Disclosing Decision Makers’ Private Interests

Antoni-Italo de Moragas †

This version: February 1, 2017
Most updated version here

Abstract

In this paper I study the effect of disclosing the private interests of decision makers on the quality of the decisions that are eventually taken. I focus on a delegation relationship where decision makers motivated by career concerns try to build up their own reputation. When private interests of decision makers are not disclosed, taking the correct decision is the only way to increase reputation and the higher the career concerns the more likely it is that correct decisions are taken. When private interests are disclosed, decisions not aligned with these private interests may also increase reputation. I find that, contrary to the common wisdom, disclosure of private interests can induce worse decisions. This happens when the salience of career concerns is high enough and decision makers are poorly informed.

JEL-Classification: C72; D72; D82.

Key-words: Disclosure; Transparency; Delegation; Conflict of Interest.

*I am grateful to my supervisors Andrea Mattozzi and David K. Levine for their help and support. I would also like to thank Andres Carvajal and Moti Michaeli for many valuable suggestions as well as the audience of the 2016 Warwick PhD Conference in Economics, the 2016 Applied Economics Workshop in Petralia and the EUI Microeconomics Working Group.

†Department of Economics, European University Institute. Email: antoni.moragas@eui.eu.
1 Introduction

“There is strong shadow where there is much light”
—Johann Wolfgang von Goethe, Götz von Berlichingen, Act I (1773)

When a delegation relationship risks having conflict of interests, disclosure of the conflict of interests is one of the remedies prescribed most often. Advocates of disclosure claim (Stark, 2003) that, by being aware of potential private interests, the public can form its own judgement on whether decision makers put their own personal benefits ahead of their duties. Such knowledge, they claim, will force decision makers to behave according to the public interest instead of their own. Although disclosure, like other accountability practices, may not solve the conflict of interest, it is broadly perceived as a good solution without the negative effects of other remedies like recusal (excluding the interested decision maker from the decision) or divestiture (removing the private interest from the decision maker). This belief explains the popularity of disclosure in different areas.

For example, in politics, the growing distrust in political power has generated a great demand for more financial and personal information about public officials (Cain, 2014). Such demand has not always translated into real changes in practice and the level of transparency and public available information still differs within countries and branches of the government. For example, (Djankov et al., 2010) analyzed the rules and practice of disclosure of MPs in 175 countries and found that despite 109 of the 175 countries of their sample had disclosure laws, only 63 of them granted public access to the disclosed information. International organizations such as Transparency International, the World Bank or the OECD stand for extending disclosure and public availability to more countries.

In academic research, the ties between academic scientists and private industry have also come under scrutiny. Many universities and peer-reviewed journals have recently adopted
disclosure policies that require their employees and authors to declare their financial interests. These policies are taken very seriously and might have severe consequences. The disclosure policy of the European University Institute (EUI) was behind the resignation of its president Josep Borrell in 2011. His failure to declare his revenues of 300,000 euros a year as a board member of the Spanish energy company Abengoa made his position as president of the EUI unsustainable.

Conflicts of interest are also common in medical practice. Despite the Hippocratic oath by which physicians promise to always act in the patient’s best interest, doctors have long faced the accusation that the gifts they receive from pharmaceutical companies influence their drugs’ prescription practices. This accusation is based on reasonable grounds. De-Jong et al. (2016) found that doctors who had enjoyed industry-sponsored meals had higher rates of prescriptions of the company’s brand-name medication. To address these concerns, the American Senate introduced the Physician Payments Sunshine Act (PPSA) in 2010, requiring medical product manufacturers to disclose any payment made to physicians.

Despite the popularity of disclosure rules to solve conflict of interests, their effects have rarely been studied. In this paper I examine the effect of disclosing the agents’ private interests on the quality of the decisions that are eventually taken. The main contribution of this paper is to identify under which conditions decision makers take more correct decisions when their private interests are revealed.

To do so, I examine a model in which an agent receives a private signal about the state of the world and has to take a decision. The correctness of this decision depends on such state of the world. All agents have private interests that can affect their preferred decision but not all of them care the same way about them. Specifically, there are two types of agents: good and bad. Good agents prefer taking the correct decision independently of their private interests. Bad agents preferred decision depends on their private interests independently of
the correctness of the decision. At the same time, agents have career concerns and they want to convince an external evaluator that they are good types because this can create a reputation that can lead to higher wages, promotions or reappointments in the future. The assessment of the evaluator depends on the information he has access to. Without disclosure of the agent’s private interests, the evaluator can only observe the state of the world and the decision taken. With disclosure, in addition to this, the evaluator also observes the private interest of the agent.

The effect of disclosure is ambiguous because it modifies the reputational incentives of bad types but it also distorts the incentives of the good ones. When their private interests are not disclosed, taking correct decisions is the only way for agents to increase their reputation. With disclosure, taking decisions against the agent’s private interests also increases reputation. I derive the conditions for the existence of three different equilibria: an equilibrium such that disclosure has no effect, an equilibrium such that disclosure leads decision makers to take correct decision more often and an equilibrium such that disclosure decreases the probability of taking correct decisions.

The existence of this last equilibrium shows that disclosure of decision makers’ private interests can lead to what Gersen and Stephenson (2014) calls over-accountability. That is, a situation where accountability decreases rather than increases an agent’s likelihood of taking correct decisions. Disclosure should therefore be used wisely. Specifically, when the agent’s career concerns are expected to be high in relation with the importance of the decision, disclosure can be negative because it induces decision makers to take decisions against their private interest independently of the signal they receive. This would be a justification for disclosing only the private interests that exceed a monetary threshold and keep less salient private interest secret.

Finally, I extend the model to a two-period political agency model to see if disclosure
can help the voters to screen politicians in the second period. More precisely, I assume that after the decision is taken in period one, voters can keep the incumbent politician or replace him with another, randomly drawn, challenger. In period 2 voters prefer to have a good politician because they are not subject to reelection and the incumbent politician always take his preferred decision. I show that disclosure of the private interests not only can decrease the utility of voters in the first period but it can also decrease the probability of selecting good politicians in period 2, which decreases voters’ utility in the second period too. The reason is that, when career concerns are high enough, good and bad types take decisions against their private interest and citizens cannot screen them.

The rest of the paper is organized as follows. In the next section, I discuss how this paper is related to the existing literature. Section 3 presents the model. The main results are in Section 4 and 5. Section 6 analyzes a political agency model, Section 7 discusses some of the assumptions of the model. I conclude in Section 8. All formal proofs can be found in the appendix.

2 Related Literature

This paper is linked to three different literatures. First, there is now a sizable literature on the effects of transparency of the decisions taken by careerist decision makers\footnote{In Prat (2005), Visser and Swank (2007), Levy (2007), Meade and Stasavage (2008), Gersbach and Hahn (2008), Swank et al. (2008), Fox and Van Weelden (2012) and Swank and Visser (2013) decision makers are heterogeneous in the precision of their signals and they want to signal that their information is precise. On the contrary, in Stasavage (2007) and Fox (2007) decision makers differ in their preferences and they want to signal to the voters that they share their same preferences. Finally in Mattozzi and Nakaguma (2016), decision makers differ both in the precision of the information and on their preferences. In all these models, transparency can be negative because if agents care about their reputation and their action is observed, they may choose the action that makes them appear smarter which is not always the best action.} From this literature, the paper most closely related to mine is the one by Fox (2007). In his political agency model, a lawmaker has to choose a policy and a representative voter is uncertain about
the lawmaker’s preferences. In particular, the voter does not know whether the lawmaker is unbiased and his preferences are aligned with voters’ preferences or biased and with different preferences. When the policy choice is not observed by the voter, lawmakers choose their preferred policy. When the choice is observed, lawmakers select policies that lead the voter to believe that they are unbiased.

The main difference with [Fox (2007)] is the information that is disclosed. In my model, the information disclosed is not the action taken by the decision maker, but the direction of his bias which, in [Fox (2007)] is always disclosed. This difference sheds light on the mechanism behind [Fox (2007)] finding. In particular, I show that the negative effects of transparency in [Fox (2007)] depend critically on the disclosure of the private interest. In particular, I show that when private interest is known, transparency of the action can decrease the probability of a correct decision, a finding that is in line with [Fox (2007)] and [Stasavage (2007)]. However, when the private interest is unknown, transparency of the action can only lead to more correct decisions.

Second, my model also relates to the literature on the disclosure of biases in expert advice settings. In particular, [Li and Madarász (2008)] extend the cheap-talk model of [Crawford and Sobel (1982)] and study how the results change when the receiver is informed about the bias of the sender. In their model, disclosure can hurt the receiver because it can reduce the quality of the communication. The intuition is that when receiver does not know the direction of the bias of the sender, he does not know whether the sender had incentives to overstate or to understate the state of the world and, if his preferences are concave, it is optimal for the receiver to follow the advice of the sender. This induces the sender to communicate always his preferred decision and, if the preferences of the sender and the receiver are sufficiently close, the receiver will benefit from this communication. On the

---

contrary, when the receiver knows the bias of the sender, the sender has to pool messages which reduces communication and hurts the sender and, in some circumstances, also the receiver.

The difference between this paper and Li and Madarász (2008) is that in the model presented here, the biased agent takes a pay-off relevant action (delegation) and in Li and Madarász (2008), the agent’s utility depends exclusively on how the receiver interprets his message (advice). The logic of the bad equilibrium is therefore qualitatively different and, more important, the situations that can be captured by both models differ too. For instance, expert advice models are not adequate to capture the effect of disclosure of private interests in politics because politicians do not advise but take take decisions on our behalf. Given the attention that disclosure in politics has attracted, we devote an extension of the model to carefully characterize the effects of disclosure in a simple political agency model.

Finally, the paper is also related to the pandering literature that studies under which conditions agent’s career concerns leads to worse decisions. I show that the bad effects of career concerns are only present when the private interests of the agent are observed by the evaluator. When they are not, higher career concerns always increase the probability of taking correct decisions.

3 The Model

There is an agent $D$ who has to take a decision $d$. For simplicity we will assume that there are only two possible decisions $d \in \{a, b\}$. In politics, this can represent a politician choosing between contracting firm $a$ or firm $b$ for a public project or, in medical practice it

---

3This is shown in Morris (2001), Ely and Välimäki (2003), Maskin and Tirole (2004), Ely et al. (2008), Acemoglu et al. (2013) and Che et al. (2013).

In section 6 we will study what happens when the number of actions is $n > 2$. 

---
can be a doctor prescribing a drug from two different pharmaceutical companies. Before the
decision is taken, the agent observes a private signal $s \in \{a, b\}$ about the state of the world
$w \in \{a, b\}$. The state of the world is unknown but it is common knowledge that both states
can occur with equal probability. The signal is only partially informative of the realized
state of the world. In particular, we will assume that $q = Pr(s = w|w) > \frac{1}{2}$ and we will refer
to $q$ as the precision of the signal.

The correctness of the decision always depends on the state of the world. Specifically, a
decision is correct when it matches the state of the world ($d = w$). The agent always has
a private interests $\beta \in \{a, b\}$ towards one of the alternatives. You can think that when $D$
is linked with firm $\beta$ and, consequently, he can benefit from contracting it. It is common
knowledge that both private interests are equally likely and uncorrelated with the state of
the world $w$. Even if all decision makers have some private interest, not all of them care the
same way about it. In particular we will assume that a fraction $\mu \in (0, 1)$ of decision makers
care about taking correct decisions regardless of their private interests and the remaining
$1 - \mu$ care about taking the decision that coincides with their private interest regardless of
the state of the world.\footnote{We could relax the assumption such that both types care about taking the decisions that coincide with their private interest but they differ in the intensity of the preferences.} We will refer to the former as good decision makers ($\theta = 1$) and to
the latter as bad decision makers ($\theta = 0$).\footnote{Notice that the type of the decision maker is defined by the couple ($\beta, \theta$), however for exposition convenience we will use exclusively $\theta$ as a type.}

In addition to the decision maker $D$, there is an evaluator $E$. As in other career concerns
models\footnote{See Levy (2004), Levy (2007) or Mattozzi and Nakaguma (2016).} the evaluator does not have any utility function and his task is simply to update
his beliefs about the type of the decision maker rationally (i.e. using Bayes rule whenever
possible). If $E$ believes that $D$ is a good decision maker, that is a decision maker that cares
about taking correct decisions, $D$ can be rewarded. In politics we can think about $E$ as the
electorate, in medical care we can interpret him as a patient that can choose to stay with the same doctor or convincing his relatives to consult him in the future. We will refer to these beliefs as the reputation $R$ of the decision maker.

The utility of a decision maker of type $\theta$ with private interest $\beta$, when taking a decision $D$ in a state of the world $w$ is given by:

$$U_\theta(\beta, w, d) = \theta 1_{d=w} + (1 - \theta) 1_{d=\beta} + \phi R$$

where $\phi \in \mathbb{R}_+$ measures the importance of reputation $R$.

We will study two different institutional settings: one with disclosure of the private interest and the other without. The only difference between these institutional settings is precisely the information available to the evaluator when he updates his beliefs about the type of the decision maker. In particular, with opacity, the evaluator only observes the decision taken and the state of the world. Thus, his beliefs are $R(w, d) = E[\theta | w, d]$. With disclosure, in addition to this, the evaluator also observes the private interest of the decision maker and $R(w, \beta, d) = E[\theta | w, \beta, d]$.

The strategy of $D$ is to choose $d$ conditional on his type $\theta$, his private interest $\beta$ and the signal $s$, thus it is a mapping $\alpha_\theta(\beta, s) \to \{a, b\}$. The following pure strategies will be useful to characterize the equilibria. We will say that a decision maker follows his signal if $\alpha_\theta(\beta, s) = s$ and he contradicts his signal if $\alpha_\theta(\beta, s) \neq s$. Analogously, a decision maker follows his private interest if $\alpha_\theta(\beta, s) = \beta$ and contradicts it if $\alpha_\theta(\beta, s) \neq \beta$.

To summarize, the timing is as follows:

1. The state of the world $w$ and the private interest $\beta$ and the type $\theta$ of the agent are realized.
2. The agent observes the signal \( s \) and takes a decision \( d \).

3. Without disclosure, the evaluator observes \( d \) and \( w \). With disclosure, the evaluator observes \( d \), \( w \) and \( \beta \). The evaluator forms a posterior \( R \) on the agent’s type.

4. Payoffs are realized.

We will solve the model using the concept of Perfect Bayesian Equilibrium. I will only consider equilibria such that good types play pure strategies \(^8\). Moreover, when multiple equilibria exist I will focus on equilibria that maximize the probability of taking correct decisions.

4 No Disclosure

Before analysing the behaviour of decision makers with disclosure we need to study what happens without it as a benchmark. Without disclosure, the reputation of decision makers cannot depend on their private interest \( \beta \) because the evaluator does not observe it. Thus, the beliefs of the evaluator given the observation \((d, w)\) and his conjecture of \( \alpha \) are:

\[
R(w, d) = \frac{\mu \sum_{\beta', s'} Pr(s = s'|w)Pr(\alpha(1, \beta', s') = d)}{\mu \sum_{\beta', s'} Pr(s = s'|w)Pr(\alpha(1, \beta', s') = d) + (1 - \mu) \sum_{\beta', s'} Pr(s = s'|w)Pr(\alpha(0, \beta', s') = d)}
\]

(2)

Now, given that both states and both private interests are equally likely and uncorrelated it is straight-forward to see that in equilibrium good types always follow the signal\(^9\).

\(^8\)This rules out equilibria that are not attractive. If good types played non-degenerate mixed strategies in equilibrium, they could get higher reputation and higher utility from any deviation if the evaluator updated his beliefs. Notice that this is not the case for bad types.

\(^9\)Clearly there are other equilibria. In these equilibria either good types always take the same decision independently of the signal and the private interest or they contradict the signal they receive. It is straight-forward to check that all these equilibria reduce the utility with respect to the equilibrium we analyse.
**Proposition 1.** For all \((q, \mu)\), in equilibrium: (i) When \(\phi \leq \bar{\phi}_S(q, \mu)\), good decision makers follow their signal and bad decision makers follow their private interest. (ii) When \(\phi > \bar{\phi}_S(q, \mu)\) good decision makers follow their signal and bad decision makers mix between following the signal and the private interest.

Without disclosure, good types follow their signal and maximize the likelihood of taking the correct decision. This happens because the reputation incentives are aligned with their preferred decision. This is not the case for bad types. When they receive a signal opposed to their private interest, bad types experience a trade-off between their reputation (following the signal) and their present utility (following the private interest). Unsurprisingly, when the career concerns are low enough, they follow their private interest but when they are higher, they mix between following the private interest and following their signal.

**Corollary 2.** When the private interests of the agent are not disclosed, the probability of a correct decision is always increasing in the career concerns of the agent.

As the previous corollary shows, higher career concerns always increase the probability of correct decisions. Given that good types always follow the signal, the results are driven by the behaviour of bad types. In particular, when bad types are mixing the probability of following the signal increases when the reputational concerns increase.\(^{[10]}\) The reason why higher career concerns increase the probability of correct decisions is that, without disclosure, the only way to increase reputation is by taking correct decisions. As we will see in the next section, this is not going to be the case with disclosure.

\(^{[10]}\) Notice, however, that no matter how much decision makers care about reputation, bad types, in equilibrium, will not perfectly mimic good types because if they did, the reputation from taking correct decisions would coincide with the reputation from taking any other decision and all career incentives would vanish.
Once we have analysed the behaviour of the agents without disclosure of their private interests, we can study the effect of its disclosure. When the evaluator knows the private interest $\beta$ of the agent, taking the correct decision is not the only relevant information the evaluator has in order to assess the reputation of the decision maker. In particular, the evaluator will also take into account whether the decision taken coincides with the private interest of the decision maker. The beliefs of the evaluator given the observation $(\beta, d, w)$ and his conjecture of $\alpha$ are:

$$R(\beta, w, d) = \frac{\mu \sum_{s'} Pr(s = s'|w) Pr(\alpha(1, \beta, s') = d)}{\mu \sum_{s'} Pr(s = s'|w) Pr(\alpha(1, \beta, s') = d) + (1 - \mu) \sum_{s'} Pr(s = s'|w) Pr(\alpha(0, \beta, s') = d)}$$

(3)

As a first step, we will derive reputation incentives that will characterize the equilibria with disclosure of the private interest of decision makers:

**Lemma 3.** In equilibrium, $R(\beta, \beta^c, \beta) \leq R(\beta, \beta, \beta) \leq R(\beta, \beta, \beta^c) \leq R(\beta, \beta^c, \beta^c)$

With disclosure, given that the cost of contradicting the private interest is always larger for bad agents than for good ones, contradicting the private interest always signals higher reputation than following it. Moreover, conditional on contradicting or following the private interest, taking the correct decision also increases the reputation given that good types incur a cost from not taking correct decisions. Recall that, without disclosure, good agents always maximized both their reputation and their present utility by following the signal. With disclosure, this is only the case when the signal and the private interest do not coincide:

**Lemma 4.** In equilibrium, $\alpha(1, \beta, s) = s$ when $s \neq \beta$,

12
When good agents receive a signal different from their private interest, they experience no trade-off because following the signal maximizes both the probability of taking their preferred decision and their reputation. Thus, when the signal does not coincide with their private interest, good agents always follow the signal. However, when a good type receives a signal that coincides with his private interest, he experiences a trade-off between his preferred decision and his reputation. In the rest of this section we will prove the conditions for the existence of three different types of equilibria. First we will study the “Nothing Changes” equilibria, that is, equilibria where good types follow the signal and bad types follow their private interest. Second we will study the the Disciplining Effect Equilibria such that good types follow their signal and bad types mix between following the signal and their private interest. Finally we will study the Pandering Equilibria such that good types contradict their private interest and ignore the signal. Let’s start analysing the “Nothing Changes” equilibria;

**Proposition 5.** There exists an equilibrium such that good agents follow their signal and bad agents follow their private interest if and only if $\phi < \bar{\phi}_D(q, \mu)$

When reputational incentives are low enough, reputation does not induce either good nor bad decision makers to shift from their preferred decision and, therefore, good decision makers follow the signal and bad decision makers follow the private interest. Recall that this was already the case in the equilibrium without disclosure. However, disclosure tightens the conditions for its existence:

**Corollary 6.** $\bar{\phi}_D < \bar{\phi}_S$

Disclosure induces larger reputational incentives to depart from the preferred decision. On the one hand, good decision makers have incentives to contradict their private interest (without disclosure they had no incentives to do it) and, on the other hand, bad decision
makers anticipate a larger drop of his reputation if they take a wrong decision aligned with their private interest. Let $\phi_1$ and $\phi_0$ be the maximum reputational concerns that make good and bad decision makers not deviate from this equilibrium (notice that $\bar{\phi}_D = \min\{\phi_0, \phi_1\}$).

When $\phi_0 < \phi_1$, bad types require a lower $\phi$ than good types to deviate from the equilibrium, that is, following the signal when it does not coincide with their private interest. When $\phi_0 > \phi_1$, good types are the ones more prone to deviate and contradict their signal when it coincides with their private interest.

**Lemma 7.** $\phi_0 < \phi_1$ if and only if $\mu > \bar{\mu}(q)$ and $\bar{\mu}(q)$ is decreasing in $q$.

The previous lemma characterizes the $(\mu, q)$ that guarantees that $\phi_0 < \phi_1$. In particular we have that $\mu$ and $q$ have to be large enough. The intuition of the result is the following. When the precision of the signal is high, good decision makers have high opportunity costs of contradicting private interest. In addition to this, high precision of the signal also induces good decision makers to take the appropriate decision with higher probability and, therefore,
the reputational incentives of contradicting the private interest instead of following the signal are lower. With respect to the proportion of good decision makers, the larger the fraction of good decision makers is, the larger is the reputation of taking the appropriate decision which again decreases the incentives of good decision makers to contradict their private interest. Once we have found the conditions such that \( \phi_0 < \phi_1 \), we can characterize the conditions for the existence of the Disciplining Effect Equilibrium:

**Proposition 8.** There exists Disciplining Effect Equilibrium such that good decision makers follow their signal and bad decision makers mix between following the signal and the private interest if and only if \( \mu \geq \bar{\mu}(q) \) and \( \phi \in [\phi_0, \phi_1] \).

The previous proposition states that when the precision of the signal is high enough and career concerns are neither too large nor too low, there exists an equilibrium such that good decision makers follow their signal and bad decision makers mix between following their signal and their private interest. Recall that, without disclosure when career concerns were high enough, there already existed an equilibrium such that bad decision makers mixed between their private interest and their signal. The next corollary shows in which equilibrium bad types follow their signal more often.

**Corollary 9.** In the Disciplining Effect Equilibrium, bad decision makers follow the signal with higher probability than without disclosure.

The previous result shows that disclosure can be effective at disciplining bad decision makers and induce them to follow their signal more often. The intuition is that when good types follow the signal, taking a decision aligned with the private interest of the decision maker that does not match the state of the world is a strong signal of being a bad type. Therefore, the incentives for following the signal are higher with this equilibrium than without disclosure. Once we have presented the Disciplining Effect Equilibrium we can present to the Pandering Equilibrium:
Proposition 10. There exists an equilibrium such that good decision makers always contradict their private interest if and only if $2q - 1 < \phi$. In particular, this equilibrium is such that: (i) If $\phi \in (2q - 1, 1]$, good decision makers contradict their private interest and bad decision makers follow it. (ii) If $\phi \in (1, \frac{1}{\mu})$, good decision makers contradict their private interest and bad decision makers mix between contradicting and following it. (iii) If $\phi \geq \frac{1}{\mu}$, good and bad decision makers contradict their private interest.

When the precision of signal received by decision makers is low with respect to their reputational concerns, it is optimal for good decision makers to contradict their private interest in equilibrium. Intuitively, contradicting the private interest always increases the reputation of a decision maker. Thus, the higher the career concerns are, the higher the benefits of contradicting the private interest. This strategy comes at the cost of taking their preferred decision less often. Notice that this cost depends on the precision of the signal. In particular, the lower the precision of the signal is, the lower the cost of contradicting the private interest and therefore, it is more profitable to contradict the private interest. Actually, when $q$ tends to $\frac{1}{2}$, this cost vanishes.

Regarding bad types, recall that when good decision makers do not follow their signal but they just contradict their private interest, they have no incentives to follow the signal but, in any case, to contradict their private interest. Of course, bad types have less incentives than good types to contradict their private interest.

By construction, the first equilibrium cannot coexist with the other two but this is not the case for the last two. The next corollary characterizes the coexistence of Disciplining Effect equilibrium and Pandering equilibrium:

Corollary 11. For any $(\mu, q)$, if a Disciplining Effect Equilibrium exists, a Pandering Equilibrium exists too.
This result is not surprising given that the utility of a good type when following the signal is lower than the utility of a bad type when following his private interest. For bad types it is more costly to depart from their preferred decision than to good types. This would not be the case if we make different informational assumptions. For instance we could assume that decision makers observe the state of the world and the evaluator only observes a signal of it. This assumption would be beyond the scope of this models. We leave this exercise for future research.

Once we have derived all the equilibria with disclosure, we can provide the main result of the paper. In order to compare the institutional setting with disclosure of private interests and the one without, we select the equilibria that maximize the probability of correct decisions in both settings.

**Proposition 12.** (i) When $\phi < \phi_0$, disclosure has no effect. (ii) When $\phi \in [\phi_0, \phi_1]$, disclosure increases the probability of a correct decision. (iii) When $\phi > \phi_1$, disclosure decreases the probability of a correct decision.

When career concerns are low, disclosing the private interest does not make any difference. When career concerns are intermediate and agents are sufficiently informed, the disclosure of the agent’s private interests increases the probability of taking correct decisions. In the rest of the situations, the probability of a correct decision without disclosure is at least as large as with disclosure.

### 6 The Political Agency model

In this section we reframe the model into a standard political agency model where the decision maker instead of caring about reputation per se, cares about reelection. In such a political agency model we can capture the sorting effect of disclosure, that is, we can analyse
if disclosure helps voters to reelect good politicians which could be another rationale for disclosure of politicians’ private interests. Notice that if decision makers did not change their strategies with disclosure, disclosure would always improve sorting because, given that the strategy of bad types depends on their private interest, knowing the private interest gives information about the type of the decision maker. The problem, however, is that disclosure changes the strategies of decision makers. Therefore the effect of disclosure on sorting is not trivial.

We will assume that there are two periods $t = 1, 2$. At period 1, there is an incumbent politician and between period 1 and 2 voters decides whether to reelect the incumbent or appoint a new politician. For simplicity we will assume that the reelection probability of the incumbent is linear in his reputation and the career concerns parameter $\phi$ now measures how much politicians value reelection.

In each period $t$, the office-holder takes a decision $d_t \in \{a, b\}$ and the state of the world $(w_1)$ of the first period is not correlated with the state of the world $(w_2)$ of the second one. However, if the incumbent is reelected, his type is the same. We may think that the decision to be taken in the second period is not related to the decision taken in period 1.

Notice that in period 2, the optimal strategy of a decision maker is to take his preferred decision because he is not subject to reelection and he does not have incentives to maximize his reputation. In particular, in period 2, good decision makers will take $d_2 = s_2$ and bad decision maker will take $d_2 = \beta_2$. Voters always prefer to have good politicians in the second period than in the first one.

**Lemma 13.** When the private interests are not disclosed, the probability of having a good decision maker in the second period is: (i) constant when $\phi < \phi_S$ and (ii) decreasing in $\phi$ when $\phi \geq \phi_S$. 

18
When the private interest of decision makers is not disclosed we have a trade-off between the disciplining effect and the sorting effect: when bad types mimic the decision taken by good types, the decision taken in the first period will benefit the voters but it will be more difficult for the voters to screen the incumbent politician in order to increase the likelihood of having a good type in the following period. This is the usual trade-off between discipline and sorting: disciplining bad types improves the decision but it makes more difficult to screen them.

**Lemma 14.** When private interests are disclosed, the probability of having a good decision maker in the second period is: (i) constant when $\phi < \phi_0$. (ii) decreasing when $\phi \in [\phi_0, \phi_1)$, (iii) increasing when $\phi \in [\phi_1, 1)$ and (iv) decreasing when $\phi \geq 1$.

With disclosure, the trade-off between the disciplining effect and the sorting effect is not present anymore when $\phi$ is large enough. The reason is that when $\phi > 1$, bad types start mimicking good types. This does not increase the probability of taking correct decisions because good types pander and contradict their private interest independently of their signal. Thus by mimicking good types, voters are less able to screen good politicians and sorting is reduced. In particular, when $\phi > \frac{1}{\mu}$, there is no sorting at all and the probability of having a good decision maker in the second period coincides with the probability of having a good decision maker in the first one.

Once we have studied the sorting effect of each informational setting we can compare both to see when disclosure increases the probability of having a good decision maker in the second period and when it decreases it.

**Proposition 15.** Disclosure of the politician’s private interest, (i) When $\phi < \phi_D$, increases the probability of having a good politician. (ii) When $\phi \in (\phi_0, \phi_1)$ decreases the probability of having a good politician. (iii) When $\phi \in (\phi_1, \phi_m)$ increases the probability of having a good politician. (iv) When $\phi > \phi_m$ decreases the probability of having a good politician.
When career concerns are low, disclosure does not change the behaviour of politicians but voters can use the information on their private interests to better screen politicians and, consequently, the effect of disclosure on voter’s utility is positive. When career concerns are intermediate, disclosure increases the probability of correct decisions in the first period but decreases the probability of having a good politician in the second one or increases the probability of having a good politician in the second period at the cost of having less correct decisions in the first one. When this happens the effect of disclosure is ambiguous. Finally, when career concerns are high enough, disclosure decreases both the probability of correct decisions in the first period and the probability of having a good politician in the second one and the effect of disclosure on voters’ utility is negative.

7 Extensions

7.1 Multiple decisions

So far we have assumed that there were only two possible states of the world and, consequently, two only possible private interests, actions and signals. In this section I study what happens when there are \( n > 2 \) of them. More precisely, the space of states (and private interests) is given by \( \Omega = \{ w_1, \cdots, w_n \} \), all states are ex-ante equally likely \( Pr(w = w_i) = \frac{1}{n} \) and, the signal is informative, that is, the precision of the signal is such that \( q = Pr(s = w|w) > \frac{1}{n} \).

The first consequence of expanding the space of the states is that it creates higher incentives to follow the signal. From the perspective of good agents, contradicting the signal is more costly because the probability of taking their preferred decision when doing so is \( \frac{1-q}{n-1} \) which is decreasing in \( n \). From the perspective of bad agents, the cost is also higher because the probability of obtaining good reputation contradicting the signal is lower too. Thus,\(^{11}\) It depends on how much voters value the second period with respect to the first one.
Figure 2: Values of $q$ and $\mu$ such that a Disciplining Effect equilibrium exists. From $n = 2$ (dark) to $n = 6$ (light)

the larger is the space the larger is the area where that the disclosure Disciplining Effect equilibrium exists:

**Proposition 16.** If $\mu > \mu(q, n)$, there always exists a $\phi > 0$ such that a Disciplining Effect equilibrium exists. Moreover, $\mu(q, n)$ is decreasing in $n$.

The second consequence of expanding the set of states is that the Pandering Equilibrium is leads to more correct decisions than with only two possible states. The reason is that, when $n = 2$, contradicting the private interest implies ignoring the signal because there is only one way to contradict the private interest (the complement of the private interest is a singleton). This is not the case when $n > 2$. In particular, among all the possible decisions that involve contradicting the private interest, good decision makers can choose the one that matches the signal except when the private interest and the signal coincide. Less information is lost.
As $n$ increases, the probability that the private interest coincides with the signal decreases but the probability of taking a wrong decision conditional on contradicting the signal increases. The next proposition shows that the first effect always dominates the second.

**Lemma 17.** The probability of taking an appropriate decision contradicting the private interest is increasing in $n$.

When $n = 2$, the Pandering equilibria always reduces the correct decisions with respect to the equilibrium without disclosure because given that good types ignored the signal, bad types did it too. When $n > 2$, good types do not ignore the signal in the Pandering Equilibrium and this can serve as an incentive for bad type to follow the signal. Thus, as the next proposition shows, the decrease of correct decisions driven by having good types pandering instead of following the signal can be compensated by the disciplining effect on bad types.

**Proposition 18.** When $n > 2$, there exists a $\hat{\phi} > 0$ such that the Pandering equilibrium with disclosure increases the probability of a correct decision. $\phi > \hat{\phi}$.

Thus, interestingly when $n$ is large enough disclosure increases the probability of a correct decision even when good types pander. Notice that this happens precisely when the career concerns are high enough because bad types need high career concerns to contradict their private interest.

### 7.2 Observability of the state of the world

In this section we will discuss what happens when we relax the assumption that the evaluator perfectly observes the state of the world when he assesses the reputation of the agent. First, notice that if the state of the world was not observed, given that the distribution of private
interests and states of the world is symmetric, the evaluator could only infer the type of
the decision maker based on whether he followed or contradicted his private interest. In
particular, given that the evaluator can not assess whether the decision maker took the
correct decision, following the signal cannot lead to high reputation.

Without disclosure, all types ignore their career concerns and good types follow their
signal and appropriate decisions and bad types follow their private interest. Disclosing does
not change the behaviour of bad types as long as good types keep following the signal.
Suppose all types played the same strategy they would play without disclosure. Then the
reputation from following the private interest is lower than the reputation from contradicting
the private interest. Now, notice that good types are more prone than bad types to contradict
their private interest when the signal is aligned with the private interest because their loss of
immediate utility is \((2q - 1)\) whereas the loss of bad types would be 1. Thus, any equilibrium
with disclosure, is either a nothing changes equilibrium or a Pandering equilibrium and
disclosure can only reduce efficiency.

The same reasoning applies if instead of assuming that the state of the world is not ob-
served, we assume that the observability is noisy for example if we assume that the evaluator
receives a signal \(s' \in \Omega\) of the state of the world such that \(q' = Pr(s' = w|w) \in (.5, 1)\). When
\(q\) is large enough we are in the situation studied in sections 4 and 5 and when \(q\) is low enough
we are in the one studied in the previous paragraph. To wrap up, if the evaluator does not
observe perfectly the state of the world, the bad effects of disclosure are more pronounced
than when it does.
7.3 Strictly careerist good decision makers

One of the critical assumptions of the model was that good decision makers care about taking the correct decision. In this section I study what happens when good decision makers only have career concerns, that is their utility is simply $\phi R$. Without disclosure, we can still sustain the equilibrium such that good decision makers follow their signal and bad decision makers either follow their private interest or they mix between following the signal and the private interest.

When we analyze the disclosure case, it is straightforward to see that good decision makers will always pander and propose the policy that goes against their private interest. Thus, the only possible equilibria with disclosure is the pandering equilibrium of the previous section and, when good decision makers do not care about the decision they take, disclosure always reduces the probability of a correct decision.

**Proposition 19.** When good decision makers only have career concerns, in equilibrium, good decision makers always contradict their private interest and disclosure always reduce the probability of a correct decision.

8 Conclusion

In this paper, I have shown that disclosure of careerist decision makers’ private interests can have different consequences. In particular, when reputational concerns are intermediate and decision makers are sufficiently informed, disclosure of decision makers’ private interests increases the probability of taking a correct decision. However, when career concerns are higher or decision makers are less informed, disclosure can decrease the probability of taking correct decisions. These results hold with different model specifications. Moreover, I extend the model into a political agency model where I show that disclosure can help voters to select
better politicians but it can also hinder their ability to do it. In particular, disclosure can simultaneously worsen the decision taken and the selection of politicians.

When private interests are disclosed, decision makers can be evaluated not only on the correctness of their decisions but also on whether these decisions were aligned with their interests. This can prevent decision makers from putting their interests in front of their duties because they can be judged more severely. However, this also distorts the accountability process because decision makers instead of being only accountable for the correctness of the decisions they are also accountable for the profits they obtain.

Disclosure increases accountability but it comes at the cost of distorting the accountability mechanism. This does not mean that disclosure of decision makers’ private interest should never be enforced. The benefits can exceed the costs when decision makers are sufficiently informed or when the decision maker has multiple possible decisions. However, in other situations disclosure is not desirable.

References


Appendix

Proof of Proposition 1

Proof. (i) Suppose good types follow their signal and bad types follow their private interest. Consistency of the beliefs requires:

\[ R(w, w) = \frac{\mu q}{\mu q + (1 - \mu)\frac{1}{2}}, \quad R(w, wc) = \frac{\mu(1 - q)}{\mu(1 - q) + (1 - \mu)\frac{1}{2}} \]  

(4)

Subtracting both reputations we have that taking the appropriate decision leads to higher reputation than taking the wrong one:

\[ R(w, w) - R(w, wc) = \frac{2(2q - 1)(1 - \mu)\mu}{1 - (2q - 1)^2\mu^2} > 0 \]  

(5)

For good types, the expected utility of following the signal is given by \( q + \phi(qR(s, s) + (1 - q)R(s^c, s)) \) and the expected utility from contradicting the signal is \( (1 - q) + \phi((1 - q)R(s^c, s^c) + qR(s, s^c)) \). Therefore good decision makers follow their signal if and only if:

\[ q + \phi(qR(s, s) + (1 - q)R(s^c, s)) > (1 - q) + \phi((1 - q)R(s^c, s^c) + qR(s, s^c)) \]  

(6)

And rearranging we get

\[ 2q - 1 > \phi(1 - 2q)(R(s^c, s^c) - R(s, s^c)) \]  

(7)

and, from \( q > \frac{1}{2} \) we have that the left-hand-side is positive and the right-hand-side is negative. Thus \( R(w, w) > R(w, wc) \) and good decision makers always follow their signal.

Regarding bad types, when the signal is aligned with their private interest it is immedi-
ate to see that they prefer to follow the signal. However, when the signal contradicts the private interest, the utility of following the private interest is $1 + \phi((1 - q)R(s^c, s^e) + qR(s, s^c))$ and the utility of contradicting the private interest is $\phi((1 - q)R(s^c, s) + qR(s, s))$ and bad decision makers follow their private interest if and only if

$$1 + \phi((1 - q)R(s^c, s^e) + qR(s, s^c)) > \phi((1 - q)R(s^c, s) + qR(s, s))$$  \hspace{1cm} (8)$$

And rearranging we get:

$$\phi < \frac{1}{(2q - 1)(R(s, s) - R(s^c, s))}$$ \hspace{1cm} (9)$$

Finally, plugging in the expression of the reputation functions, we obtain:

$$\phi < \bar{\phi}_S(q, \mu) = \frac{1 - (2q - 1)^2\mu^2}{2(2q - 1)^2(1 - \mu)\mu}$$ \hspace{1cm} (10)$$

Thus, when $\phi < \bar{\phi}_S(q, \mu)$, there exists an equilibrium such that good types follow their signal and bad types follow their private interest.

(ii) Now, suppose good types follow their signals and bad types mix between following their signal and their private interest. In particular, let $\alpha(0, \beta, \beta) = \beta$ and $Pr(\alpha(0, \beta, \beta^c) = \beta^c) = x \in (0, 1]$. Notice that, in equilibrium $x < 1$ because for $x = 1$, we have that $R(w, d) = \mu$ and without reputational incentives, bad types strictly prefer to follow their private interest than following the signal. But $x \in (0, 1)$ requires bad types to be indifferent between following the signal and the private interest when $s = \beta^c$. Consistency of the beliefs requires: $R(w, w) = \frac{\mu q}{\mu q + (1 - \mu)\frac{1}{2}\left[\mu + (1 - \mu)\frac{1}{2}\right]}$ and $R(w, w^c) = \frac{\mu(1 - q)}{\mu(1 - q) + x(1 - \mu) + (1 - x)q}$. And the $x$ that makes a bad type indifferent between
following the signal and following the private interest when $s = \beta^c$ is:

$$x = \frac{\mu(2q - 1)(\phi - 1) - \sqrt{1 - \mu \phi(1 - 2q)^2(2 - \mu \phi)}}{(1 - \mu)(2q - 1)}$$

(11)

And $x \in (0, 1)$ if and only if $\phi < \bar{\phi}_S(q, \mu)$.

Finally notice that $\bar{\phi}_S(q, \mu) > 1$:

$$\bar{\phi}_S(q, \mu) = \frac{1 - (2q - 1)^2 \mu^2}{2(2q - 1)^2(1 - \mu)\mu} > 1 \leftrightarrow \frac{1}{\mu(2 - \mu)} > (2q - 1)^2$$

(12)

Now, from $\mu \in (0, 1)$ we have that the left-hand-side is larger than one and from $q \in (.5, 1)$ the right-hand-side is smaller than one.

Proof of Corollary 2

Proof. Good types always follow the signal independently of $\phi$. When $\phi \leq \phi_S$, bad types follow always their private interest and the probability of taking a correct decision is constant for all $\phi \leq \phi_S$. When $\phi > \phi_S$, the probability that a bad type follows the signal is increasing in $\phi$. Therefore the probability of taking a correct decision is also increasing in $\phi$.

Proof of Lemma 3

Proof. First we will prove that for every $(\beta, s)$, $Pr(\alpha(0, \beta, s) = \beta) \geq Pr(\alpha(1, \beta, s) = \beta)$. Let $u(\theta, s, d)$ be the expected utility of a decision maker of type $\theta$ and private interest $\beta$ when
he observes a signal $s$ and takes a decision $d$.

\[
u(1, \beta, \beta) - u(1, \beta, \beta^c) = 2q - 1 + \phi (q(R(\beta, \beta) - R(\beta, \beta^c)) + (1 - q)(R(\beta^c, \beta) - R(\beta^c, \beta^c)) < 1 + \phi (q(R(\beta, \beta) - R(\beta, \beta^c)) + (1 - q)(R(\beta^c, \beta) - R(\beta^c, \beta^c)) = u(0, \beta, \beta) - u(0, \beta, \beta^c) \tag{13}\]

\[
u(1, \beta^c, \beta) - u(1, \beta^c, \beta^c) = 2q - 1 + \phi ((1 - q)(R(\beta, \beta) - R(\beta, \beta^c)) + q(R(\beta^c, \beta) - R(\beta^c, \beta^c)) < 1 + \phi ((1 - q)(R(\beta, \beta) - R(\beta, \beta^c)) + q(R(\beta^c, \beta) - R(\beta^c, \beta^c)) = u(0, \beta^c, \beta) - u(0, \beta^c, \beta^c) \tag{14}\]

Thus, since low types always find more profitable than high types to follow the private interest than high types, if high types follow the private interest, low types to. In particular if $x^*_\theta = Pr(\alpha(\theta, \beta, s) = \beta)$, then $x^*_1 \leq x^*_0$.

\[
R(\beta, \beta) = \frac{\mu(q x^{\beta}_1 + (1 - q)x^{\beta^c}_1)}{\mu(q x^{\beta}_1 + (1 - q)x^{\beta^c}_1) + (1 - \mu)(q x^{\beta}_0 + (1 - q)x^{\beta^c}_0)} \tag{15}\]

\[
R(\beta, \beta^c) = \frac{\mu(q(1 - x^{\beta}_1) + (1 - q)(1 - x^{\beta^c}_1))}{\mu(q(1 - x^{\beta}_1) + (1 - q)(1 - x^{\beta^c}_1)) + (1 - \mu)(q(1 - x^{\beta}_0) + (1 - q)(1 - x^{\beta^c}_0))} \tag{16}\]

And, rearranging we get that $R(\beta, \beta) > R(\beta, \beta^c)$ if and only if

\[
\frac{q x^{\beta}_1 + (1 - q)x^{\beta^c}_1}{q x^{\beta}_0 + (1 - q)x^{\beta^c}_0} > \frac{q(1 - x^{\beta}_1) + (1 - q)(1 - x^{\beta^c}_1)}{q(1 - x^{\beta}_0) + (1 - q)(1 - x^{\beta^c}_0)} \tag{17}\]
The left-hand-side is smaller than 1 and the right-hand-side is larger. Therefore, it has to be that $R(\beta, \beta) \leq R(\beta, \beta^c)$. Analogously, we get $R(\beta^c, \beta) \leq R(\beta^c, \beta^c)$. Finally notice that $R(\beta^c, \beta) \leq \mu \leq R(\beta^c, \beta^c)$.

**Proof of Lemma 4**

Proof. From the previous lemma we know that, in equilibrium, for every realization of the state of the world, reputation is higher when the decision maker contradicts the private interest than when he follows it. Given that when $s = \beta^c$, $w = \beta^c$ is more likely than $w = \beta$, good decision makers always choose $\beta^c$.

**Proof of Proposition 5**

Proof. Suppose $\alpha(1, s) = 1_{s=\beta}$ and $\alpha(0, s) = 1$. Consistency of the beliefs requires $R(w, \beta, \beta^c) = 1$, $R(\beta, \beta, \beta) = \frac{\mu q}{\mu q + (1 - \mu)}$ and $R(\beta, \beta, \beta) = \frac{\mu(1 - q)}{\mu(1 - q) + (1 - \mu)}$. The incentive compatibility condition for good types is:

$$
\phi < \phi_1 = \frac{(2q - 1)(\mu + \mu^2(q - 1)q - 1)}{(1 - \mu)(\mu(2q - 1)q + 1) - 1}
$$

(18)

and the incentive compatibility condition for low types is:

$$
\phi < \phi_0 = \frac{1 - \mu - \mu^2(q - 1)q}{(1 - \mu)(2\mu(q - 1)q + 1)}
$$

(19)

Let $\bar{\phi}_D := \min(\phi_0, \phi_1)$ Therefore this equilibrium exists if and only if $\phi < \bar{\phi}_D$. Finally from $\mu \in (0, 1)$ and $q \in (.5, 1)$, it follows that both $\phi_0$ and $\phi_1$ are strictly positive.

**Proof of Corollary 6**

32
Proof. We will prove that \( \varphi_0 < \bar{\varphi} \).

\[
\text{sign} (\varphi_0 - \bar{\varphi}) = \text{sign} \left( \frac{\mu(\mu + 2(2\mu - 3)(q - 1)q - 2) + 1}{2(\mu - 1)\mu(1 - 2q)^2(2\mu(q - 1)q + 1)} \right)
= -\text{sign} (\mu(\mu + 2(2\mu - 3)(q - 1)q - 2) + 1)
= -1
\] (20)

\[\square\]

Proof of Lemma 7

Proof.

\[
\varphi_1 - \varphi_0 = \left( \frac{\mu + \mu^2(q - 1)q - 1}{\mu - 1}(\mu + 2q(2\mu(q - 1)q + 1) - 2) \right)
\] (21)

And, rearranging we get that \( \varphi_1 - \varphi_0 > 0 \) if and only if

\[
\mu > \bar{\mu} = \frac{2 - 2q}{4(q - 1)q^2 + 1}
\] (22)

And \( \frac{\partial \mu}{\partial q} < 0 \). \[\square\]

Proof of proposition 8

Proof. Suppose there exists an equilibrium such that good types follow the signal and the strategy of bad types is such that \( \alpha(0, \beta) = 1 \) and \( Pr(\alpha(0, \beta^c) = \beta) = x \in (0, 1) \). Let \( \Delta R(s) \) be the reputational gains from contradicting the private interest when the signal is \( s \). In particular,

\[
\Delta R(\beta) = q(R(\beta, \beta^c) - R(\beta, \beta)) + (1 - q)(R(\beta^c, \beta^c) - R(\beta^c, \beta))
\] (23)

\[
\Delta R(\beta^c) = q(R(\beta^c, \beta^c) - R(\beta^c, \beta)) + (1 - q)(R(\beta^c, \beta^c) - R(\beta^c, \beta))
\] (24)
Since in both signals, good types contradict their private interest with higher probability, \( R(w, \beta^c) > R(w, \beta) \) and \( \Delta R(s) > 0 \). Thus, when \( s = \beta^c \) good types follow the signal because it increases both their reputation and their present utility and for good types we only have to show that they follow the signal when \( s = \beta \). Regarding bad types, if good types follow the signal when \( s = \beta^c \), bad types will follow the signal too. Therefore we only have to check that when \( s = \beta^c \), there exists an \( x \) such that they are indifferent.

Regarding good types, notice that from \( \phi < \phi_1 \) we know that, when \( x = 0 \), good types strictly prefer to follow the private interest when \( s = \beta \). Now, notice that \( \frac{\partial}{\partial x} R(s, \beta) > 0 \) and \( \frac{\partial}{\partial x} R(s, \beta^c) < 0 \). Therefore \( \frac{\partial}{\partial x} \Delta R(s) < 0 \), that is, the reputational gains of contradicting the private interest are decreasing on \( x \) and, therefore good types strictly prefer to follow the private interest when \( s = \beta \) for any \( x \in (0, 1) \).

Finally, regarding bad types notice that when \( x = 0 \), \( \phi R(\beta^c) > 1 \) and when \( x = 1 \), \( \phi R(\beta^c) = 0 \). Therefore by Bolzano’s theorem there exists an \( x \in (0, 1) \) such that \( \phi R(\beta^c) - 1 = 0 \).

Proof of Corollary 9

\[ \text{Proof.} \] The expected reputation gain from following the signal when \( s = \beta^c \) without disclosure is \( \phi ((2q - 1)(R(w, w) - R(w, w^c)) \) and the expected reputation gain with disclosure is \( \phi (q(R(\beta, \beta^c, \beta^c) - R(\beta, \beta^c, \beta)) + (1-q)((R(\beta, \beta, \beta^c) - R(\beta, \beta, \beta)))) \) and the second expression always exceeds the first one.

\[ \square \]

Proof of proposition 10

34
Proof. (i) Suppose \(\alpha(1, s) = \beta^c\) and \(\alpha(0, s) = \beta\). Consistency of the beliefs requires \(R(w, \beta, d)\) to be such that \(R(w, \beta, \beta) = 0\) and \(R(w, \beta, \beta^c) = 1\). With these reputational incentives, good types only experience a trade-off when the signal coincides with the private interest and the condition to sustain their strategy in equilibrium is \(q < (1 - q + \phi \leftrightarrow 2q - 1 < \phi)\). The problem of bad types does not depend on the signal they receive and the condition to sustain their strategy in equilibrium is \(\phi < 1\). Therefore this equilibrium exists when \(2q - 1 < \phi < 1\).

(ii) Suppose \(\alpha(1, s) = \beta^c\), \(Pr(\alpha(0, \beta) = \beta) = x\) and \(Pr(\alpha(0, \beta^c) = \beta) = y\). First of all notice that \(x = y\). Suppose without loss of generality that \(x < y\), then \(R(\beta, \beta^c) - R(\beta, \beta) < R(\beta^c, \beta^c) - R(\beta^c, \beta)\), and low types would have higher reputational incentives to contradict the private interest when \(s = \beta^c\) which is a contradiction with the indifferent condition in both signals. Then, when \(x = y\), Consistency of the beliefs requires \(R(w, \beta, \beta) = 0\), \(R(w, \beta, \beta^c) = \frac{\mu}{\mu + (1-x)(1-\mu)}\), bad types are indifferent between following and contradicting the private interest if \(1 = \frac{\phi \mu}{\mu + (1-x)(1-\mu)}\) and \(x = 1 - \frac{\mu(\phi-1)}{1-\mu}\). The reputation incentives of contradicting the private interest are \(R(w, \beta, \beta^c) = \frac{1}{\phi}\) and good types always contradict the private interest. This equilibrium exists if \(x \in (0, 1)\) and this happens when \(1 < \phi < \frac{1}{\mu}\).

(iii) Suppose \(\alpha(1, s) = \beta^c\) and \(\alpha(0, s) = \beta^c\). Consistency of the beliefs requires \(R(w, \beta, \beta^c) = \mu\) and the less restrictive out of equilibrium belief is \(R(w, \beta, \beta) = 0\). The binding incentive compatibility condition of good decision makers is \(2q - 1 < \phi \mu\) and the incentive compatibility condition of bad decision makers is \(1 < \phi \mu\). Therefore, this equilibrium exists when \(1 < \frac{1}{\mu} < \phi\).

\(\square\)

Proof of corollary 11
Proof. A Disciplining Effect Equilibrium exists if $\phi \in [\phi_0, \phi_1]$ and a Pandering Equilibrium exists if $\phi > 2q - 1$. In order to prove that for any $(\mu, q)$, if a Disciplining Effect equilibrium exists, a Pandering Equilibrium exists too, it is sufficient to prove that for any $(\mu, q)$, $2q - 1 < \phi_j$ for $j \in \{0, 1\}$.

Notice that $\phi_0 < 2q - 1$ if and only if:

$$\frac{\mu(2q - 1)(1 - \mu + (3\mu - 2)(1 - q)q)}{(1 - \mu)(1 - \mu(2(q - 1)q + 1))} > 0 \quad (25)$$

Which always holds because the denominator is positive and the numerator too. And $\phi_1 < 2q - 1$ if and only if:

$$1 < \frac{(1 - \mu(1 - q))(1 - \mu q)}{(1 - \mu)(1 - \mu(1 - 2(1 - q)q))} \quad (26)$$

Which is always satisfied for $\mu \in (0, 1)$ and $q \in (.5, 1)$.

Proof of Proposition 12

Proof. (i) Follows from Proposition 5. (ii) follows from Proposition 8. (iii) follows from Proposition 10.

Proof of Lemma 13

Proof. When $\phi \leq \phi_S$, an increase in $\phi$ does not change the behaviour of any of the types and the probability of having a good type in the second period is constant. However, when $\phi > \phi_S$, the probability that a bad type follows the signal is increasing and $R(w, w)$ decreases and $R(w, w^c)$ increases, therefore, the probability that a good type is reelected decreases and the probability that a bad type is reelected increases. Thus, the probability of having a good
Proof of Lemma 14

Proof. When $\phi < \phi_D$, good types follow the signal and bad types follow their private interest both with disclosure and without disclosure. is

Without disclosure, the probability that a good type is reelected is:

$$E[R|\theta = 1, ND] = qR(w, w) + (1 - q)R(w, w^c)$$ (27)

And the probability that a bad type is reelected is:

$$E[R|\theta = 0, ND] = \frac{1}{2}(R(w, w) + R(w, w^c))$$ (28)

The probability of having a good type in period 2 is:

$$Pr[\theta_2 = 1|ND] = \mu(E[R|\theta = 1, ND] + (1 - E[R|\theta = 1, ND])\mu) + (1 - \mu)(1 - E[R|\theta = 0, ND])\mu$$ (29)

And, plugging-in the reputation expression we get:

$$Pr[\theta_2 = 1|ND] = \frac{\mu(-\mu((\mu - 3)\mu + 1)(1 - 2q)^2 - 1)}{\mu^2(1 - 2q)^2 - 1}$$ (30)

Now, with disclosure, the probability that a good type is reelected is:

$$E[R|\theta = 1, D] = \frac{1}{2}(q(R(\beta, \beta, \beta) + R(\beta, \beta^c, \beta^c)) + (1 - q)(R(\beta, \beta, \beta^c) + R(\beta, \beta^c, \beta))$$ (31)
And the probability that a bad type is reelected is:

$$E[R|\theta = 0, D] = \frac{1}{2}(R(\beta, \beta, \beta) + R(\beta, \beta^c, \beta)) \quad (32)$$

The probability of having a good type in period 2 is:

$$Pr[\theta_2 = 1|D] = \mu(E[R|\theta = 1, D] + (1 - E[R|\theta = 1, D])\mu) + (1 - \mu)(1 - E[R|\theta = 0, D])\mu \quad (33)$$

And, plugging-in the reputation expression we get:

$$Pr[\theta_2 = 1|D] = \frac{\mu(\mu(-\mu - 2((\mu - 3)\mu + 1)(q - 1)q - 4) - 3)}{2(\mu + \mu^2(q - 1)q - 1)} \quad (34)$$

Finally,

$$Pr[\theta_2 = 1|D] - Pr[\theta_2 = 1|ND] = \frac{(\mu - 1)^2\mu(\mu + 2(2\mu - 3)(q - 1)q - 2) + 1}{2(\mu^2(1 - 2q)^2 - 1)(\mu + \mu^2(q - 1)q - 1)} > 0 \quad (35)$$

The proof of the rest of the cases is analogous.

\[\square\]

Proof of proposition 15

**Proof.** Follows from Lemma 13 and 14.

\[\square\]

Proof of proposition 16

**Proof.** Analogously to the case $n = 2$, we compute $\phi_0$ and $\phi_1$ and the condition such that
\( \phi_0 < \phi_1 \) is given by \( \mu > \mu(q, n) \) where:

\[
\mu(q, n) := \frac{(n - 1)^2 n (1 - q)}{n (n (2(q - 1) q + 1) + q (q - 5) q + 3) - 2 + 2q + 1) - 1}
\] (36)

Notice that \( \mu(q, 2) = \overline{\mu}(q) \) that we computed previously and \( \mu(q, n) \) is decreasing in \( q \) and in \( n \).

**Proof of lemma 17**

*Proof.* When contradicting the private interest, the probability of taking an appropriate decision is:

\[
\Pi(n) = \frac{1}{n} \left( (n - 1) q + \frac{1 - q}{n - 1} \right)
\] (37)

Notice that \( \Pi(2) = 2 \) which is the expression we already obtained in the model. Finally, we compute the partial derivative of \( \Pi \) with respect to \( n \) and we obtain:

\[
\frac{\partial \Pi}{\partial n} = \frac{1 + n (-2 + nq)}{(n - 1)^2 n^2}
\] (38)

which is positive if and only if \( \frac{2n - 1}{n^2} < q \) which always hold because

\[
\frac{2n - 1}{n^2} = \frac{2 - \frac{1}{n}}{n} < \frac{1}{n} < q
\] (39)

**Proof of proposition 18**

*Proof.* First of all we will derive the equilibrium with secrecy for \( n > 2 \). Analogously to the case \( n = 2 \), it is always sustainable for good types to follow the signal. With this strategy, in equilibrium, bad types can either follow follow their private interest or mix between following
the private interest and the signal. Any other action would be suboptimal because it would reduce the expected reputation with respect to following the signal without increasing the present utility. Now, notice that the equilibrium is such that, there exists a \( \tilde{\phi}_S \) such that when \( \phi < \tilde{\phi}_S \), bad types follow the private interest and when \( \phi \geq \tilde{\phi}_S \), bad types follow the signal with probability \( x \) and follow the private interest with probability \( 1 - x \).

It is immediate to see that:

\[
\tilde{\phi}_S = \frac{(n - 1)(\mu(nq - 1) + 1)(\mu q - 1) + 1 - \mu}{(1 - \mu)nq(nq - 1)^2} \tag{40}
\]

\[
x = \frac{n^2 - 3n + 2 + \mu(nq - 1)(n(\phi - 2) + 2) - n\sqrt{\mu^2 \phi^2(nq - 1)^2 - 2\mu(n - 1)(2q - 1)\phi(nq - 1) + (n - 1)^2}}{2(1 - \mu)(n - 1)(nq - 1)} \tag{41}
\]

And, in terms of welfare, the probability that an appropriate decision is taken is:

\[
W_S = \frac{n - 1 + \mu\phi(nq - 1) - \sqrt{\mu^2 \phi^2(nq - 1)^2 - 2\mu(n - 1)(2q - 1)\phi(nq - 1) + (n - 1)^2}}{2(n - 1)} \tag{42}
\]

Now, regarding the Pandering equilibrium, let’s focus on the case such that all types follow the signal when it does not coincide with their private interest and they mix between the remaining actions when it does. We can impose the out of equilibrium belief \( R(w = W, d = \beta) = 0 \) and consistency of the beliefs requires \( R(w = W, d \neq \beta) = \mu \). Now, the only binding IC condition is for bad types and it is simply:

\[
\tilde{\phi}_D := \frac{1}{\mu} < \phi \tag{43}
\]
Now, in terms of welfare, the probability that an appropriate decision is taken is:

\[ W_D = \frac{(n-2)nq + 1}{(n-1)n} \]  \hspace{1cm} (44)

Notice that \( \bar{\phi}_D < \bar{\phi}_S \), that is when there exists a mixed equilibrium with secrecy, there exists the disclosure Pandering equilibrium too. Finally, \( W_D > W_S \) if and only if:

\[ \phi > \hat{\phi} := \frac{(n^2(1-q) + n(2q-1) - 1)((n-2)nq + 1)}{\mu n(nq - 1)^2} \]  \hspace{1cm} (45)

Notice that \( \frac{\partial \hat{\phi}}{\partial n} \), \( \frac{\partial \hat{\phi}}{\partial \mu} \) and \( \frac{\partial \hat{\phi}}{\partial q} < 0 \)

**Proof of proposition 19**

*Proof.* Since good types only maximize their reputation and do not care about the decision, they will simply take the decision that gives them higher reputation. We will distinguish between two situations. Suppose \( R \) is constant, when this is the case, bad types always follow the private interest. Now, in order to satisfy the beliefs, good types should also follow their private interest. However, this is PBE is not attractive. Therefore it has to be that the reputation from contradicting the private interest is larger than the reputation from following it and then, good types always contradict their private interest.

\[ \square \]