0|g, x)}{\Pr(y = 0|x)}\right)\Pr(g)\\
&= \left(\frac{\Pr(y = 1|g, x)}{\Pr(y = 1|x)} - \frac{1 - \Pr(y = 1|g, x)}{1 - \Pr(y = 1|x)}\right)\Pr(g)\\
&= \frac{\Pr(y = 1|g, x)(1 - \Pr(y = 1|x)) - (1 - \Pr(y = 1|g, x))\Pr(y = 1|x)}{\Pr(y = 1|x)\Pr(y = 0|x)}\Pr(g)\\
&= \frac{\Pr(y = 1|g, x) - \Pr(y = 1|x)}{\Pr(y = 1|x)\Pr(y = 0|x)}\Pr(g)
\end{aligned}$$

First,

$$\begin{aligned}
\Pr\left(x|y=0\right)\left(\pi\left(1,x\right)-\pi\left(0,x\right)\right) &= \frac{\Pr(y=0|x)\Pr(x)}{\Pr(y=0)}\frac{\Pr(y=1|g,x)-\Pr(y=1|x)}{\Pr(y=1|x)\Pr(y=0|x)}\Pr(g) \\
&= \frac{\Pr(g)}{\Pr(y=0)}\Pr(x)\frac{\Pr(y=1|g,x)-\Pr(y=1|x)}{\Pr(y=1|x)}
\end{aligned}$$

$$\begin{aligned}
&\Leftrightarrow\ \Pr\left(1|y=0\right)\left(\pi\left(1,1\right)-\pi\left(0,1\right)\right)+\Pr\left(0|y=0\right)\left(\pi\left(1,0\right)-\pi\left(0,0\right)\right)\leq 0 \\
&\Leftrightarrow\ \frac{\Pr(g)}{\Pr(y=0)}\left(\Pr(1)\frac{\Pr(y=1|g,1)-\Pr(y=1|1)}{\Pr(y=1|1)}+\Pr(0)\frac{\Pr(y=1|g,0)-\Pr(y=1|0)}{\Pr(y=1|0)}\right)\leq 0 \\
&\Leftrightarrow\ \frac{\Pr(g)}{\Pr(y=0)}\left(\Pr(1)\frac{\Pr(y=1|g,1)}{\Pr(y=1|1)}+\Pr(0)\frac{\Pr(y=1|g,0)}{\Pr(y=1|0)}-1\right)\leq 0 \\
&\Leftrightarrow\ \begin{array}{l}\Pr(1)\Pr(y=1|g,1)\left(\Pr(y=1|g,0)\Pr(g)+\Pr(y=1|b,0)\Pr(b)\right)+\Pr(0)\left(\Pr(y=1|g,0)\Pr(y=1|g,1)\right. \\ -\Pr(y=1|0)\Pr(y=1|1)\leq 0\end{array} \\
&\Leftrightarrow\ \Pr(1)q_{1g}\left(q_{0g}\Pr(g)+q_{0b}\Pr(b)\right)+\Pr(0)q_{0g}\left(q_{1g}\Pr(g)+q_{1b}\Pr(b)\right)-\left(q_{0g}\Pr(g)+q_{0b}\Pr(b)\right)\left(q_{1g}\Pr(g)+\right. \\
&\Leftrightarrow\ \Pr(1)q_{1g}\left(q_{0g}\Pr(g)+q_{0b}\Pr(b)\right)+\Pr(0)q_{0g}\left(q_{1g}\Pr(g)+q_{1b}\Pr(b)\right)-\left(\Pr(0)+\Pr(1)\right)\left(q_{0g}\Pr(g)+q_{0b}\Pr(\right. \\
&\Leftrightarrow\ \Pr(1)\left(q_{1g}-\left(q_{1g}\Pr(g)+q_{1b}\Pr(b)\right)\right)\left(q_{0g}\Pr(g)+q_{0b}\Pr(b)\right)+\Pr(0)\left(q_{0g}-\left(q_{0g}\Pr(g)+q_{0b}\Pr(b)\right)\right)\left(q_{1g}\Pr\right. \\
&\Leftrightarrow\ \Pr(1)\left(q_{1g}-q_{1b}\right)\Pr(b)\left(q_{0g}\Pr(g)+q_{0b}\Pr(b)\right)+\Pr(0)\left(q_{0g}-q_{0b}\right)\Pr(b)\left(q_{1g}\Pr(g)+q_{1b}\Pr(b)\right)\leq 0 \\
&\Leftrightarrow\ \Pr(1)\left(q_{1g}-q_{1b}\right)\left(q_{0g}\Pr(g)+q_{0b}\Pr(b)\right)\leq\Pr(0)\left(q_{0b}-q_{0g}\right)\left(q_{1g}\Pr(g)+q_{1b}\Pr(b)\right) \\
&\Leftrightarrow\ \frac{q_{0b}-q_{0g}}{q_{1g}-q_{1b}}\geq\frac{p}{1-p}\frac{\Pr(b)q_{0b}+\Pr(g)q_{0g}}{\Pr(g)q_{1g}+\Pr(b)q_{1b}}
\end{aligned}$$

Second,

$$\begin{aligned}
\Pr\left(x|y=1\right)\left(\pi\left(1,x\right)-\pi\left(0,x\right)\right) &= \frac{\Pr(y=1|x)\Pr(x)}{\Pr(y=1)}\frac{\Pr(y=1|g,x)-\Pr(y=1|x)}{\Pr(y=1|x)\Pr(y=0|x)}\Pr(g) \\
&= \frac{\Pr(g)}{\Pr(y=1)}\Pr(x)\frac{\Pr(y=1|g,x)-\Pr(y=1|x)}{\Pr(y=0|x)}
\end{aligned}$$

The condition holds if and only if

$$\begin{aligned}
&\Leftrightarrow\ \Pr\left(1|y=1\right)\left(\pi\left(1,1\right)-\pi\left(0,1\right)\right)+\Pr\left(0|y=1\right)\left(\pi\left(1,0\right)-\pi\left(0,0\right)\right)\geq 0 \\
&\Leftrightarrow\ \frac{\Pr(g)}{\Pr(y=1)}\left(\Pr(1)\frac{\Pr(y=1|g,1)-\Pr(y=1|1)}{\Pr(y=0|1)}+\Pr(0)\frac{\Pr(y=1|g,0)-\Pr(y=1|0)}{\Pr(y=0|0)}\right)\geq 0 \\
&\Leftrightarrow\ \Pr(1)\left(\Pr(y=1|g,1)-\Pr(y=1|1)\right)\Pr(y=0|0)+\Pr(0)\left(\Pr(y=1|g,0)-\Pr(y=1|0)\right)\Pr(y=0|1)\geq 0 \\
&\Leftrightarrow\ \Pr(1)\left(\Pr(y=1|g,1)-\Pr(y=1|1)\right)\left(1-\Pr(y=1|0)\right)+\Pr(0)\left(\Pr(y=1|g,0)-\Pr(y=1|0)\right)\left(1-\Pr(y\right. \\
&\Leftrightarrow\ \begin{array}{l}\Pr(1)\left(\Pr(b)\Pr(y=1|g,1)-\Pr(b)\Pr(y=1|b,1)\right)\left(1-\Pr(g)\Pr(y=1|g,0)-\Pr(b)\Pr(y=1|b,0)\right) \\ +\Pr(0)\left(\Pr(b)\Pr(y=1|g,0)-\Pr(b)\Pr(y=1|b,0)\right)\left(1-\Pr(g)\Pr(y=1|g,1)-\Pr(b)\Pr(y=1|b,1)\right)\geq\end{array} \\
&\Leftrightarrow\ \Pr(1)\left(q_{1g}-q_{1b}\right)\left(1-\Pr(g)q_{0g}-\Pr(b)q_{0b}\right)+\Pr(0)\left(q_{0g}-q_{0b}\right)\left(1-\Pr(g)q_{1g}-\Pr(b)q_{1b}\right)\geq 0 \\
&\Leftrightarrow\ \Pr(1)\left(q_{1g}-q_{1b}\right)\left(1-\Pr(g)q_{0g}-\Pr(b)q_{0b}\right)\geq\Pr(0)\left(q_{0b}-q_{0g}\right)\left(1-\Pr(g)q_{1g}-\Pr(b)q_{1b}\right) \\
&\Leftrightarrow\ \frac{q_{0b}-q_{0g}}{q_{1g}-q_{1b}}\leq\frac{p}{1-p}\frac{1-\Pr(b)q_{0b}-\Pr(g)q_{0g}}{1-\Pr(g)q_{1g}-\Pr(b)q_{1b}}.
\end{aligned}$$

■

Proposition 6 is best understood in connection with the conditions for the virtuous signal found in Proposition 1. Both impose bounds on the term

$$\frac{q_{0b} - q_{0g}}{q_{1g} - q_{1b}},$$

which is the relative informativeness of the two $y$'s. The virtuosity condition establishes which signal is more informative. The condition in Proposition 6 says whether one signal is *much* more informative than the other. This is because

$$\frac{1 - \Pr(b)q_{0b} - \Pr(g)q_{0g}}{1 - \Pr(g)q_{1g} - \Pr(b)q_{1b}} > 1 \qquad \text{and} \qquad \frac{\Pr(b)q_{0b} + \Pr(g)q_{0g}}{\Pr(g)q_{1g} + \Pr(b)q_{1b}} < 1.$$

If, for instance, $y = 1$ is the virtuous signal, then

$$\frac{q_{0b} - q_{0g}}{q_{1g} - q_{1b}} < \frac{p}{1-p} \tag{9}$$

We can disregard the second inequality in Proposition 6 because it is implied by (9). Instead, the inequality

$$\frac{q_{0b} - q_{0g}}{q_{1g} - q_{1b}} \geq \frac{p}{1-p} \frac{\Pr(b)q_{0b} + \Pr(g)q_{0g}}{\Pr(g)q_{1g} + \Pr(b)q_{1b}} \tag{10}$$

can hold or not. If it holds, there is no informative equilibrium because $y = 1$ is "too virtuous" to allow for separation. If the equilibrium were informative, the agent who observes $y = 0$ would always want to pretend he observed $y = 1$. If instead the inequality (10) does not hold, separation is possible because the agent who observes $y = 0$ prefers to increase his likelihood to get $u = 1$ rather than pretend he has $y = 1$.

If instead $y = 0$ is the virtuous signal, we can disregard the first inequality in Proposition 6. The second inequality tells us if $y = 0$ is too virtuous to allow for separation.

If we revisit the example presented earlier, we can now formally verify the result that there is no informative equilibrium. Recall that in that example $\gamma = \frac{1}{2}$, $p = \frac{1}{2}$, $q_{0b} = q_{1b} = \frac{1}{2}$, $q_{0g} = \frac{1}{2}$, and $q_{1g} = 1$. The virtuosity condition (9) is

$$\frac{0}{\frac{1}{2}} < 1.$$

The virtuous signal is $y = 1$. There exists an informative equilibrium if and only if (10) is satified. That is,

$$0 \geq 1 \frac{\frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2}}{\frac{1}{2}1 + \frac{1}{2}\frac{1}{2}} = \frac{2}{3},$$

which shows that informative equilibria are impossible.

If instead the virtuous signal had been less virtuous, an informative equilibrium would have been possible. For instance, modify the example by assuming that if $x = 0$, the good type receives an informative signal: $q_{0g} = \frac{1}{6}$. The existence condition (10) becomes

$$\frac{\frac{1}{2} - \frac{1}{6}}{1 - \frac{1}{2}} \geq 1 \frac{\frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{6}}{\frac{1}{2}1 + \frac{1}{2}\frac{1}{2}},$$

that is, $\frac{2}{3} \geq \frac{4}{9}$. Indeed, one can show that, holding the other parameters constant, there exists an informative equilibrium if and only if $q_{0g} \leq \frac{1}{4}$.

To complete the analysis, we examine the welfare effects of revealing the action. As usual, for each information scenario we focus on th best equilibrium from the viewpoint of the principal's expected payoff.

First suppose that the agent signal is decision-relevant. From Proposition 3, the best equilibrium with concealed action is a separating equilibrium with $a = y$. What happens with revealed action depends on condition (8). If the condition holds, there exists a separating equilibrium with $a = y$. The agent behavior is thus the same as with concealed action but the principal gets more information. The variance of the agent posterior increases and the principal's payoff, which is convex in the posterior, goes up.[6] Compared to concealed action, the discipline effect is the same but the screening effect improves. Thus, the principal is better off. If instead condition (8) fails, there is no informative equilibrium and the best equilibrium is one where the agent chooses the action that corresponds to the most likely state. The discipline effect worsens because the agent disregards decision-relevant information. Screening too is affected negatively because in an informative equilibrium the posterior is equal to the prior. Thus, the principal is worse off.

Second, if the agent signal is not decision-relevant, the best equilibrium with concealed action is the uninformative equilibrium where the agent always chooses the action that is associated to the state that is ex ante more likely. This is still an equilibrium when the action is revealed. Thus, revelation cannot make the principal worse off in the sense that we are using here.

[6] Proof that screening is better when $a$ is observed:

$$\begin{aligned} \tilde{\pi}(u) &= E(\pi(a,x)|u) \\ &\rightarrow v(\tilde{\pi}(u)) \leq E(v(\pi(a,x))|u) \end{aligned}$$

$$\begin{aligned} v &= \sum_{x,y} \Pr(x,y) v(\pi(y,x)) \\ &= \Pr(y=x) \sum_x \Pr(x) v(\pi(x,x)) + \Pr(y \neq x) \sum_x \Pr(x) v(\pi(1-x,x)) \\ &\geq \Pr(y=x) \sum_x \Pr(x) v(\tilde{\pi}(1)) + \Pr(y \neq x) \sum_x \Pr(x) v(\tilde{\pi}(0)) \\ &= \sum_{x,y} \Pr(x,y) v(\pi(u(x,y))) \end{aligned}$$

**Proposition 7** *Suppose the agent signal is decision-relevant. If (8) holds and/or the agent signal is not decision-relevant, revealing the agent's action does not affect discipline and improves screening. If (8) fails and the signal is decision-relevant, revealing the agent's action worsens both discipline and screening. Hence, the principal prefers to reveal the action if and only if (8) holds.*

# 6 The Difficulty of Committing

So far, we have assumed that the principal can commit to keeping the agent action concealed. However, in reality commitment could be imperfect and the principal could, perhaps at a cost, break her commitment and observe the agent's action once the agent has taken it. For instance, the agent could be a civil servant who operates in an opaque environment but can ex-post be forced to make his actions public by setting up a parliamentary commission inquiry. Or is could be the manager of a firm who can be compelled by shareholders to make the details of his decisions public by legal action.

It is immediate to see that once the decision is made the principal always prefers to renege on her promise and observe the agent's action. The problem of commitment has two sides: How *useful* is it for the principal to commit to keep the agent's action concealed? How *difficult* is it to keep the commitment ex post? In this section we shall see that the two sides are tightly connected: the more useful commitment is, the harder it is not too break it ex post.

To explore this tension, we make the commitment technology explicit. Committing to keep the action concealed means that, if the principal wants to observe the action ex post, she has to pay a "break-the-piggybank" cost $r$. If $r = 0$ we have no commitment; if $r \to \infty$ we have perfect commitment. Let $r^*$ be the minimal cost at which the principal does not want to observe the agent's action ex post assuming that the agent plays $a = y$. The threshold $r^*$ can be viewed as the difficulty of committing. To simplify the analysis, we assume that the principal chooses whether to break her commitment after the agent takes his action but *before* the consequence is known.

Suppose that the virtuous signal is $y = 1$. As we saw above, a policy of concealment is optimal if and only if (10) fails. The incentive to commit is simply represented by the variable $c$ which takes value 0 if (10) holds and 1 if it fails.

To make a simple comparative analysis exercise, we fix $p$ and $\gamma$. We also hold constant the following expressions:

$$\begin{aligned}\Pr(y &= x|g) = pq_{1g} + (1-p)(1-q_{0g}) \equiv s; \\ \Pr(y &= 1|x) = q_{xg}\gamma + q_{xb}(1-\gamma) \equiv q_x \qquad \text{for } x = 1, 2.\end{aligned}$$

Note that this also implies that $\Pr(y=x)$ and $\Pr(y=x|b)$ are constant. This leaves one degree of freedom on $q$, which can be represented without loss of generality with movements of the ratio $\frac{q_{0b}-q_{0g}}{q_{1g}-q_{1b}}$.

**Proposition 8** *A decrease in* $\frac{q_{0b}-q_{0g}}{q_{1g}-q_{1b}}$ *causes an increase in the incentive to commit c and in the difficulty of committing* $r^*$*.*

**Proof.** The incentive to commit $c$ depends on whether (10) holds The right-hand side of the inequality is constant. Hence, $c$ is nondecreasing in $\frac{q_{0b}-q_{0g}}{q_{1g}-q_{1b}}$.

For the difficulty of committing $r^*$, not that, as $s$ is constant, an increase in $q_{1g}$ must be accompanied by an increase in $q_{0g}$. As $q_x$ is constant, an increase in $q_{xg}$ is associated to a decrease in $q_{xb}$. Thus, a decrease in $\frac{q_{0b}-q_{0g}}{q_{1g}-q_{1b}}$ corresponds to an increase in $q_{1g}$ and $q_{0g}$ and a decrease in $q_{1b}$ and $q_{0b}$.

Given that $a=y$, the posteriors are

$$\begin{aligned}
\pi(1,0) &= \frac{q_{0g}\gamma}{q_{0g}\gamma+q_{0b}(1-\gamma)};\\
\pi(0,0) &= \frac{(1-q_{0g})\gamma}{(1-q_{0g})\gamma+(1-q0)(1-\gamma)};\\
\pi(1,1) &= \frac{q_{1g}\gamma}{q_{1g}\gamma+q_{1b}(1-\gamma)};\\
\pi(0,1) &= \frac{(1-q_{1g})\gamma}{(1-q_{1g})\gamma+(1-q_{1b})(1-\gamma)}.
\end{aligned}$$

A decrease in $\frac{q_{0b}-q_{0g}}{q_{1g}-q_{1b}}$ generates an increase in $\pi(1,1)$ and $\pi(1,0)$ and a decrease in $\pi(0,1)$ and $\pi(0,0)$. The assumption that $\Pr(y=x|g)$ is constant means that $\tilde{\pi}(u)$ is constant as well.

The benefit of observing $a$ when the agent plays $a=y$ is

$$\sum_{x,y}\Pr(x,y)v(\pi(y,x))-\sum_{x,y}\Pr(x,y)v(\tilde{\pi}(u(y,x))).$$

Given that $\tilde{\pi}(u)$ is constant, we only consider the first part, which we rewrite as

$$V=\Pr(y=x)\sum_{\tilde{x}}\Pr(\tilde{x})v(\pi(\tilde{x},\tilde{x}))+\Pr(y\neq x)\sum_{\tilde{x}}\Pr(\tilde{x})v(\pi(1-\tilde{x},\tilde{x})).$$

Note that

$$\sum_{\tilde{x}}\Pr(\tilde{x})\pi(\tilde{x},\tilde{x})=\tilde{\pi}(1) \qquad \sum_{\tilde{x}}\Pr(\tilde{x})\pi(1-\tilde{x},\tilde{x})=\tilde{\pi}(0).$$

It is easy to check that $\pi(1,1)>\pi(0,0)$. Hence a decrease in $\frac{q_{0b}-q_{0g}}{q_{1g}-q_{1b}}$ increases $\sum_x\Pr(x)v(\pi(x,x))$. Similarly $\pi(1,0)>\pi(0,1)$. A decrease in $\frac{q_{0b}-q_{0g}}{q_{1g}-q_{1b}}$ increases $\sum_x\Pr(x)v(\pi(1-x,x))$. ■

There is a tension between the ex ante benefit of concealed action and the ex post of revealed action. If the agent plays according to their signal, the principal can infer whether the agent received the virtuous signal or not. The more "virtuous" the virtuous signal is, the more useful the information. In a highly asymmetric situation, ex post the principal stands to gain a lot from knowing the agent's action. But the agent realizes this and wants to hides the fact that he receives the non-virtuous signal. This kills the separating equilibrium and damages the principal in terms both of discipline and sorting. Thus, if the virtuous signal is very virtuous, the principal is better off if she can commit to concealed action. In this case, we should expect the principal to try to develop powerful commitment technology, which makes it very costly to observe the agent's action ex post.

# 7 Complementarity between Observing Action and Consequence

We have so far asked whether revealing the agent's action is a good idea, but we have maintained the assumption that consequences are always observed. In some cases, especially in the political arena, the principal may not be able to fully evaluate the consequences of the agent's behavior or may be able to do it with such a time lag that the information is of limited use for screening purposes. This is the case for large-scale public projects, such a reform of the health system. Its main provisions are observable right away, but it takes years for its effects to develop.

This section looks at what happens when consequences are not necessarily observed. First, we examine the simple case in which $u$ is not observed. The game is as in the reduced form except that at stage 2 the principal observes either only $a$ or nothing at all.

If $a$ is observed, in a perfect Bayesian equilibrium the choice of action must be uncorrelated with the agent's signal.If $\pi(a=1) \neq \pi(a=0)$, the agent plays only the action associated with the higher posterior, independently of $y$. Thus, in equilibrium one of the following must be true: only one action is played or the posteriors are the same. In both cases, the action choice provides the principal with no information. The best equilibrium for the principal is an uninformative equilibrium in which the incumbent chooses always chooses the most likely action. No screening occurs. If the action is not observed, the best equilibrium is one in which the agent chooses $a = y$. No screening occurs but the first-period decision is better. The principal's expected payoff is higher when $a$ is concealed.

This observation contrasts with the result obtained in the previous section that revealing the action may a good idea when the consequence is observed and it seems to point to a complementarity between observing consequences and revealing actions. We shall now show that this complementarity is indeed present in a general way. Let $\rho_u \in [0,1]$ be the probability that $u$ is observed and $\rho_a \in [0,1]$ be the probability that $a$ is observed. At stage 2 there are thus four possible information scenarios according to whether the consequence and/or the action is observed. The previous section considered the cases ($\rho_u = 1, \rho_a = 1$) and ($\rho_u = 1, \rho_a = 0$).

To simplify matters, we disregard the possibility of semiseparating equilibria or perverse equilibria. We assume that $y = 1$ is the virtuous signal and we restrict attention to the separating equilibrium in which $a = y$ and the pooling equilibrium in which the agent plays the most likely action. The pooling equilibrium always exists. For every pair $(\rho_u, \rho_a)$, we ask whether the separating equilibrium exists.

**Proposition 9** *For every $\rho_u$ there exists $\rho_a^*(\rho_u) \in (0, 1]$ such that the game has a separating equilibrium if and only if $\rho_a \leq \rho_a^*$. The threshold $\rho_a^*$ is nondecreasing in $\rho_u$.*

**Proof.** Suppose that the agent chooses $a = y$. Let $\pi(a, x)$, $\pi(u(a, x)$, $\pi(a)$, and $\gamma$ be the posterior evaluated by the principal in the four possible information scenarios. Note that because we hold fixed the agent equilibrium strategy ($a = y$), these posteriors do not depend on $(\rho_u, \rho_a)$ but only on the information scenario that is realized. Given $a$ and $y$, the expected posterior for the agent is

$$E(\pi|a, y) = \rho_u\rho_a E_x(\pi(a, x)|y) + \rho_u(1 - \rho_a) E_x(\pi(u(a, x)|y) + (1 - \rho_u)\rho_a\pi(a) + (1 - \rho_u)(1 - \rho_a)\gamma.$$

Note that the last two addends do not depend on $x$, and therefore on $y$. A necessary and sufficient condition for the existence of a separating equilibrium is $E(\pi|0, 0) \geq E(\pi|1, 0)$, which rewrites as:

$$\begin{aligned} \rho_u\rho_a E_x(\pi(0, x)|y &= 0) + \rho_u(1 - \rho_a) E_x(\pi(u(0, x)|y = 0) + (1 - \rho_u)\rho_a\pi(a = 0) + (1 - \rho_u)(1 - \rho_a)\gamma \\ &\geq \rho_u\rho_a E_x(\pi(1, x)|y = 0) + \rho_u(1 - \rho_a) E_x(\pi(u(1, x)|y = 0) + (1 - \rho_u)\rho_a\pi(a = 1) + (1 - \end{aligned}$$

or

$$\begin{aligned} & (1 - \rho_a)\rho_u(E_x(\pi(u(0, x)|y = 0) - E_x(\pi(u(1, x)|y = 0)) \\ \geq\ & \rho_a(\rho_u(E_x(\pi(1, x)|y = 0) - E_x(\pi(0, x)|y = 0)) + (1 - \rho_u)(\pi(a = 1) - \pi(a = 0))) \end{aligned}$$

or

$$(1 - \rho_a)\rho_u\Delta_1 \geq \rho_a(\rho_u\Delta_2 + (1 - \rho_u)\Delta_3).$$

Note that $\Delta_1$, $\Delta_2$, and $\Delta_3$ do not depend on $(\rho_u, \rho_a)$. It is easy to see that $\Delta_1 > 0$ and that $\Delta_3 > 0$. Because

$$E_x(\pi(1, x)|y = 0) - E_x(\pi(0, x)|y = 0) < E_y E_x(\pi(1, x)|y) - E_y E_x(\pi(0, x)|y) < E_x(\pi(1, x)|y = 1) - E_x(\pi(0, x)$$

and $E_y E_x(\pi(a,x)|y) = \pi(a)$, we have that $\Delta_3 > \Delta_2$. We rewrite () as

$$\frac{1-\rho_a}{\rho_a} \geq \frac{\rho_u \Delta_2 + (1-\rho_u)\,\Delta_3}{\rho_u \Delta_1}.$$

On the right-hand side, the numerator is decreasing in $\rho_u$ and the denominator is increasing. The left-hand side is decreasing in $\rho_a$. The proposition is proven. ■

Observing the action is more likely to be useful when the precision of the signal on the consequence increases. If $\rho$ is very low, revealing $a$ is always a bad idea because it is certain to induce conformism. As $\rho$ increases, the principal is more able to confront the action with the consequence and to "punish" an agent who acts against his own signal. It may or may not be the case that for $\rho = 1$ revelation is the optimal policy (see Proposition 7). If it is the case, then there exists a threshold $\bar{\rho}$ below which concealing the action is optimal and above which revealing it is optimal.

# 8 Political Correctness

So far, the agent has been an expert who wants to show his skills. We now assume that skills are known but what is uncertain are preferences (Crawford-Sobel, Morris).

The model is modified as follows (See Morris). An agent of type $g$ has utility $u(a,x) + k\pi$, where $k$ is a positive number and $\pi$ is the posterior $(\Pr(\theta = g|I))$. An agent of type $b$ has utility $a + h\pi$, where $h$ is a positive number. An agent, irrespective of his type, receives signal $y$ with distribution

$$\begin{aligned} \Pr(y=1|x=1) &= q_1; \\ \Pr(y=1|x=0) &= q_0. \end{aligned}$$

The principal's utility is $u(a,x) + v(\pi)$, where $v$ is a convex and increasing function of the posterior. The rest of the model is as before. The principal must choose whether to obsreve only $u$ or also $a$.

The analysis is quite similar to the previous case. To make matters simpler, we focus on the "reasonable" informative equilibrium in which the an agent of type $g$ chooses $a = x$ and an agent of type $b$ chooses $a = 1$.

If the principal observes only $u$, such equilibrium exists if agent with $g$ does not want to deviate

$$\begin{aligned} E\left(u(1,x)|y=1\right) + kE(\pi\left(u=1\right)|a=1,y=1) &\geq E\left(u(0,x)|y=1\right) + kE(\pi\left(u=1\right)|a=0,y=1)\,; \\ E\left(u(0,x)|y=0\right) + kE(\pi\left(u=1\right)|a=0,y=0) &\geq E\left(u(1,x)|y=0\right) + kE(\pi\left(u=1\right)|a=1,y=0)\,; \end{aligned}$$

and the agent with $b$ does not want to deviate

$$1 + hE(\pi(u=1)|a=1, y=1) \geq hE(\pi(u=1)|a=0, y=1);$$
$$1 + hE(\pi(u=1)|a=1, y=0) \geq hE(\pi(u=1)|a=0, y=0);$$

The first two conditions are equivalent to (**??**). The third condition is implied by the first. The fourth condition imposes an upper bound to $h$. It is satisfied if $h$ is below a threshold $\bar{h}$. We assume that these conditions are satisfied. If only $u$ is observed, there is a "reasonable" separating equilibrium.

[TO BE CONTINUED]

## 9 When the agent knows his own type

[TO DO]

## 10 Conclusion

This paper has identified a precise set of circumstances under which committing to concealing a certain kind of information can make the principal better off. First, we must be in a world in which performance-related contracts cannot be signed. Second, the agent should be an expert, in the sense that his career depends on how able he is perceived to understand the state of the world. Third, the information about the agent's behavior should be separable into a part that is directly utility-relevant for the principal and a part that is not. If these conditions are met, then revealing the non-directly utility-relevant signal may make the agent behave in a more conformist way, which worsens both discipline and sorting.

Are the theoretical lessons learnt in this paper useful for understanding existing institutional arrangements? The idea that more information about non-directly utility-relevant information may induce the agent to behave in a suboptimal way because of career concerns is clearly present in political writings. In its famous 1974 ruling related to the Watergate case (US *vs.* Nixon), the US Supreme Court uses the following argument to defend the principle behind executive privilege: "Human experience teaches us that those who expect public dissemination of their remarks may well temper candor with a concern for appearances and for their own interest to the detriment of the decision-making process." Britain's Open Government code of practice uses a similar rationale when it provides that "internal discussion and advice can only be withheld where disclosure of the information *in question* would be harmful to the frankness and candour of future discussions." (Campaign for Freedom Information [5, p. 3]).

More precise implications can be extracted from Proposition 9. The optimal extent to which action is revealed is increasing in the extent to which consequence is revealed. We should expect transparency on decisions to go hand in hand with transparency on consequences. In particular, an action, or the intention to take an action, should not be revealed *before* the consequences of the action are observed. Indeed, this simple principle governs open government policy in the 30-plus countries that have adopted freedom of information legislation ( [11]). For instance, Sweden, the country with the oldest freedom of information act, does not recognize the right for citizens to obtain information about a public decision until that decision is implemented. Thus, working papers and internal recommendations that lead to a decision are released only when voters also have a chance to form an opinion on the consequence of the decision in question.[7]

The result on complementarity has also another implication. If, for exogenous reasons, citizens are less likely to observe the consequence, optimal institutional design dictates less transparency with regards to action. This may help explain that EU-level bodies are less transparent than the corresponding institututions at the national level. For instance, the meetings of the highest legislative body of each EU country are usually public, while, as we saw earlier, the Council of the European Union meets behind closed doors. There is no doubt that Europeans find it easier to evaluate the consequences of policy in areas that are typically under national jurisdiction (health, pensions, education, transports, etc...) rather than areas mainly under EU control (harmonization policy, competition policy, agricoltural subsidies, etc...). According to ourreults, the exogenous differential of information on payoff-relevant observables (laws) is optimally associated to a differential of information on non-payoff-relevant observables (positions during meetings). A Council in which debates were public would risk to give its members so strong an incentive to conform to citizens' expectations that its meetings would lose their information aggregating function.[8]

A competing explanation of why the meetings of the Council are secret has to do with bargaining costs (Samuelson...). Council members can be seen as agents of their respective countries and decision-making in the council as a bargaining game. If bargaining positions are publicly observable and bargainers have career concerns, then they may try to signal their type by acting tougher. This creates inefficiencies and may induce countries to commit not

[7]A historical example of this transparency policy is the US Constitutional Convention. George Mason refers to the secrecy of the Convention meetings as "a proper precaution" because it averted "mistakes and misrepresentations until the business shall have been completed, when the whole may have a very different complexion from that in which the several parts might in their first shape appear if submitted to the public eye" (Farrand [9, 3:28,32])

[8]The view that keeping Council meetings secret is desirable is often found in the writings of scholars of European politics. For instance, Calleo [4, p. 270-271] states that "Whether making Council debates more open is, of course, debatable. Discrete decision making, dominated by expert advisers, has its advantages, especially in periods of prolonged economic difficulty."

to observe bargaining positions. Note, however, that there is nothing in this theory that says that, as the decisions taken by the Council become more important, bargaining should become more transparent (if anything, secrecy should be more valuable). Instead, in our theory, the more Europeans care about the consequences of the Council decisions (and presumably they get informed about them in the same way they follow national law making), the more likely it is that a full transparency policy becomes optimal.

Lastly, we briefly relate our theory to transparency in corporate governance. Shareholders receive information about the management of their firm from the accounting reports that the firm makes. Clearly, accounting involves a great deal of aggregation both across time and across areas. Accounting research has been very active on the issue of the optimal degree of disaggregation. One point that is particularly debated, both among researchers and policy-makers, is whether a firm should provide disaggregated data about its productive segments (*segment disclosure*) on a quarterly basis or just on a yearly basis (Leuz and Verrecchia **??**). Currently, in the US there is no legal requirement for quarterly segment disclosure: some firms follow a disclosure policy and others do not. Evidence on whether segment disclosure improves firm performance is inconclusive (Botosan and Harris [3]). Without quarterly segment disclosure, shareholders still have information about short term consequences (from quarterly aggregated reports). What they have difficulty with is inferring the strategy that the firm is following, especially with regard to resource allocation across productive areas. Segment disclosure can then be seen as an improvement in transparency over action. Thus, the present theory provides an additional angle to evaluate the optimality of segment disclosure.[9]

## References


[1] Timothy Besley and Robin Burgess. The political economy of government responsiveness: Theory and evidence from India. Working paper, 2001.

[2] Arnoud W. A. Boot, Todd T. Milbourn, and Anjan V. Thakor. Sunflower management and capital budgeting. Working paper, March 2001.

[3] Christine A. Botosan and Mary S. Harris. Motivations for a change in disclosure frequency and its consequences: An examination of voluntary quarterly segment disclosures. *Journal of Accounting Research* 38(2): 329–353, 2000.

[4] David P. Calleo. *Rethinking Europe's Future.* Princeton University Press, 2001.


[9]Most existing work in accounting theory predicts that firms should adopt transparency policies, but see Nagar **??** for a reason why risk averse managers may want to limit disclosure.

[5] The Campaign for Freedom of Information. *Freedom of Information: Key Issues*. 1997 (available on www.cfoi.org.uk/pdf/keyissues.pdf).

[6] Jacques Crémer. Arm's length relationships. Quarterly Journal of Economics 110(2): 275–295.

[7] Mathias Dewatripont, Ian Jewitt, and Jean Tirole. The economics of career concerns, Part I: Comparing information structures. *Review of Economic Studies* 66(1): 183–198, 1999.

[8] Alexander Dyck and Luigi Zingales. Why are private benefits of control so large in certain countries and what effects does this have on their financial development? Working paper, 2001.

[9] Max Farrand (ed.). *The Records of the Federal Convention of 1787*. Yale University Press, 1967.

[10] John Fingleton and Michael Raith. Career concerns for bargainers. Working paper, October 2001.

[11] Maurice Frankel. Freedom of information: Some international characteristics. Working paper, The Campaign for Freedom of Information, 2001 (available on www.cfoi.org.uk/pdf/amsterdam.pdf).

[12] Bengt Holmström. Managerial incentive problems: A dynamic perspective. *Review of Economic Studies* 66(1): 169–182, 1999.

[13] Bengt Holmström. Moral hazard and observability. *Bell Journal of Economics* 10: 74–91, 1979.

[14] Christian Leuz and Robert E. Verrecchia. The economic consequences of increased disclosure. *Journal of Accounting Research* 38(supplement): 91–124, 2000.

[15] Stephen Morris. Political correctness. *Journal of Political Economy*, forthcoming.

[16] Venky Nagar. The role of the manager's human capital in discretionary disclosure. *Journal of Accounting Research* 37(supplement): 167–185, 1999.

[17] Marco Ottaviani and Peter Sørensen. Information aggregation in debate: Who should speak first? *Journal of Public Economics* 81: 393–421, 2001.

[18] Marco Ottaviani and Peter Sørensen. Professional advice. Working paper, September 2001.

[19] Motty Perry and Larry Samuelson. Open- versus close-door negotiations. *RAND Journal of Economics* 25(2): 348–59, 1995.

[20] Canice Prendergast. A theory of "Yes Men". *American Economic Review* 83(4): 757–770, 1993.

[21] Canice Prendergast and Lars Stole. Impetuous youngsters and jaded oldtimers. *Journal of Political Economy* 104: 1105–34, 1996.

[22] Mark J. Rozell. *Executive Privilege: The Dilemma of Secrecy and Democratic Accountability.* Johns Hopkins University Press, 1994.

[23] David Scharfstein and Jeremy Stein. Herd behavior and investment. *American Economic Review* 80: 465–479, 1990.

[24] Russell B. Stevenson, Jr. *Corporations and Information: Secrecy, Access, and Disclosure.* Johns Hopkins University Press, 1980.