The Wrong Kind of Transparency[1]

Andrea Prat

London School of Economics[2]

March 22, 2005

[1] I thank Heski Bar-Isaac, Roman Inderst, Alessandro Lizzeri, George Mailath, Margaret Meyer, Stephen Morris, Marco Ottaviani, David Stavasage, two anonymous referees, an editor, and a number of seminar participants for useful comments.

[2] Department of Economics and STICERD, Houghton Street, London WC2A 2AE, United Kingdom. Email: a.prat@lse.ac.uk. Homepage: econ.lse.ac.uk/staff/prat.

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. When the latter kind of information is available, the agent faces an incentive to disregard useful private signals and act according to how an able agent is expected to act a priori. This conformist behavior hurts the principal in two ways: the decision made by the agent is less likely to be the right one (discipline) and ex post it is more difficult to evaluate the agent's ability (sorting). The paper identifies a necessary and sufficient condition on the agent signal structure under which the principal benefits from committing not to observe the agent's action. The paper also shows the existence of strategic complementarities between information on action and information on consequence.

1 Introduction

There is a widespread perception, especially among economists, that agency relationships should be as transparent as possible. By transparency, we mean the ability of the principal to observe how the agent behaves and what the consequences of such behavior are. The idea is that transparency improves accountability, which in turn aligns the interests of the agent with those of the principal. Bengt Holmström (1979) has shown that in moral hazard problems more information about the agent is never detrimental to the principal, and, under mild assumptions, it is strictly beneficial. Should one conclude that whenever it is technologically feasible and not extremely expensive the principal should observe everything that the agent does?

Before asking what the optimal policy is, let us note that in practice we observe systematic deviations from full transparency in agency relationships in delegated portfolio management, corporate governance, and politics.

In delegated portfolio management, one might expect a high degree of transparency between the principal (the fund manager) and the agent (the investor). Instead, investors are typically supplied with limited information on the composition of the fund they own. Currently, the US Securities and Exchange Commission requires disclosure every six months, which consists of a portfolio snapshot at a particular point in time and can easily be manipulated by re-adjusting the composition just before and after the snapshot is taken – a practice known as "window dressing". It would be easy and almost costless to have more frequent disclosure by requiring mutual funds to publicize their portfolio composition on the internet. Yet there is strong resistance from the industry to proposals in the direction of more frequent disclosure (Craig S. Tyle, 2001).

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 shareholders and managers (Russell B. Stevensons, Jr. , 1980): "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'."

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 United States, the "executive privilege" allows the president to withhold information from the Congress, the courts, and the public (Mark J. Rozell, 1994). 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. Over thirty countries have passed Open Government codes, which establish the principle that a citizen should be able to access any public document. There are, however, important types of information, such as pre-decision material, that are often exempt from this requirement (Maurice Frankel, 2001).

Are the observed deviations from transparency in some sense optimal, or are they just due to inefficient arrangements that survive because of institutional inertia or resistance from entrenched interests? To answer this question, we need to establish what arguments can be made against transparency.

One obvious candidate explanation is that information revealed to the principal would also be revealed to a third party who will make use of it in ways that hurt the principal. In the political arena, voters may choose to forego 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 competitors learn it. In delegated portfolio management, *real time* disclosure could damage a fund because its investment strategy could be mimicked or even anticipated by competitors.[1]

The "third-party rationale" for keeping information secret presumably entails a tradeoff

[1] However, the SEC proposed reform allows for a time lag – usually sixty days – that is judged to be sufficient to neutralize free riding and front running.

between damage from information leaks and weaker incentives for the agent. This paper will instead look for an "agency rationale": a desire 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. Holmström's (1979) results on the optimality of information revelation in moral hazard problems suggest that the agency rationale should be explored in contexts in which, for exogenous reasons, there is no full contracting on observables. We will focus our attention on career concern models (Bengt Holmström, 1999), in which the principal and the agent can sign only short-term non-contingent contracts.[2]

The agency literature has already identified instances in which more information can hurt the principal. Holmström (1999) noted that more precise information about the agent's type reduces the incentive for the agent to work hard in order to prove his worth. Mathias Dewatripont et al. (1999) present examples in which the agent works harder if the principal receives a coarser signal on agent performance rather than observing performance directly. Jacques Crémer (1995) shows that in a dynamic contracting model where renegotiation is possible the principal may be hurt by observing a precise signal on agent performance because it makes the commitment to non-renegotiation less credible. In these three instances, transparency is bad for discipline (the agent works less) but it is good for sorting (it is easier to identify agent type).

The rationale for secrecy considered in the present paper is entirely different. It does not hinge on the danger that the agent exerts less effort, as in the papers above, but rather on the possibility that the agent disregards useful private signals. We employ a model of career concerns for "experts": the agent's type determines his ability to understand the state of the world.[3] We distinguish between two types of information that the principal can have about his agent: information about the *consequences* of the agent's action and

[2] As Robert Gibbons and Kevin J. Murphy (1992) show, there are still strong career concern incentives even when contracts are contingent on observables. Thus, the crucial assumption we make is that long-term contracts are not available.

[3] See David Scharfstein and Jeremy Stein (1994), Jeffrey Zwiebel (1995), Canice Prendergast and Lars Stole (1996), Colin Campbell (1998), Marco Ottaviani and Peter Sørensen (2001, 2003), Gilat Levy (2004), Jeffrey Ely and Juuso Välimäki (2003), and Jeffrey Ely et al. (2002).

information directly about the *action*. The main contribution of this paper is to show that, while transparency on consequences is beneficial, transparency on action can have detrimental effects. When the latter kind of information is available, the agent faces an incentive to disregard useful private signals and to act according to how an able agent is expected to act a priori. This conformist behavior hurts the principal both through discipline (the agent's action is less aligned with the principal's interest) and sorting (it is impossible to discern the agent's ability). If that is the case, the principal wants to commit to keep the agent's action secret.

The present work is particularly related to two papers on experts. Canice Prendergast (1993) 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 bias his report toward the principal's signal. Misreporting on the part of the agent causes a loss of efficiency. For this reason, the principal may choose to offer the agent a contract in which pay is independent of action. This will induce minimum effort exertion but also full honesty. Christopher Avery and Margaret M. Meyer (1999) ask whether in a model of career concerns for advisors who may be biased it is beneficial from the point of view of principal to keep track of the advisor's past recommendations. They argue that in certain circumstances observing past recommendations worsens discipline and does not improve sorting.

The plan of the paper is as follows. Section 2 introduces the career concern game. Section 3 begins with a simple example to illustrate why transparency on action can be damaging. We then prove the main technical result, a characterization of the set of perfect Bayesian equilibria under the two information scenarios, concealed action and revealed action, and we use this result to perform a welfare analysis. Section 4 studies the complementarity between action observation and consequence observation. Section 5 concludes.

2 A Model of Career Concerns for Experts

To make our main point, it is sufficient to consider a simple model in which the agent's action, type, signal, and consequence are all binary. There are a principal and an agent. The agent's type $\theta \in \{g, b\}$ is unknown to both players. The prior 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 θ 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.[4]

The principal does not know 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 $q_{x\theta} = \Pr(y = 1|x,\theta)$. We assume that

$$0 < q_{0g} < q_{0b} < q_{1b} < q_{1g} < 1. \tag{1}$$

This means that the signal is informative (because $\Pr(x = 1|y)$ is increasing in y and $\Pr(x = 0|y)$ is decreasing in y) and that the signal is more informative for the better type (because $\Pr(x = y|y,g) > \Pr(x = y|y,b)$).

These assumptions alone are not sufficient to guarantee that the signal is useful. For instance, if the prior p on x is very high or very low, it is optimal to disregard y. To make the problem interesting, we also assume that the signal y is *decision-relevant*, that is:

$$(q_{1g}\gamma + q_{1b}(1-\gamma))\,p + ((1-q_{0g})\,\gamma + (1-q_{0b})(1-\gamma))(1-p) > \max(p, 1-p). \tag{2}$$

It is easy to check that (2) implies that an agent who observes realization y knows that the probability that the signal is correct is greater than 50%. Formally, for $y \in \{0,1\}$, $\Pr(x = y|y) > \Pr(x = 1-y|y)$.

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. The principal's belief that the agent's type is g is $\pi(I)$, where I is the information

[4] A more general version of the model, in which variables are not binary, is available in Andrea Prat (2003).

available to the principal. With concealed action, if the principal observes consequence $\hat{u}$, the belief is

$$\tilde{\pi}(\hat{u}) = \Pr(\theta = g | u = \hat{u}) = \frac{\gamma \Pr(u = \hat{u} | \theta = g)}{\Pr(u = \hat{u})}.$$

With revealed action, the principal is able to infer x from a and u. The agent's belief, assuming that a is played in equilibrium with positive probability, is

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

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 principal's belief $\pi(I)$. The payoff to the principal depends on the consequence and on the posterior distribution: $u(a, x) + v(\pi(I))$, where v is a convex function of π. This model should be taken as a reduced form of a two-period career concerns model in which the principal can choose to retain the first-period agent or hire another one. Convexity is then a natural assumption because the principal's expected payoff is the upper envelope of the expected payoffs provided by the incumbent agent and the challengers. More information about the incumbent can only be beneficial.

Given any equilibrium strategy α^*, the ex ante expected payoff of the agent must be γ, 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 α^*, the expected payoff of the principal can also be taken as total welfare.

We sometimes refer to a perfect Bayesian equilibrium simply as an "equilibrium". An equilibrium is *informative* if $\alpha_0^* \neq \alpha_1^*$ and *pooling* if $\alpha_0^* = \alpha_1^*$. An informative equilibrium is *separating* if either $\alpha_0^* = 0$ and $\alpha_1^* = 1$ or $\alpha_0^* = 1$ and $\alpha_1^* = 0$. An informative equilibrium is *semi-separating* if it is not separating, i.e. if at least one of the two types of agents uses a mixed strategy. An informative equilibrium is *perverse* if the agent chooses the 'wrong' action given his signal: $\alpha_0^* > \alpha_1^*$.

Let $E_{revealed}$ and $E_{concealed}$ be the sets of perfect Bayesian equilibria in the two possible information scenarios. Given the existence of babbling equilibria, it is clear that the sets are nonenmpty. Let $W_{revealed}$ be the supremum of $w(\alpha^*)$ in $E_{revealed}$ and let $W_{concealed}$ the corresponding value when the action is concealed. The main question that we shall ask is whether $W_{revealed} \geq W_{concealed}$.

Attention should be drawn to two assumptions. First, assuming that the agent's payoff is belief $\pi(I)$, rather than an arbitrary function of belief $\pi(I)$, is not without loss of generality (see Ottaviani and Sørensen, 2003 for a discussion of this point). The assumption is made by most papers in career concerns because it makes the analysis simpler. Second, the analysis is also facilitated by assuming that the agent does not know his own type (again, Ottaviani and Sørensen, 2003 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.[5]

Finally, we introduce a notion that corresponds to a mental experiment. Suppose the principal could observe the agent signal y directly. Which of the two realizations of the signal y is better news about the agent type? This corresponds to comparing $\Pr(\theta = g|y = 1)$ with $\Pr(\theta = g|y = 0)$. We exclude the nongeneric case in which the two probabilities are identical. If $\Pr(\theta = g|y = 1) > \Pr(\theta = g|y = 0)$ we say that $y = 1$ is the *smart realization* of the agent signal. If $\Pr(\theta = g|y = 1) < \Pr(\theta = g|y = 0)$, we say that $y = 0$ is the smart realization.

Smartness can be related to the primitives of the problem. The smart realization is $y = 1$ if and only if

$$\frac{q_{0b} - q_{0g}}{q_{1g} - q_{1b}} < \frac{p}{1-p}. \tag{3}$$

If the two states of the world are equiprobable, this means that

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

That is, the difference between the probability that the good type gets a correct signal and the probability that the bad type gets a correct signal must be greater if $x = 1$ than if $x = 0$. Then, observing $y = 1$ raises the principal's belief above γ while observing $y = 0$ decreases it.

[5] Prat (2003) analyzes the case in which the agent has information about his type. While the incentive to behave in a conformist way is softened, the main results are confirmed.

3 The Effects of Transparency

In this section, we begin with a simple example of how disclosing the agent's action generates conformism. We then analyze separately the concealed action scenario and the revealed action scenario. The main result is a necessary and sufficient condition on the primitives of the game under which the principal is better off committing to keep the action concealed.

3.1 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 an uninformative signal. A good agent observes the state $x = 1$ with certainty and gets pure noise if the state is $x = 0$.[6]

The key assumption is that the importance of ability is highly state-specific. A good agent is better than a bad agent at recognizing $x = 1$ but not $x = 0$. As a result, the signal realization $y = 1$ is good news about the agent's type and realization $y = 0$ is bad news. Indeed, one can check that the smart realization is $y = 1$ because $\Pr(\theta = g|y = 0) = \frac{1}{3}$ and $\Pr(\theta = g|y = 1) = \frac{3}{5}$.

We now argue that in this example transparency on action induces complete conformism and it damages the principal.[7] First, consider the revealed action scenario and suppose that there exists a separating equilibrium in which the agent plays $a = y$. The principal's belief $\pi(a, x)$ in such a separating equilibrium is:

$$\pi(1,1) = \frac{2}{3}; \quad \pi(1,0) = \frac{1}{2}; \quad \pi(0,1) = 0; \quad \pi(0,0) = \frac{1}{2}.$$

The belief when $a = 1$ dominates the one when $a = 0$, in the sense that for any realization of x, $\pi(1, x) \geq \pi(0, x)$. The agent who observes $y = 0$ has a strict incentive to report $a = 1$ instead of $a = 0$. Therefore, this cannot be an equilibrium

[6] The example violates some of the strict inequalities in (1). A slight modification of this example would satisfy (1) and it would lead to similar conclusions.

[7] We say "argue" rather than "prove" because in this section we restrict attention to pure-strategy equilibria (separating or pooling). The next section will provide a full analysis, including semi-separating equilibria.

A similar argument shows that there is no perverse separating equilibrium (in which $a = |1 - y|$). The only remaining pure-strategy equilibria are then pooling equilibria in which no information is revealed (either the agent always plays $a = 0$ or he always plays $a = 1$).[8] 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). Thus, with revealed action, the best equilibrium for the principal is one in which her expected payoff in the current period is $\frac{1}{2}$ and her posterior belief is the same as her prior.

Instead, in the concealed action scenario there exists a separating equilibrium in which the agent plays $a = y$. To see this, compute the belief $\tilde{\pi}(u)$ about the agent's type in such an equilibrium: $\tilde{\pi}(1) = \frac{3}{5}$ and $\tilde{\pi}(0) = \frac{2}{5}$. The agent maximizes the expected belief by maximizing the expected value of u. As the signal y is decision-relevant, this means that the optimal strategy is $a = y$. In this separating equilibrium, the probability that the principal gets utility 1 in the first period is $\Pr(u = 1) = \frac{5}{8}$. Thus, with concealed action, the principal receives an expected payoff of $\frac{5}{8}$ in the first period and she learns something about the agent type.

To sum up, by committing to keep the action concealed, the principal gets a double benefit. On the discipline side, she increases her expected payoff in the current period because the agent follows his signal. On the sorting side, she improves the precision of her posterior distribution on her agent type.

3.2 Equilibria

We now return to the general game introduced in Section 2. We begin by looking at what happens when the principal observes only the consequence, which turns out to be the easier part.

The principal's belief after observing u is $\tilde{\pi}(u) = \Pr(\theta = g|u)$. The agent observes his signal y and maximizes $E_x[\tilde{\pi}(u(a, x))|y]$. If the agent plays $a = y$, by (1), $\tilde{\pi}(u = 1) >$

[8] Pooling equilibria are supported by beliefs that are independent of the consequence: $\Pr(\theta = g|a, u) = \Pr(\theta = g|a)$ for $u \in \{0, 1\}$. For instance, in the equilibrium where the agent always plays $a = 1$ the on-the-equilibrium-path belief $\Pr(\theta = g|a = 1)$ equals the prior γ and the out-of-equilibrium belief $\Pr(\theta = g|a = 0)$ can be set at any value in $[0, \gamma]$.

$\gamma > \tilde{\pi}(u = 0)$. As the signal y is by assumption decision-relevant, it is a best response for teh agent to play $a = y$. Therefore,

Proposition 1 *With concealed action, there exists a non-perverse separating equilibrium.*

The analysis of the concealed action case is straightforward. In a non-perverse separating equilibrium the principal's belief is higher in case of success than in case of failure. But then the agent should maximize the probability of success, which means choosing $a = y$. Hence, a separating equilibrium exists. There may be other equilibria: uninformative, perverse separating, semi-separating. But the non-perverse separating equilibrium above is clearly the best from the viewpoint of the principal.

The more difficult case is when the principal observes the action as well, because we need to deal with semi-separating equilibria. First, note that, as the consequence does not enter the agent's utility directly, there are always pooling equilibria. These equilibria are akin to babbling equilibria in cheap talk games: the agent uses the same strategy (pure or mixed) independently of his type, and in equilibrium the principal learns nothing about the agent's private information.[9]

Next, we ask whether there exist equilibria that are, at least partly, informative:

Proposition 2 *With revealed action, there exists a non-perverse separating equilibrium if and only if*

$$\frac{p}{1-p}\frac{\gamma q_{0g} + (1-\gamma)\, q_{0b}}{\gamma q_{1g} + (1-\gamma)\, q_{1b}} \leq \frac{q_{0b} - q_{0g}}{q_{1g} - q_{1b}} \leq \frac{p}{1-p}\frac{\gamma\,(1-q_{0g}) + (1-\gamma)\,(1-q_{0b})}{\gamma\,(1-q_{1g}) + (1-\gamma)\,(1-q_{1b})}. \qquad (4)$$

There exists an informative equilibrium if and only if there exists a non-perverse separating equilibrium.

Proof. See Appendix. ■

[9] The present model is not, strictly speaking, a cheap-talk game because the principal's utility is directly affected by the agent's action. Still, the fact that the action is costless to the agent makes standard refinements, like the Intuitive Criterion, irrelevant. See Narvin Kartik (2003) for a discussion of the role of refinements in cheap talk.

To understand the result, note that

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

We can link condition (4) with the smartness condition in (3). Both impose bounds on the informativeness ratio $\frac{q_{0b}-q_{0g}}{q_{1g}-q_{1b}}$. The smartness condition establishes which realization is better news on the agent's type. The condition in Proposition 2 says whether one realization is *much* better news than the other.

Suppose for instance that $y = 1$ is the smart signal. In that case, we can disregard the second inequality in Proposition 2 because it is implied by the smartness condition in (3). Instead, the first inequality may or may not hold. If it fails, there is no informative equilibrium: if the equilibrium were informative, the agent who observes $y = 0$ would always want to pretend he observed $y = 1$. If instead the first inequality holds, 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 we revisit the example presented earlier, we can now formally verify the result that there is no informative equilibrium. The smart signal is $y = 1$. There exists an informative equilibrium if and only if the first inequality is satisfied. 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 smart signal had been "less smart", 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: $q_{0g} \in \left(0, \frac{1}{2}\right)$. The existence condition (4) shows that, holding the other parameters constant, there exists an informative equilibrium if and only if $q_{0g} \leq \frac{1}{4}$.

3.3 When should the action be revealed?

We are now in a position to compare the expected payoff of the principal in the best equilibrium under concealed action with her expected payoff in the best equilibrium with revealed action. As we saw in Section 2, ex ante social welfare corresponds to the expected payoff of the principal because the expected payoff of the agent is constant.

From Proposition 1, the best equilibrium with concealed action is a separating equilibrium with $a = y$. What happens with revealed action depends on condition (4). If the condition holds, there exists a separating equilibrium with $a = y$. The agent's behavior is thus the same as with concealed action but the principal gets more information. The variance of the posterior belief increases and the principal's payoff goes up. Compared to concealed action, the discipline effect is the same but the sorting effect improves. If instead condition (4) 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 useful information. Sorting too is affected negatively because in an uninformative equilibrium the posterior belief is equal to the prior. Thus, the principal is worse off. We summarize the argument as follows:

Proposition 3 *If (4) holds, revealing the agent's action does not affect discipline and improves sorting. If (4) fails, revealing the agent's action worsens both discipline and sorting. Hence, the principal prefers to reveal the action if and only if (4) holds.*

There exists a fundamental tension between what is optimal ex ante and what is optimal ex post. Suppose we are in a separating equilibrium. After the agent has chosen his action, the principal always benefits from observing the action because she can use the additional information for sorting purposes. Moreover, the sorting benefit is particularly large when the informativeness ratio in (4) is very high or very low. However, before the agent has chosen his action, the principal may want to commit not to observe the action ex post. Indeed, a separating equilibrium is unlikely to exist when the informativeness ratio is very high or very low. The principal opposes action disclosure ex ante exactly when she benefits most from action disclosure ex post.

4 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 sorting purposes. Take for instance a large-scale public project, such as a reform of the health system. Its main provisions are observable right away, but it takes years for its effects to develop. In the medium term, the public knows the characteristics of the project that has been undertaken (the action) but cannot yet judge its success (the consequence).

This section looks at what happens when consequences are imperfectly observed. 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 restrict attention to pooling and separating equilibria. We look at 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 (ρ_u, ρ_a), we ask whether the separating equilibrium exists.[10]

Proposition 4 *For every ρ_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 ρ_a^* is nondecreasing in ρ_u.*

Proof. Appendix. ■

Conformism is deterred only by the fear of failure. An agent who has observed the non-smart realization is tempted to try to fool the principal by playing the action that corresponds to the smart consequence. If the consequence is not observed, the trick succeeds. If, however, the consequence is observed, the agent is likely to generate $u = 0$ and to obtain a low posterior belief. Hence, the incentive to pool on the smart realization is decreasing in the degree of transparency on consequence. If, for exogenous reasons, the

[10] We cannot exclude (although we have no example) that there are semi-separating equilibria that exist even when there exists no separating equilibrium. For that reason, we cannot use Proposition 4 to draw normative conclusions on sorting (but we can use it for discipline).

consequence is easy to observe, the principal can afford to have more transparency on action as well, without creating incentives for conformism.

5 Conclusion

We have identified circumstances in which a career-driven agent who knows that his action is observed has an incentive to behave in a conformist manner. The principal is damaged by such behavior and she may want to commit to keeping the agent's action secret. Is this theoretical finding useful for understanding existing institutional arrangements?

Our results resonate with informal arguments that have been used to justify certain forms of secrecy.[11] 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, 1997).

More precise empirical implications can be extracted from Proposition 4. 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, Frankel (2001) reports that all the 30-plus countries that have adopted an open government code allow for some form of short-term secrecy while the decision process is still ongoing. For instance, Sweden, the country with the oldest and, perhaps the most forceful, freedom of information act, does not recognize the right for citizens to obtain information about a public decision until that decision is implemented. Working papers and internal recommendations that lead to a

[11] See Eric Maskin and Jean Tirole (forthcoming) for a general analysis of career concerns in public decision-making.

decision are released only when voters begin to have a chance to form an opinion on the consequence of the decision in question.[12]

With regards to delegated portfolio management, there is a potential link with Josef Lakonishok et al. (1992). They compare returns for the equity-invested portion of mutual funds and pension funds in United States. Their evidence suggests that pension funds underperform mutual funds. This is a surprising finding because pension funds are typically monitored by professional investors with large stakes (the treasury division of the company that sponsors the pension plan), while mutual funds are held by a very large number of individuals who presumably exert less monitoring effort. One of the hypotheses that Lakonishok et al. advance is that the ability of pension fund investors to monitor their funds closely actually creates an agency problem. Mutual fund investors typically choose funds only based on yearly returns, while pension fund investors select funds only after they communicate directly with fund managers who explain their investment strategy. The results of Lakonishok et al. are consistent with our prediction that transparency over action (the investment strategy in this case) can lead to worse performance.

References

Avery, Christopher and Meyer, Margaret M. "Designing Hiring and Promotion Procedures When Evaluators Are Biased." Working paper, 1999.

The Campaign for Freedom of Information. *Freedom of Information: Key Issues.* 1997.

Campbell, Colin. "Learning and the Market for Information." Working paper, Ohio State University, 1998.

[12] A historical example of a policy of short-term secrecy is supplied by 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" (Max Farrand, 1967)

Crémer, Jacques. "Arm's Length Relationships." *Quarterly Journal of Economics*, May 1995, 110(2), pp. 275–295.

Dewatripont, Mathias; Jewitt, Ian and Tirole, Jean. "The Economics of Career Concerns, Part I: Comparing Information Structures." *Review of Economic Studies*, January 1999, 66(1), pp. 183–198.

Ely, Jeffrey; Fudenberg, Drew and Levine, David K. "When is Reputation Bad?" Working paper, Northwester University, 2002.

Ely, Jeffrey and Välimäki, Juuso. "Bad Reputation." *Quarterly Journal of Economics*, August 2003, 118(3), pp. 785–813.

Farrand, Max (ed.). *The Records of the Federal Convention of 1787.* New Haven: Yale University Press, 1967.

Frankel, Maurice. "Freedom of Information: Some International Characteristics." Working paper, The Campaign for Freedom of Information, 2001.

Gibbons, Robert and Murphy, Kevin J. "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence." *Journal of Political Economy*, June 1992, 100(3), pp. 468–505.

Holmström, Bengt. "Moral Hazard and Observability." *Bell Journal of Economics*, Spring 1979, 10, pp. 74–91.

Holmström, Bengt. "Managerial Incentive Problems: A Dynamic Perspective." *Review of Economic Studies*, January 1999, 66(1), pp. 169–182.

Kartik, Navin. "Information Transmission with Cheap and Almost-cheap Talk." Working paper, Stanford University, 2003.

Lakonishok, Josef; Shleifer, Andrei and Vishny, Robert W. "The Structure and Performance of the Money Management Industry." *Brookings Papers on Economic Activity*, 1992, Vol 1992, pp. 339–379.

Levy, Gilat. "Anti-Herding and Strategic Consultation." *European Economic Review*, June 2004, Vol. 48(3), pp. 503-525.

Maskin, Eric and Tirole, Jean. "The Politician and the Judge: Accountability in Government." *American Economic Review*, September 2004, Vol 94(4), pp. 1034–1054.

Ottaviani, Marco and Sørensen, Peter. "Information Aggregation in Debate: Who Should Speak First?" *Journal of Public Economics*, September 2001, 81(3), pp. 393–421.

Ottaviani, Marco and Sørensen, Peter. "Professional Advice." *Journal of Economic Theory*, forthcoming.

Prat, Andrea. "The Wrong Kind of Transparency." April 2003, CEPR Discussion Paper 3859.

Prendergast, Canice. "A Theory of 'Yes Men'." *American Economic Review*, September 2003, 83(4), pp. 757–770.

Prendergast, Canice and Stole, Lars. "Impetuous Youngsters and Jaded Oldtimers." *Journal of Political Economy*, December 1996, 104(6), pp. 1105–34.

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

Scharfstein, David and Stein, Jeremy. "Herd Behavior and Investment." *American Economic Review*, June 1990, 80(3), pp. 465–479.

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

Tyle, Craig S. "Letter to the SEC on the Frequency of Mutual Fund Portfolio Holdings Disclosure." Investment Company Institute, 2001.

Zwiebel, Jeffrey. "Corporate Conservatism and Relative Compensation." *Journal of Political Economy*, February 1995, 103(1), pp. 1–25.

Appendix: Proofs

Proof of Proposition 2

The proposition is proven through three lemmas. We begin by excluding informative equilibria in fully mixed strategies:

Lemma 5 *There cannot exist an informative equilibrium in which* $\alpha_0 \in (0,1)$ *and* $\alpha_1 \in (0,1)$.

Proof. Assume that there exists an equilibrium in which $\alpha_0 \in (0,1), \alpha_1 \in (0,1), \alpha_0 \neq \alpha_1$. The agent must be indifferent between the two actions for both realizations of y :

$$\Pr(x=0|y=1)(\pi(0,0)-\pi(1,0)) = \Pr(x=1|y=1)(\pi(1,1)-\pi(0,1)), \quad (5)$$

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

There are two cases:

$$(\pi(0,0)-\pi(1,0))(\pi(1,1)-\pi(0,1)) \leq 0 \quad (7)$$

$$(\pi(0,0)-\pi(1,0))(\pi(1,1)-\pi(0,1)) > 0 \quad (8)$$

If (7) holds, note that in an informative equilibrium it cannot be that both $\pi(0,0)=\pi(1,0)$ and $\pi(1,1)=\pi(0,1)$. But then we have a contradiction because the two sides of (5) have different signs. If (8) holds, subtract (6) from (5)

$$\begin{aligned} & (\Pr(x=0|y=1)-\Pr(x=0|y=0))(\pi(0,0)-\pi(1,0)) \quad (9) \\ = \; & (\Pr(x=1|y=1)-\Pr(x=1|y=0))(\pi(1,1)-\pi(0,1)). \end{aligned}$$

But by assumption (1) signals are informative on x:

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

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

Then, (8) creates a contradiction in (9). ■

We further characterize the equilibrium set by showing that, if there exists an informative equilibrium, then there must also exist a (non-perverse) separating equilibrium:

Lemma 6 *There exists an equilibrium in which $\alpha_0 \neq \alpha_1$ if and only if there exists an equilibrium in which $\alpha_0 = 0$ and $\alpha_1 = 1$.*

Proof. We begin by expressing beliefs in terms of primitives and strategies. It is useful to make the dependence on strategies explicit (we use $\Pi(a, x, \alpha_0, \alpha_1)$ rather than $\pi(a, x)$):

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

To simplify notation in the proof, we use the following (slightly abusive) notation for special cases of $\Pi(a, x, \alpha_0, \alpha_1)$:

$$\begin{aligned}
\Pi(a, x) &\equiv \Pi(a, x, \alpha_0 = 0, \alpha_1 = 1) \\
\Pi(a, x, \alpha_1) &\equiv \Pi(a, x, \alpha_0 = 0, \alpha_1) \\
\Pi(a, x, \alpha_0) &\equiv \Pi(a, x, \alpha_0, \alpha_1 = 1)
\end{aligned}$$

Throughout the proof, assume without loss of generality that $y = 1$ is the smart realization. If $y = 0$ is the smart realization, just switch 0 and 1 for a, x, and y.

We begin by considering perverse informative equilibria. Suppose there exists an equilibrium in which $\alpha_0 > \alpha_1$, with beliefs $\Pi(a, x, \alpha_0, \alpha_1)$. For $y \in \{0, 1\}$, if a is played in equilibrium it must be that:

$$a \in \arg\max_{\tilde{a}} \sum_{x \in \{0,1\}} \Pr(x|y)\, \Pi(\tilde{a}, x, \alpha_0, \alpha_1).$$

But note that for every a, x, α_0, and α_1,

$$\Pi(a, x, \alpha_0, \alpha_1) = \Pi(1 - a, x, 1 - \alpha_0, 1 - \alpha_1).$$

Therefore, if the perverse equilibrium exists, there also exist a non-perverse equilibrium in which the agent plays $\hat{\alpha}_0 = 1 - \alpha_0$ and $\hat{\alpha}_1 = 1 - \alpha_1$, and beliefs are a mirror image of the initial beliefs: $\Pi(a, x, \hat{\alpha}_0, \hat{\alpha}_1) = \Pi(1 - a, x, \alpha_0, \alpha_1)$. The rest of the proof focuses on the existence of non-perverse informative equilibria ($\alpha_0 < \alpha$). We begin with a technical result that is useful later:

Claim 1: The smart realization is $y = 1$ if and only if

$$\begin{aligned} & \Pr(x=1)\Pr(y=1|x=1)\Pr(y=0|x=1)(\Pi(1,1)-\Pi(0,1)) \\ > \ & \Pr(x=0)\Pr(y=1|x=0)\Pr(y=0|x=0)(\Pi(0,0)-\Pi(1,0)). \end{aligned} \tag{10}$$

Proof of Claim 1: Note that

$$\begin{aligned} & \Pr(x)\Pr(y=1|x)\Pr(y=0|x)(\Pi(1,x)-\Pi(0,x)) \\ = \ & \frac{1}{\Pr(x)}(\Pr(g,y=1,x)\Pr(y=0,x)-\Pr(g,y=0,x)\Pr(y=1,x)) \\ = \ & \frac{1}{\Pr(x)}(\Pr(g,y=1,x)(\Pr(g,y=0,x)+\Pr(b,y=0,x))) \\ & -\frac{1}{\Pr(x)}(\Pr(g,y=0,x)(\Pr(g,y=1,x)+\Pr(b,y=1,x))) \\ = \ & \frac{1}{\Pr(x)}(\Pr(g,y=1,x)\Pr(b,y=0,x)-\Pr(g,y=0,x)\Pr(b,y=1,x)) \\ = \ & \Pr(b)\Pr(g)\Pr(x)(q_{xg}(1-q_{xb})-(1-q_{xg})q_{xb}) \\ = \ & \Pr(b)\Pr(g)\Pr(x)(q_{xg}-q_{xb}). \end{aligned} \tag{11}$$

By (3), the smart realization is $y = 1$ if and only if $p(q_{1g} - q_{1b}) > (1-p)(q_{0b} - q_{0g})$, which can be rewritten as

$$\Pr(b)\Pr(g)\Pr(x=1)(q_{1g}-q_{1b}) > \Pr(b)\Pr(g)\Pr(x=0)(q_{0b}-q_{0g}),$$

which, by the argument above, is equivalent to (10). Claim 1 is proven.

We now discuss separating equilibria. The necessary and sufficient conditions for the existence of a non-perverse separating equilibrium are:

$$\Pr(x=1|y=0)(\Pi(1,1)-\Pi(0,1)) \leq \Pr(x=0|y=0)(\Pi(0,0)-\Pi(1,0)) \tag{12}$$

$$\Pr(x=1|y=1)(\Pi(1,1)-\Pi(0,1)) \geq \Pr(x=0|y=1)(\Pi(0,0)-\Pi(1,0)) \tag{13}$$

Claim 2: The inequality (13) is always satisfied. There exists a separating equilibrium if and only if (12) holds.

Proof of Claim 2: Recall that $y = 1$ is the smart realization and note that $\Pr(x)\Pr(y=1|x) = \Pr(y=1)\Pr(x|y=1)$. Therefore, by Claim 1,

$$\begin{aligned} & \Pr(x=1|y=1)\Pr(y=0|x=1)(\Pi(1,1)-\Pi(0,1)) \\ > \; & \Pr(x=0|y=1)\Pr(y=0|x=0)(\Pi(0,0)-\Pi(1,0)). \end{aligned}$$

But (1) implies that $\Pr(y=0|x=0) > \Pr(y=0|x=1)$. Therefore (13) holds a fortiori. Claim 2 is proven.

From Lemma 5, there cannot exist an equilibrium in which $0 < \alpha_0 < \alpha_1 < 1$. There can be two forms of informative equilibria: either $\alpha_0 = 0$ and $\alpha_1 \in (0,1]$ or $\alpha_0 \in [0,1)$ and $\alpha_1 = 1$. Claims 3 and 4 deal with the two cases separately. Together, the claims prove that there exists an equilibrium with $\alpha_0 < \alpha_1$ only if there exists an equilibrium with $\alpha_0 = 0$ and $\alpha_1 = 1$.

Claim 3: There cannot exist an equilibrium in which $\alpha_0 = 0$ and $\alpha_1 \in (0,1)$.

Proof of Claim 3: Suppose there exists an equilibrium in which $\alpha_0 = 0$ and $\alpha_1 \in (0,1)$. It must be that

$$\Pr(x=1|y=1)(\Pi(1,1,\alpha_1)-\Pi(0,1,\alpha_1)) = \Pr(x=0|y=1)(\Pi(0,0,\alpha_1)-\Pi(1,0,\alpha_1)). \tag{14}$$

Because an agent who observes $y = 0$ never plays $a = 1$, we have $\Pi(1,x,\alpha_1) = \Pi(1,x)$. Note that

$$\begin{aligned} \Pi(0,x,\alpha_1) &= \frac{\Pr(a=0,g,x)}{\Pr(a=0,x)} = \frac{\Pr(y=0,g,x)+(1-\alpha_1)\Pr(y=1,g,x)}{\Pr(y=0,x)+(1-\alpha_1)\Pr(y=1,x)} \\ &= \frac{\Pr(y=0,x)\frac{\Pr(y=0,g,x)}{\Pr(y=0,x)}+(1-\alpha_1)\Pr(y=1,x)\frac{\Pr(y=1,g,x)}{\Pr(y=1,x)}}{\Pr(y=0,x)+(1-\alpha_1)\Pr(y=1,x)} \\ &= A(x,\alpha_1)\Pi(0,x)+(1-A(x,\alpha_1))\Pi(1,x), \end{aligned}$$

where

$$\begin{aligned} A(x,\alpha_1) &\equiv \frac{\Pr(y=0,x)}{\Pr(y=0,x)+(1-\alpha_1)\Pr(y=1,x)} \\ &= \frac{\Pr(y=0|x)}{\Pr(y=0|x)+(1-\alpha_1)\Pr(y=1|x)}. \end{aligned}$$

Condition (14) rewrites as

$$\begin{aligned} & \Pr\left(x=1|y=1\right) A\left(1,\alpha_1\right)\left(\Pi\left(1,1\right)-\Pi\left(0,1\right)\right) \\ = \; & \Pr\left(x=0|y=1\right) A\left(0,\alpha_1\right)\left(\Pi\left(0,0\right)-\Pi\left(1,0\right)\right). \end{aligned}$$

which in turn holds only if

$$\begin{aligned} & \Pr\left(x=1|y=1\right)\left(\Pi\left(1,1\right)-\Pi\left(0,1\right)\right) \\ \leq \; & \Pr\left(x=0|y=1\right) \max_{\alpha_1\in[0,1]} \frac{A\left(0,\alpha_1\right)}{A\left(1,\alpha_1\right)}\left(\Pi\left(0,0\right)-\Pi\left(1,0\right)\right). \end{aligned} \tag{15}$$

Note that

$$\begin{aligned} \max_{\alpha_1\in[0,1]} \frac{A\left(0,\alpha_1\right)}{A\left(1,\alpha_1\right)} &= \max_{\alpha_1\in[0,1]} \frac{\Pr(y=0|x=0)}{\Pr(y=0|x=1)} \frac{\Pr(y=0|x=1)+(1-\alpha_1)\Pr(y=1|x=1)}{\Pr(y=0|x=0)+(1-\alpha_1)\Pr(y=1|x=0)} \\ &= \frac{\Pr(y=0|x=0)}{\Pr(y=0|x=1)} \frac{\Pr(y=0|x=1)+\Pr(y=1|x=1)}{\Pr(y=0|x=0)+\Pr(y=1|x=0)} \\ &= \frac{\Pr(y=0|x=0)}{\Pr(y=0|x=1)}. \end{aligned}$$

Inequality (15) can thus be rewritten as

$$\begin{aligned} & \Pr\left(x=1|y=1\right)\Pr\left(y=0|x=1\right)\left(\Pi\left(1,1\right)-\Pi\left(0,1\right)\right) \\ \leq \; & \Pr\left(x=0|y=1\right)\Pr\left(y=0|x=0\right)\left(\Pi\left(0,0\right)-\Pi\left(1,0\right)\right), \end{aligned}$$

or

$$\begin{aligned} & \Pr\left(x=1\right)\Pr\left(y=1|x=1\right)\Pr\left(y=0|x=1\right)\left(\Pi\left(1,1\right)-\Pi\left(0,1\right)\right) \\ \leq \; & \Pr\left(x=0\right)\Pr\left(y=1|x=0\right)\Pr\left(y=0|x=0\right)\left(\Pi\left(0,0\right)-\Pi\left(1,0\right)\right), \end{aligned}$$

which is impossible by Claim 1. Claim 3 is proven.

Claim 4: If there exists an equilibrium in which $\alpha_0 \in [0,1)$ and $\alpha_1 = 1$, there exists an equilibrium in which $\alpha_0 = 0$ and $\alpha_1 = 1$.

Proof of Claim 4: A necessary condition for the existence of an equilibrium in which $\alpha_0 \in [0,1)$ and $\alpha_1 = 1$ is that for some $\alpha_0 \in [0,1)$,

$$\Pr\left(x=1|y=0\right)\left(\Pi\left(1,1,\alpha_0\right)-\Pi\left(0,1,\alpha_0\right)\right) \leq \Pr\left(x=0|y=0\right)\left(\Pi\left(0,0,\alpha_0\right)-\Pi\left(1,0,\alpha_0\right)\right). \tag{16}$$

With an argument analogous to the one in the proof of the previous claim, we can rewrite (16) as

$$\Pr\left(x=1|y=0\right)B\left(1,\alpha_0\right)\left(\Pi\left(1,1\right)-\Pi\left(0,1\right)\right)\leq\Pr\left(x=0|y=0\right)B\left(0,\alpha_0\right)\left(\Pi\left(0,0\right)-\Pi\left(1,0\right)\right),$$

where

$$B\left(x,\alpha_0\right)=\frac{\Pr(y=1|x)}{\Pr(y=1|x)+\alpha_0\Pr(y=0|x)},$$

which in turn holds only if

$$\Pr\left(x=1|y=0\right)\left(\Pi\left(1,1\right)-\Pi\left(0,1\right)\right)\min_{\alpha_0}\frac{B\left(1,\alpha_0\right)}{B\left(0,\alpha_0\right)}\leq\Pr\left(x=0|y=0\right)\left(\Pi\left(0,0\right)-\Pi\left(1,0\right)\right). \tag{17}$$

But

$$\min_{\alpha_0}\frac{B\left(1,\alpha_0\right)}{B\left(0,\alpha_0\right)}=\min_{\alpha_0}\frac{\Pr(y=1|x=1)}{\Pr(y=1|x=0)}\frac{\Pr(y=1|x=0)+\alpha_0\Pr(y=0|x=0)}{\Pr(y=1|x=1)+\alpha_0\Pr(y=0|x=1)}=1,$$

Then (17) rewrites as (12). If (16) holds, (12) holds, and by Claim 2 there exists an equilibrium in which $\alpha_0=0$ and $\alpha_1=1$. Claim 4 is proven. ■

Lemma 6 says that if the equilibrium set contains some kind of informative equilibrium then it must also contain a non-perverse separating equilibrium. This is a useful characterization because the existence conditions for separating equilibria are – as we shall see in the next Lemma – easily derived.

Lemma 7 *There exists a non-perverse separating equilibrium if and only if (4) holds.*

Proof. Suppose first that $y=1$ is the smart realization. From Claim 2 in the proof of the previous lemma, the necessary and sufficient condition for the existence of an equilibrium in which $\alpha_0=0$ and $\alpha_1=1$ is (12). Note that

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

and that $\Pr\left(y=1|x\right)=\gamma q_{xg}+\left(1-\gamma\right)q_{xb}$. Then, by (11), we have

$$\Pr\left(x|y=0\right)\left(\Pi\left(1,x\right)-\Pi\left(0,x\right)\right)=\frac{\Pr\left(b\right)\Pr\left(g\right)\Pr\left(x\right)\left(q_{xg}-q_{xb}\right)}{\left(\gamma q_{xg}+\left(1-\gamma\right)q_{xb}\right)\Pr\left(y=0\right)}.$$

Therefore, (12) holds if and only if

$$\frac{\Pr(x=1)(q_{1g}-q_{1b})}{\gamma q_{1g}+(1-\gamma)q_{1b}} \leq \frac{\Pr(x=0)(q_{0b}-q_{0g})}{\gamma q_{0g}+(1-\gamma)q_{0b}},$$

which is equivalent to the first inequality in (4). Also, if $y = 1$ is the smart realization, the second inequality in (4) is always satisfied.

If instead $y = 0$ is the smart realization, an analogous line of proof shows that there exists a non-perverse separating equilibrium if and only if the second inequality of (4) holds (and the first inequality is always satisfied). Hence, independently of which realization is smart, the statement of the lemma is correct. ■

Combining the three lemmas, we get the proof of Proposition 2.

Proof of Proposition 4

Suppose that the agent plays $a = y$. Let $\pi(a,x)$, $\pi(u(a,x))$, $\pi(a)$, and γ be the belief formed by the principal in the four possible information scenarios. Given a and y, the expected belief 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:

$$(1-\rho_a)\rho_u\Delta_1 \geq \rho_a(\rho_u\Delta_2+(1-\rho_u)\Delta_3). \tag{18}$$

where

$$\begin{aligned}
\Delta_1 &= E_x(\pi(u(0,x))-\pi(u(1,x))|y=0);\\
\Delta_2 &= E_x(\pi(1,x)-\pi(0,x)|y=0);\\
\Delta_3 &= \pi(a=1)-\pi(a=0).
\end{aligned}$$

It is easy to see that $\Delta_1 > 0$ and that $\Delta_3 > 0$. By (1), we can verify that

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

Hence,

$$\Delta_2 < E_y E_x \left(\pi\left(1,x\right) - \pi\left(0,x\right)\right) = \pi(a=1) - \pi(a=0) = \Delta_3.$$

We rewrite (18) as

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

The left-hand side of the inequality is decreasing in ρ_a. For any given ρ_u: if the inequality is satisfied for ρ_a, it is also satisfied for any $\rho_a' < \rho_a$. As the inequality is certainly satisfied for $\rho_a \to 0$, there exists a threshold $\rho_a^*\left(\rho_u\right) \in (0,1]$ such that the for all $\rho_a \leq \rho_a^*\left(\rho_u\right)$ the inequality is satisfied. The first part of the proposition is proven. As $\Delta_2 < \Delta_3$, the right-hand side of the inequality is decreasing in ρ_u. Therefore $\rho_a^*\left(\rho_u\right)$ is nondecreasing in ρ_u.