The Wrong Kind of Transparency

January 30, 2004

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. The results on the distinction between action and consequence are then used to interpret existing disclosure policies in politics, corporate governance, and delegated portfolio management.

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. Holmström [12] 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 (Tyle [29]).

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 (Stevensons [28, 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'."

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 (Rozell [26]). 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 (Frankel [10]).[1]

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 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 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.[2]

[1] Section 5 returns to these non-disclosure policies and re-interprets them in the context of the present model.

[2] 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.

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 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 [12] 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 (Holmström [13]), in which the principal and the agent can sign only short-term non-contingent contracts.[3]

The agency literature has already identified instances in which more information can hurt the principal. Holmström [13] 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. Dewatripont, Jewitt and Tirole [6] present examples in which the agent works harder if the principal receives a coarser signal on agent performance rather than observing performance directly. Crémer [4] 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, more information 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 risk that the agent exerts less effort, like in the papers above, but rather on the possibility that the agent disregards useful private signals. In a nutshell, we show that the availability of a certain kind of information may induce the agent to behave in a conformist way. This 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). In the following paragraphs, we provide a brief, informal description of the model

[3] As Gibbons and Murphy [11] 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.

and the main findings.

This paper employs a model of career concerns for experts.[4] What differentiates a good agent from a bad agent is his ability to understand the state of the world, which can be interpreted as expertise, intelligence, or vision. There are two periods: the current period and the future period. In the current period, an agent (the expert) is in charge of taking an action on behalf of the principal.[5] The agent has no intrinsic preferences among possible actions, i.e. there is no moral hazard in a classical sense. The agent receives a signal about the state of the world, whose precision depends on the agent's type. For now we assume that the agent does not know his own 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 belief 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) and the consequence in the second period, which in turn depends on the ability of the principal to screen agents by type (sorting).

We distinguish between two kinds of information that the principal can obtain: observing the action that the agent takes (*transparency on action*) and observing the consequence of the agent's action (*transparency on consequence*). Suppose for now that the principal always knows the consequence but may or may not observe the action (and that a consequence can be generated by more than one action-state pair, so the principal cannot deduce the action from the consequence).

For example, in delegated portfolio management the relevant state of the world is a vector of future asset prices. The agent (the fund manager) receives a private signal on changes in asset prices and selects a portfolio of assets on behalf of the investor. A good

[4] See Scharfstein and Stein [27], Zwiebel [30], Prendergast and Stole [25], Campbell [3], Ottaviani and Sørensen [20] [21], Levy [17], Ely and Välimäki [8], and Ely, Fudenberg, and Levine [7].

[5] The formalization we use can be interpreted as a reduced form of either a model with only one principal or a model with a market of principals. For now, we adopt the first interpretation.

fund manager differs from a bad one in his ability to make correct predictions on future asset prices, through which he can generate higher returns for his investor. In the second period, the investor is more likely to retain the fund manager if the belief on his predictive ability is high.[6] In the first period, the fund manager selects the portfolio in order to show his predictive ability. If the investor is risk neutral, the distinction between action and consequence is straightforward. The action is the portfolio that the fund manager selects; the consequence is the return on the portfolio. Suppose for now that the investor always observes the return (which is available on newspapers for major funds). The question we ask is: should the investor also observe the composition of the fund she owns?

A first result is that more information about the action can hurt the principal. To understand this, 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 because knowing which particular action-state pair has generated the observed consequence helps the principal understand the agent's type. 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, but clearly an agent who realizes that his action is going to be observed faces a different incentive structure. A key observation is that, in a generic model, the possible realizations of the agent's signal can be ranked in order of *smartness*, that is, according to the posterior probability that the agent's type is high given the realization of the signal. Good agents are more likely than bad agents to receive smart signals. If in equilibrium the agent's action is informative of his signal, then also all the possible actions can be ranked in order of smartness. The belief on the agent's type depends on the consequence but also on the smartness of the action. This can create a contradiction. If the smartness component is too strong, the only possible equilibrium is one in which actions cannot be ranked in order of smartness, 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.

[6]Delegated portfolio management in the US fits well with the career concern setup because the Investment Advisers Act of 1940 prevents mutual fund managers from receiving additional payments contingent on good returns (Das and Sundaram [5]).

To make this line of reasoning more concrete, let us return to the delegated portfolio management example. Suppose that the fund manager has two possible investment strategies: a portfolio oriented toward blue chips and one which is heavy on high-tech smaller firms.[7] There are two states of the world, according to whether blue chips or high techs will yield better returns. The agent receives a private signal on asset prices with two possible realizations, "blue chips" or "high techs," which tells him which portfolio choice maximizes expected return. Assume that, in the terminology introduced above, the high-tech realization is the smart one. This means that the fund manager's ability is more important in understanding when to invest in high-tech stocks (perhaps because prospects of small technology firms are more difficult to evaluate). If the investor could observe the fund manager's signal directly, she would form a higher belief when the realization is "high tech" rather than "blue chip". If the portfolio composition is observed and if the fund manager acts according to the realization of his signal, for any possible return the investor forms a higher belief if the investor chooses high techs rather than blue chips. But then, the fund manager may have an incentive to choose high techs even when her signal suggests blue chips. In other words, transparency on portfolio induces the fund manager to behave not according to his private signal but according to the investor's prior on how an able fund manager is likely to behave. If this is the case, the only equilibrium is for the fund manager to always pick the same portfolio – a very negative situation for the investor who gets an uninformed choice in the first period and is not able to sort fund managers based on ability. Under these circumstances, the investor should commit not to observe the portfolio composition, in which case the fund manager follows his private signal in making the investment decision. In Section 5 we will relate this theoretical result to evidence suggesting that mutual funds (in which investors only observe returns and bi-annual snapshots) may outperform pension funds (in which investors have access to

[7]With only two assets, an investor who observes the portfolio return and stock prices can deduce the portfolio composition. So, if we want to keep open the possibility that the investor does not learn portfolio choices, we need to assume that she does not observe stock prices. With at least three assets, this kind of deduction is in general no longer possible. In practice, fund managers are able to construct portfolios by combining hundreds of stocks, and the presence of an inference problem becomes unrealistic.

portfolio composition information).

The core result of the paper is a necessary and sufficient condition under which revealing the agent's action leads to conformism. The condition has to do with the relative smartness of the realizations of the agent's signal. If one realization is much more smart than the others, then the chain of negative effects described 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 action for sorting purposes.

We also show that there is complementarity between transparency on action and transparency on consequence. The optimal probability that action is observed is nondecreasing in the probability that the consequence is observed. This is because an agent who pretends to have observed the smart realization by playing the action corresponding to the smart realization has a lower probability of obtaining a good consequence than an agent who actually observed the smart realization. Thus the cost of pretending to have observed the smart realization is increasing in the probability that consequence is observed.

The present work is particularly related to two papers on experts. Prendergast [24] 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. While the setup is entirely different, the present work shares Prendergast's insight that when the principal attempts to gather information on the agent's signal the agent may have an incentive to distort his signal report. The two works are complementary in that Prendergast focuses on comparing compensation schemes while we are interested

in comparing information structures. Avery and Meyer [1] 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. Although the setup is quite different, the intuition bears a connection to the present paper. If the advisor knows that his recommendations affect his future career prospects, he may have an incentive to pool on one type of recommendation independently of his private information.

The plan of the paper is as follows. Section 2 introduces the career concern game. Section 3 begins with a simple example in which transparency on action is detrimental. 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 by using the results of the paper to interpret some existing institutional arrangements.

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 agent does not know his type). 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.[8]

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

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

[8] A more general version of the model, in which variables are not binary, is available in Prat [22].

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:

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

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 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.[9]

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 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 [21] 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 [21] 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.[10]

Finally, we introduce a notion that corresponds to a mental experiment. Suppose

[9] Prat [22] proves that this model is indeed the reduced form of two "long form" models: retrospective voting and labor market. In particular, it is shown that the agent's payoff is linear in the belief and the principal's expected payoff is convex in the belief.

[10] Prat [22] 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.

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. The following result – which is easy to check – relates smartness to the primitives:

Proposition 1 *The smart realization is $y = 1$ if and only if*

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

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 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 agent's belief above γ while observing $y = 0$ decreases it.

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 $x = 0$.

This setup may be used to represent the delegated portfolio management example discussed in the introduction. The state $x = 0$ corresponds to the "boring" situation in which blue chips are the optimal investment strategy, while $x = 1$ is the "exciting" scenario in which high tech smaller firms do better. To make the point as clear as possible, everything is symmetric except the agent signal. The key assumption is that the ability of the agent is state-specific. In the boring blue-chip case both types of agents have the same precision while in the exciting high-tech case the good agent performs better.

The smart realization is $y = 1$ because $\Pr(\theta = g|y = 0) = \frac{1}{3}$ and $\Pr(\theta = g|y = 1) = \frac{3}{5}$. The signal $y = 0$ is bad news for the ability of the agent, and the agent will try to conceal this information from the principal. As we shall see shortly, this leads to conformism: the agent has an incentive to act as if he had observed the smart realization $y = 1$ even when he observes $y = 0$.

We now argue that in this example transparency on action induces complete conformism and it damages the principal.[11] 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:

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

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

A similar non-existence argument applies to the 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 $a = 1$). 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

[11] 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.

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 $\tilde{\pi}(u)$ on 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 begin the analysis of the game introduced in Section 2 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), $\tilde{\pi}(u = 1) > \gamma > \tilde{\pi}(u = 0)$. As the signal y is by assumption decision-relevant, it is a best response for teh agent to play $a = y$. Therefore,

Proposition 2 *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. Still, we can show.

Proposition 3 *There exists a non-perverse separating equilibrium if and only if*

$$\frac{p}{1-p}\frac{\gamma q_{0g}+(1-\gamma)\,q_{0b}}{\gamma q_{1g}+(1-\gamma)\,q_{1b}} \leq \frac{q_{0b}-q_{0g}}{q_{1g}-q_{1b}} \leq \frac{p}{1-p}\frac{\gamma\,(1-q_{0g})+(1-\gamma)\,(1-q_{0b})}{\gamma\,(1-q_{1g})+(1-\gamma)\,(1-q_{1b})}. \tag{3}$$

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

Proof. See Appendix. ■

Proposition 3 is arrived at in two steps. We first prove that if there exists a semi-separating equilibrium, there must also exist a non-perverse separating equilibrium. Then, we identify the necessary and sufficient conditions for the existence of a non-perverse separating equilibrium.

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 (3) with the smartness condition of Proposition 1. 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 3 says whether one realization is *much* better news than the other.

Suppose for instance that $y=1$ is the smart signal. We can disregard the second inequality in Proposition 3 because it is implied by the smartness condition of Proposition 1. Instead, the first inequality may or may not hold. If it fails, there is no informative equilibrium because $y=1$ is "too smart" 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 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 (3) 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 2, the best equilibrium with concealed action is a separating equilibrium with $a = y$. What happens with revealed action depends on condition (3). 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 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 (3) 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 informative equilibrium the posterior belief is equal to the prior. Thus, the principal is worse off. We summarize the argument as follows:

Proposition 4 *If (3) holds, revealing the agent's action does not affect discipline and improves sorting. If (3) fails, revealing the agent's action worsens both discipline and sorting. Hence, the principal prefers to reveal the action if and only if (3) 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 (3) 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.[12]

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 $(\rho_u = 1, \rho_a = 0)$.

[12] This tension can be captured in a formal result linking the ex ante incentive to commit with the ex post benefit of reneging on the commitment (see Prat [22]).

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.[13]

Proposition 5 *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. ■

Proposition 5 has two parts. First, given a probability that the consequence is observed, ρ_u, there exists a threshold $\rho_a^*(\rho_u)$ such that there exists a separating equilibrium if and only if the probability of observing the action is below the threshold. Second, the threshold is nondecreasing in ρ_u.

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

5 Conclusion

Are the theoretical results obtained in this paper useful for understanding existing institutional arrangements? Let us re-consider one by one the examples of non-transparent institutions that we listed in the introduction.

[13] This section has a more limited scope than the previous one because we focus attention on pooling and separating equilibria. In particular, 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 5 to draw normative conclusions on sorting (but we can use it for discipline).

In delegated portfolio management, there have been proposals to increase the frequency with which mutual funds are required to disclose their portfolio composition, which in the US is now six months. The Investment Company Institute (the fund managers' association) [29] argues that an increased frequency risks hurting investors because "[it] would focus undue attention on individual portfolio securities and could encourage a short-term investment perspective." The Institute also argues that there does not seem to be much demand by investors for more information on portfolio holdings. It is easy to use the framework developed here to back up the Institute's argument. The action of a fund manager is his investment strategy. The consequence is return to investors. Returns are observable but also volatile. In the long term they are a reliable signal of the fund manager's quality but in the short term they contain a lot of variance. If the action is observable in the short term, there is a risk that fund managers will behave in a conformist way, ignoring their private investment and following the strategy that is a priori a better signal of their competence.[14]

There is an interesting link between this paper and Lakonishok et al. [14]. 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. The present paper makes this possibility more precise. 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 present model predicts that this may create an incentive for conformism in pension fund managers, which decreases their expected return.

Moving on to corporate governance, shareholders receive information about the man-

[14]Clearly, such a risk must be traded off against a more traditional moral hazard problem caused by the availability of rent extraction opportunities such as late trading.

agement of their firm from the accounting reports that the firm prepares. Clearly, accounting involves a great deal of aggregation both across time and across areas. 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 as it is now the case in the US (Leuz and Verrecchia [16]). 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, which according to the present model can create adverse consequences.

In politics, the importance of career concerns is widely recognized.[15] For our purpose, it is interesting to not that 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 has been articulated in several contexts. 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 [2, p. 3]).

More precise implications can be extracted from Proposition 5. 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 [10] reports that all the 30-plus countries that have adopted an open government code allow for some form of short-term

[15] See Maskin and Tirole [18] for a general analysis of career concerns in public decision-making.

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 decision are released only when voters begin to have a chance to form an opinion on the consequence of the decision in question.[16]

It is important to stress that this paper also identifies circumstances in which information revelation *is* the optimal policy. In particular, supplying information on the consequence of the agent's action is unambiguously good for the principal. This is true in a direct sense, because it improves discipline and sorting, and – as the section on complementarities showed – in an indirect sense, because it allows for more information on action, which in turn improves sorting. Most of the recent corporate scandals involved distorted profit reporting. As profit is a consequence, nothing in the present paper lends support to accounting policy choices, such as the expensing of CEO options, that can lead to a less precise measure of firm profits.

We conclude by pointing to three possible extensions. First, there are two ways of modeling cheap talk with career concerns: the expert model which is used here and the biased advisor model of Morris [19]. It would be interesting to know to what extent the results presented here carry over to the biased advisor model. Second, this paper does not allow for an asymmetric disclosure policy. One action could be revealed with a higher probability than another action. Could the principal benefits from the introduction of such asymmetric information structures?[17] Finally, in many agency relationships information is generated endogenously by the players. Lobbies and media gather and distribute intelligence on government policy. Shareholders and financial analysts question company

[16] 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])

[17] Leaver [15] considers a career concern model in which the agent's actions may be observed with different precision.

management. What are the incentives of monitoring agencies and how do they reflect on transparency policy?[18]

References

[1] Christopher Avery and Margaret M. Meyer. Designing hiring and propotion procedures when evaluators are biased. Working paper, 1999.

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

[3] Colin Campbell. Learning and the market for information. Mimeo, Ohio State University, 1998.

[4] Jacques Crémer. Arm's length relationships. *Quarterly Journal of Economics* 110(2): 275–295, 1995.

[5] Sanjiv R. Das and Rangarajan K. Sundaram. On the regulation of fee structures in mutual funds. *Mathematical Modelling in Finance Vol III*. Courant Institute of Mathematical, forthcoming.

[6] 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.

[7] Jeffrey Ely, Drew Fudenberg, and David K. Levine. When is reputation bad? Mimeo, 2002.

[8] Jeffrey Ely and Juuso Välimäki. Bad reputation. *Quarterly Journal of Economics*, 118(3): 785–813, 2003.

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

[18] Prendergast [23] considers the role of consumer monitoring, and the possible biases that it may induce. If consumers' interests are distant from the principal's goals, the principal may want to restrict the customers' ability to complain about the agent's performance.

[10] Maurice Frankel. Freedom of information: Some international characteristics. Working paper, The Campaign for Freedom of Information, 2001.

[11] Robert Gibbons and Kevin J. Murphy. Optimal incentive contracts in the presence of career concerns: Theory and evidence. *Journal of Political Economy* 100(3): 468–505, 1992.

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

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

[14] Josef Lakonishok, Andrei Shleifer, and Robert W. Vishny. The structure and performance of the money management industry. *Brookings Papers on Economic Activity* Vol 1992: 339–379, 1992.

[15] Clare Leaver. Bureaucratic minimal squawk: Theory and evidence. Working paper, University College London, February 2002.

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

[17] Gilat Levy. Strategic consultation in the presence of career concerns. STICERD Discussion Paper TE/00/404, London School of Economics, 2000.

[18] Eric Maskin and Jean Tirole. The politician and the judge: Accountability in government. Working paper, Toulouse University, 2001.

[19] Stephen Morris. Political correctness. *Journal of Political Economy* 109: 231–265, 2001.

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

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

[22] Andrea Prat. The wrong kind of transparency. CEPR Discussion Paper 3859, 2003.

[23] Canice Prendergast. Consumers and agency problems. NBER Working Paper w8445, 2001.

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

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

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

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

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

[29] Craig S. Tyle. Letter to the SEC on the frequency of mutual fund portfolio holdings disclosure. Investment Company Institute, 2001.

[30] Jeffrey Zwiebel. Corporate conservatism and relative compensation. *Journal of Political Economy* 103(1): 1–25: 1995.

Appendix: Proofs

Proof of Proposition 3

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

Lemma 6 *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 :

$$\begin{aligned} \Pr(x=0|y=1)(\pi(0,0)-\pi(1,0)) &= \Pr(x=1|y=1)(\pi(1,1)-\pi(0,1)), \quad (4) \\ \Pr(x=0|y=0)(\pi(0,0)-\pi(1,0)) &= \Pr(x=1|y=0)(\pi(1,1)-\pi(0,1)). \quad (5) \end{aligned}$$

There are two cases:

$$\begin{aligned} (\pi(0,0)-\pi(1,0))(\pi(1,1)-\pi(0,1)) &\leq 0 \quad (6) \\ (\pi(0,0)-\pi(1,0))(\pi(1,1)-\pi(0,1)) &> 0 \quad (7) \end{aligned}$$

If (6) 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 (4) have different signs. If (7) holds, subtract (5) from (4)

$$\begin{aligned} &(\Pr(x=0|y=1)-\Pr(x=0|y=0))(\pi(0,0)-\pi(1,0)) \quad (8) \\ = \; &(\Pr(x=1|y=1)-\Pr(x=1|y=0))(\pi(1,1)-\pi(0,1)). \end{aligned}$$

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

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

Then, (7) creates a contradiction in (8). ■

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 7 *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

$$\begin{aligned}&\Pr(x = 1)\Pr(y = 1|x = 1)\Pr(y = 0|x = 1)(\Pi(1, 1) - \Pi(0, 1)) \qquad (9)\\ >\ &\Pr(x = 0)\Pr(y = 1|x = 0)\Pr(y = 0|x = 0)(\Pi(0, 0) - \Pi(1, 0)).\end{aligned}$$

Proof of Claim 1: Note that

$$\begin{aligned}
& \Pr(x)\Pr(y=1|x)\Pr(y=0|x)\left(\Pi(1,x)-\Pi(0,x)\right) \qquad (10)\\
= \;& \frac{1}{\Pr(x)}\left(\Pr(g,y=1,x)\Pr(y=0,x)-\Pr(g,y=0,x)\Pr(y=1,x)\right)\\
= \;& \frac{1}{\Pr(x)}\left(\Pr(g,y=1,x)\left(\Pr(g,y=0,x)+\Pr(b,y=0,x)\right)\right)\\
& -\frac{1}{\Pr(x)}\left(\Pr(g,y=0,x)\left(\Pr(g,y=1,x)+\Pr(b,y=1,x)\right)\right)\\
= \;& \frac{1}{\Pr(x)}\left(\Pr(g,y=1,x)\Pr(b,y=0,x)-\Pr(g,y=0,x)\Pr(b,y=1,x)\right)\\
= \;& \Pr(b)\Pr(g)\Pr(x)\left(q_{xg}\left(1-q_{xb}\right)-\left(1-q_{xg}\right)q_{xb}\right)\\
= \;& \Pr(b)\Pr(g)\Pr(x)\left(q_{xg}-q_{xb}\right).
\end{aligned}$$

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

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

which, by the argument above, is equivalent to (9). Claim 1 is proven.

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

$$\Pr(x=1|y=0)\left(\Pi(1,1)-\Pi(0,1)\right) \leq \Pr(x=0|y=0)\left(\Pi(0,0)-\Pi(1,0)\right) \qquad (11)$$

$$\Pr(x=1|y=1)\left(\Pi(1,1)-\Pi(0,1)\right) \geq \Pr(x=0|y=1)\left(\Pi(0,0)-\Pi(1,0)\right) \qquad (12)$$

Claim 2: The inequality (12) is always satisfied. There exists a separating equilibrium if and only if (11) 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,

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

But (1) implies that $\Pr(y=0|x=0) > \Pr(y=0|x=1)$. Therefore (12) holds a fortiori. Claim 2 is proven.

From Lemma 6, 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

$$\Pr(x=1|y=1)\left(\Pi(1,1,\alpha_1) - \Pi(0,1,\alpha_1)\right) = \Pr(x=0|y=1)\left(\Pi(0,0,\alpha_1) - \Pi(1,0,\alpha_1)\right). \tag{13}$$

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 (13) rewrites as

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

which in turn holds only if

$$\begin{aligned} & \Pr(x=1|y=1)\left(\Pi(1,1)-\Pi(0,1)\right) \\ \leq\ & \Pr(x=0|y=1)\max_{\alpha_1\in[0,1]}\frac{A(0,\alpha_1)}{A(1,\alpha_1)}\left(\Pi(0,0)-\Pi(1,0)\right). \end{aligned} \tag{14}$$

Note that

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

Inequality (14) can thus be rewritten as

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

or

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

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 $\alpha_0 \in [0,1)$,

$$\Pr(x=1|y=0)\left(\Pi(1,1,\alpha_0)-\Pi(0,1,\alpha_0)\right) \leq \Pr(x=0|y=0)\left(\Pi(0,0,\alpha_0)-\Pi(1,0,\alpha_0)\right). \tag{15}$$

With an argument analogous to the one in the proof of the previous claim, we can rewrite (15) as

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

where

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

which in turn holds only if

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

But

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

Then (16) rewrites as (11). If (15) holds, (11) holds, and by Claim 2 there exists an equilibrium in which $\alpha_0=0$ and $\alpha_1=1$. Claim 4 is proven. ■

Lemma 7 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 8 *There exists a non-perverse separating equilibrium if and only if (3) 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 (11). Note that

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

and that $\Pr(y=1|x)=\gamma q_{xg}+(1-\gamma)q_{xb}$. Then, by (10), we have

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

Therefore, (11) 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 (3). Also, if $y=1$ is the smart realization, the second inequality in (3) 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 (3) 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 3.

Proof of Proposition 5

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:

$$(1 - \rho_a)\rho_u \Delta_1 \geq \rho_a(\rho_u \Delta_2 + (1 - \rho_u)\Delta_3). \tag{17}$$

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,

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

We rewrite (17) as

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

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 $\rho_a^*(\rho_u) \in (0,1]$ such that the for all $\rho_a \leq \rho_a^*(\rho_u)$ 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 $\rho_a^*(\rho_u)$ is nondecreasing in ρ_u.