The London School of Economics and Political Science

Essays in Organisational Economics

Melania Nica

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy

London, July 2014

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 25221 words.

Abstract

This thesis consists of three chapters. The first two chapters explore the effect of career concerns on communication by multiple experts. The third chapter addresses corporate governance as a double layered moral hazard.

The first two chapters relate to a model where a decision maker acts over two periods on the advice of two imperfectly informed experts. Both experts are possibly biased, but in opposite directions. The decision maker can only rely on the experts' reports to determine a course of action, as he never observes the true state of the economy. I show that the experts may report in the opposite direction of their possible bias not only for reputational reasons, but also as a strategic response to the possibility of misreporting by their counterpart. This model also provides a new justification for conformity: an expert might send the same message as the other, not in order to look similar, but to distinguish herself. This is done by inviting comparison to the reliability of the other expert. I also show that a decision maker could discipline both experts to disclose their information by making one value the future more. Also, an expert might be made to tell the truth by being paired with another with high initial reputation. However, negative outcomes still persist, such as the possibility that unbiased experts end up misreporting their signals in order to disavow their perceived predisposition.

In the third chapter I study self-dealing in organizations where investors are aware of the existence of different participants in a project. The model involves two-layers of moral hazard, where a manager acts simultaneously as an agent to an investor and as a principal to the employees of the firm. The manager's role is to determine the allocation of the uncontractible resources at his discretion. The optimal executive compensation offered by the investor takes into account the ease with which the employees exert effort and the trade-offs that arise in the process of committing resources.

Acknowledgments

I am extremely grateful to my supervisors, Francesco Nava and Pasquale Schiraldi for making this thesis possible. Without their steadfast support and belief in me I would not be at this stage.

Francesco has given generously of his time, inspiring me with thought provoking discussions and at the same time guiding me through the process of taking an idea and developing it into a paper. Our discussions were always thorough, and although I received detailed critical advice, I was also offered the freedom to think independently. He continues to inspire me to challenge myself and to grow as a researcher.

From the moment Pasquale started supervising me, he placed an alternative point of view in front of me, motivating me to think out of the box. His extremely positive encouragement and gentle guidance made me understand better how to present my work and in turn, led me to prioritize my research questions. Pasquale has the art of asking questions that surprise me with new insights.

For my work Gilat Levy has been a special inspiration in terms of her knowledge and her kindness. I will always aspire to her standards of intellectual enquiry.

I got to know Ronny Razin as his teaching assistant for political economy. His support during a very difficult period in my personal life means a lot to me. He has not stopped supporting me since then, always offering rigorous criticism of my research and advice on career development.

Vassilis Hajivassiliou not only taught me econometrics at LSE but also gave me the opportunity to teach it. He is not only a mentor but has also become a friend.

The economic theory group at LSE has been an excellent nurturing environment, offering the opportunities one needs to develop as a scholar. Many of the faculty members have made themselves available and have taken an interest in the development of my research. I am grateful to each of the members of faculty, my colleagues in the PhD programme, and to the administrative team, especially Mark Wilbor and Sam Keenan, for the countless ways in which they made my time at LSE enriching.

Since before I came to LSE, Nikolay Gospodinov has been an important influence on my path and no work that I produce should go without acknowledging him.

From my friend Jeanne Sahni I learned that in life there are times when we must stand firm. She has been an amazing inspiration in my personal life.

Due to a series of circumstances, this PhD has been a long part of my life, and I would like to thank all my friends who stood by me during the journey.

My husband Jaideep has been constantly at my side, not only in our shared adventures but also as a sounding board and discussion partner for my ideas.

Neeta and Harinder offered me all the love and stability I needed at difficult times and I want to thank them for being a family to me. Sheena is one of my best friends - I always enjoyed all the “mature” discussion we had. Amalia is a free spirit and a wonderful sister and she has always motivated me to move forward.

In the end I would not be here without my parents. They supported me in all my decisions without ever failing to believe in me. I owe my intellectual curiosity to them both: to my father for encouraging me to read books “not really for my age” and to my mother for always dreaming big for her through me. During my PhD my mother passed away and this has left a big hole that can never be filled.

I dedicate this PhD thesis to her. *Maman I wish you were with me, at the end of this endeavour.*

Table of Contents

Introduction

This thesis focuses on questions of governance within a career concerns framework or through an optimal contracts perspective.

Many times, we are unable to gain knowledge that would guide our actions. This is why we rely on experts' opinions in various fields. We often observe that decision makers need to take action based on the opinion of experts from opposing sides. What happens when they cannot verify the veracity of the experts' claims? In the first two chapters of this thesis, I describe a situation where informed experts motivated by reputational concerns outwardly appear to agree even though they provide biased information. Under certain conditions on relative initial reputations, one of the experts is disavowing her perceived bias as a strategic reaction to possible misreporting by the other. I call this 'conformity as separation'.

This paper is connected to different strands of the career concerns literature. Morris (2001) studies reputational distortions when there is uncertain misalignment of preferences between a principal and an agent. My innovation is to incorporate strategic interaction and unverifiability of states in this context. I also connect the model to the concept of anti-herding (e.g. see Levy, 2004), developed in a setting where there is asymmetric information on experts' abilities rather than their preferences.

A political economy example of this model could refer to a situation where two leaders from opposing parties recommend policies to an uninformed electorate before an election. Both leaders could be fair and have preferences aligned with the electorate or they could be biased towards their party agenda. Once the election takes place and a policy is implemented the true state is not verifiable any longer. Another example could be taken from organizations when an employee and her supervisor report to the CEO on how well she performed a task so that she is considered for promotion. The honest employee reports her evaluation correctly, while the dishonest one exaggerates her performance. The supervisor in contrast could be biased against the employee or fair. The CEO is unable to verify the true performance.

In terms of comparative statics, I show that a principal could discipline both experts to disclose their information by making one value the future more. Further, an expert might be made to tell the truth by being paired with another with high reputation. However, negative aspects still remain, such as the fact that good experts might end up misreporting their signals in order to disavow their perceived predisposition.

In the third chapter, I study a firm in which a manager could optimally use his private benefits to incentivize the exertion of exceptional effort from firm's employees. The model involves two layers of moral hazard where a manager acts simultaneously as an agent for an investor and as a principal to the employees.

The corporate governance literature has dealt with four ways in which managers may act against the owners' best interests: insufficient effort, extravagant investment, entrenchment strategies, and self-dealing. As summarized in Tirole (2006), these are essentially all different moral hazard problems. In this paper I focus only on self-dealing, which is traditionally solved by compensating the manager sufficiently so that he does not divert investment for private benefit. However, I show that this may not always be the optimal incentive scheme once we take into account the overall structure of the firm.

I find that if the proportion of discretionary funds is within a specific range, the manager can be made to forego his private benefit and instead use the funds available to incentivize exertion of effort. Thus, in my model self-dealing loses the exclusively negative connotation that was attached to it by the contracts literature. I argue that an investor who sets the managerial compensation must take into account the trade-offs the manager faces in the process of disbursing resources, particularly keeping in mind the ease with which the workers perform their tasks. The model offers a plausible mechanism that relates self-dealing to the emergence of different forms of corporate governance across the world.

Chapter 1

Conforming to Stand Out: A Model of Career Concerns with Biased Experts

1 Introduction

The reliability of an expert's advice is centrally important in many economic settings, whether it be in public policy or organizational decision making. It is clear then that we must be concerned about the incentives faced by experts, which may include their desire for a particular outcome and for career advancement. In order to learn about these incentives we typically assume that any advice is eventually validated. However, there are many situations where we cannot observe the underlying information that is required. One way to overcome this could be to seek multiple opinions, but we often find that experts come from opposing backgrounds. How, then, do we interpret their reports? Do they always side with their respective ideologies? Do they sometimes agree, and can we infer from their agreement that their reports are correct?

Situations like this are seen in politics, in government, and in organizations in general. For instance, politicians may state views that are traditionally opposed to their asserted party positions. In the United States, the term "Republicrat" is widely used to refer to someone who belongs to one party but often supports the policies of the other. The phenomena of New Labour and of Compassionate Conservatism in the United Kingdom also come to mind. As a more specific example, the new leadership of the Labour party in the UK has recently expressed an openness to reducing the influence of unions (their biggest funder) on the party. Further, they announced their intention to be even tougher than the Conservatives in implementing welfare caps, a signature policy of the Conservative party in government. Some commentators claim that this is a response to persistent labeling of the current leadership of the party as being very left-leaning. How does the public determine whether these moves are based on a genuine analysis of policy imperatives, or are simply calculated to earn credibility with the electorate? On the flip side, is there an action the Conservatives could take that would be better for their reputation? Moreover, is their choice affected by the extent to which each party is perceived as being away from the centre?

In order to explore the questions above, I construct a reputation forming game involving a decision maker and two experts, where each of the experts may be biased in the opposite direction to the other. The experts report over two periods about the state of the world to the decision maker who then takes an action based on these reports. Career concerns enter because of the relative value that experts place on the present and the future. The experts are imperfectly informed about the state, while at the same time the decision maker is not able to verify it.

The state of the world is reflected in the two opposing points of view, and each of the states are equally probable. The decision maker chooses an action to be as close as possible to the realized state, but the only way for him to do that is to consider both reports and draw a conclusion given information about the agents' ability, potential biases, degree of career concerns, and reputation. The fact that the unverifiability of the state leads an expert's reputation to be based instead on the combined information in the two reports is a departure from the regular reputation forming models. I argue that it offers explanations for a larger pool of settings that we face in the real world.

If the experts are unbiased, their preferences are effectively aligned with the decision maker, and thus we can think of them as a good type. By contrast, the biased agents have a preference for actions that reflect one point of view, and we can think of them as the bad type. Thus, the uncertainty about types is not based on ability (signal precision is known), but on alignment of preferences with the decision maker.

The game is solved by backward induction. The objective is to identify the ways in which career concerns may distort the behavior of the experts. Naturally, in the last period, the experts have no incentive to deviate from their preferences. Thus, in the second period there exists an informative equilibrium in which the good experts report their signals while the bad experts their biases. As this situation reflects a cheap talk setting, a non-informative equilibrium also exists where no information is transmitted to the decision maker.

In the first period, the experts must trade off their respective current preferences against the incentive to report in the opposite direction of their possible bias for reputational reasons. However, they must now also strategically respond to the possibility of misreporting by their counterpart. Depending on experts' initial reputations, signal precisions, and their relative preference for the future, I show the existence of truthtelling, informative and non-informative equilibria.

In a truthtelling equilibrium the good experts disclose fully their signals while the bad ones do so partially. The informative equilibrium occurs when good experts' career concerns become more important and they also start to report their signals only partially. A limiting case is the babbling equilibrium when the good experts never give a report consistent with their perceived bias, and the bad experts of the same appartenance pool on this strategy.

In non-babbling equilibria I discover a new type of behavior: when there is a high probability that one expert will misreport the truth, the other expert will tend to offer the same report. Here, she is conforming not for the purpose of proving that she is the same as her counterpart, but in order to distinguish herself, effectively offering a comparison to the reliability of the other expert. She signals that she is of a good type but she must have committed an error (considering that her report is in the opposite direction of her own potential bias, while the counterpart's report is close to the counterpart's own bias). This is what I will call "Conformity as separation".

The flip side of the above result is that, when we observe this form of conformity, the decision maker ends up placing a higher probability on the state opposite to both reports. In other words this is a situation in which both experts transmit biased information while the decision maker is more likely not to believe it.

This paper also offers other economic implications. I show that if both experts are of a bad type, making one to value the future more will discipline both of them disclose their signals with a higher probability. This is because, going into the second period, a bad expert cannot afford to lose reputation relative to another bad expert.

Also, I find that if the precision of one expert's signal is reduced, a bad expert is more likely to lie in equilibrium as the decision maker is unable to differentiate between a good expert that got a wrong signal and a bad expert that lies.

Further, consider that a decision maker is able to pick one expert with higher reputation out of a pool. In case this expert turns out to be of a bad type she is more likely to lie in equilibrium. On the other hand, her counterpart is more likely to tell the truth. Even though these are opposing effects, in terms of the decision maker's welfare, I show that the overall effect is positive.

My paper is most closely linked to Morris (2001), in the sense that it features career concerns and misalignment of preferences. Morris shows two key results: a good expert may not disclose her signal when doing so impinges on her future reputation (political correctness) and a bad expert might be induced to tell the truth even though it is against her preference for the same reputational reason. I move away from Morris in three significant ways. First, I allow for the state not to be verifiable. I then involve another expert, but I also consider the possibility commonly observed in economic problems that there is no generally accepted reference point for being biased. This is not hard to imagine. In a recent opinion blog, Krugman (2012) complains about a "new political correctness," whereby politicians seem inclined to pander to the views of conservatives (just the same way as political correctness was originally identified with liberal values). I am therefore careful not to interpret outcomes of conformity as instances of politically correct behavior.

I nonetheless attempt to draw links between Morris' findings in the case of one agent, and also with other forms of behavior seen in the career concerns literature.

If one expert in my model is biased for sure, I show in an extension that the decision maker will ignore her. This can be interpreted as Morris' model with no state verifiability, where in order to relate it to my model I allow for the existence of any prior about the state of the world. I find that the experts report against their potential bias, but more so when the prior is in favor of their bias. Thus we may see greater distortion as the prior varies from 0 to 1, rather than remain at 0.5 as in Morris. At the same time, declaring against her bias is less valuable to the expert as the prior gets closer to this opposing value. This result is similar with the concept of anti-herding developed by Levy (2004) and others. In the case of the anti-herding papers, the asymmetric information about type is on the dimension of ability rather than alignment of preferences.

As we move further from Morris by adding another expert with uncertain opposing bias, we now see that there are settings when the decision maker may be able to infer better information even though biased reports are transmitted to him.

The combined effect of strategic interaction between experts and unverifiability of states thus offers novel insights into different motivations for conforming. The term conformity has been used widely in the literature to refer to situations where individuals comply with a social norm as in Bernheim (1994). Here it refers only to the fact that experts take similar actions.

In the context of the wider literature on career concerns, this paper is related in a behavioral sense to the two strands in which there is asymmetric information about ability or misalignment of preferences. In terms of ability, Levy (2004) shows that managers might 'anti-herd' or excessively contradict public information to distinguish themselves from the rest and increase their reputation. On the other hand, Scharfstein and Stein (1990) explore the behavior of managers when they ignore their own information and may herd on the others' actions for the purpose of being perceived informed. In these papers, the uncertain types are in the dimension of ability, unlike Sobel (1985), and Morris (2001) who construct models where the uncertain type is based on the alignment of preferences with a principal.

Other relevant papers include Prendergast and Stole (1996), where in a dynamic setting analysts initially overweight their information so that they are initially seen as fast learners; later in their career they become conservative. Ottaviani and Sorensen (2001) look for an optimal information transmission by designing a reputation model where experts act in a similar way.

This work is also related to research involving cheap talk with multiple senders in the presence of misalignment of preferences. Some relevant papers include Gilligan and Krehbiel (1989) where a committee composed of two perfectly informed experts with (sure) opposing biases offers advice to a legislator. Morgan and Krishna (2001) allow the experts not to have necessarily opposing biases while the advice is offered simultaneously or sequentially. Austen-Smith (1993) studies a cheap talk model with two imperfectly informed experts. McGee and Yang (2013) look at a decision problem when the experts have complementary information. In these papers the experts are interested only in their current payoffs without an interest in their future career.

The real world applicability of my model could be quite wide ranging. Some examples include:

1. A decision maker takes the advice of two political advisers - one from the left of the political spectrum and the other from the right. Both the advisers could be fair and have preferences aligned with the decision maker - in the sense that their advice is as close as possible to the state of the world, or they could be biased towards their party agenda. The state of the world is not verifiable in this case and once the advice is implemented the initial state of the world based on which the advice was implemented is not observable anymore.

2. An employee and her supervisor report to the head of the organization on how well the employee performed a task. Both of them can evaluate the performance correctly but not with full precision. The honest employee would report her evaluation correctly, while the dishonest employee would exaggerate

her performance. In the same way, the supervisor could be good and report truly his evaluation of the task, or he can be biased with a preference for reporting negatively on the employee. The head of the organization has to compare the reports of both employee and supervisor when taking a decision for promotion but also to update her belief on the type of the two employees for future reference. It is important to note that this game is one of information aggregation and transmission - the experts are similar in every aspect apart from the quality of their signal and their possible biases. This model does not illuminate other aspects such as the influence of power or hierarchy.

3. Financial analysts writing a report on new investment offerings may have biases that lead them to talk up or down the security values. Such biases could arise from the nature of an industry (a green technology firm whose prospects depend on concerted global action against climate change), or from views about economic policy (bonds issued by a state where there is a debate on the sustainability of government expenditures). In both cases, the final outcome may arise over a relatively long period, so a useful way to evaluate the reports of the analysts would be to consider both sides of the story in the context of the existing reputation of the analysts.

4. Further examples could include funding decisions on different research projects in a research lab. Also, when there is a jury decision in a murder trial with no confession but expert testimonies, an expert could make herself available to offer support for either the prosecution or defence teams.

The rest of the paper is organized as follows. In the next section I describe the model. As the game is set over two periods, in section 3 I find and characterize the equilibrium in the last stage game; in section 4 I characterize and show the existence of the equilibrium in the first stage and I provide a specific case with truthtelling in equilibrium; in section 5 I provide numerical solutions and show some comparative static analyses, while in section 6 I provide some potential extensions; section 7 concludes. All proofs that are not in the text appear in the appendix.

2 Model

There are three players in this game: a decision maker D, and two experts (L and R). The game is played over two periods, $t \in \{1, 2\}$.

There is an underlying state of the world x_t which can take values of 0 and 1 with equal probability. The states of the world are drawn independently each time. The decision maker is not able to verify the state of the world in either period. However, each of the experts receive a noisy but informative private signal about the true state of the world each period: $s_t^i \in \{0, 1\}$, where $i \in \{L, R\}$. The signal has precision $p^i = \Pr[s_t^i = x_t | x_t] > \frac{1}{2}$.

The decision maker receives reports l_t and r_t about x_t from L and R respectively. Based on these reports, he takes an action $a_t \in [0, 1]$. His objective is to be as close as possible to the true state of the world, so I set his expected payoff to be: $-\mu_1 E(x_1 - a_1)^2 - \mu_2 E(x_2 - a_2)^2$.

There are two types of experts: 'good' (G) and 'bad' (B) and the decision maker D is uncertain of their type. D's prior probability that R is of type G is $\lambda_1^R \in (0, 1)$ while his prior probability that L is of type G is $\lambda_1^L \in (0, 1)$. $\Lambda_1 \equiv (\lambda_1^L, \lambda_1^R)$.

The good experts have preferences aligned with the decision maker, which is reflected in their payoff structure.

If the expert R is good, her payoff is $-\mu_1^{RG} E\left[(x_1 - a_1)^2 | s_1^R\right] - \mu_2^{RG} E\left[(x_2 - a_2)^2 | s_2^R\right]$.

A good expert L has exactly the same payoff as a good expert R, adjusted however to the signal she observes: $-\mu_1^{LG} E\left[(x_1 - a_1)^2 | s_1^L\right] - \mu_2^{LG} E\left[(x_2 - a_2)^2 | s_2^L\right]$.

The bad experts are biased towards either 1 or 0. If R is bad, she has a higher utility when the action taken by D is closer to 1. Her payoff is $\mu_1^{RB} a_1 + \mu_2^{RB} a_2$, while the bad expert L's payoff is: $-\mu_1^{LB} a_1 - \mu_2^{LB} a_2$, which reflects a bias towards 0.

The experts could value the present different than the future by assigning different weights to current and future payoffs: $\mu_1^{ik} > 0$,and $\mu_2^{ik} > 0$ with $i \in \{L, R\}, k \in$

$\{G, B\}$. These weights reflect different time preferences between experts and allow for situations in which any of the parties involved could value the future payoff more than the current one.

After observing l_1 and r_1, D updates his beliefs on the type of the experts and on the state of the world x_1. The posterior reputations are denoted $\Lambda_2 \equiv \left(\lambda_2^L, \lambda_2^R\right)$ and the belief on the state of the world $\Gamma\left(x_1|l_1, r_1\right) = \Pr\left(x_1|l_1, r_1\right)$. For simplicity of notations I denote the posterior belief that the state of the world is 1 with $\Gamma\left(l_1, r_1\right)$.

If the state of the world were verifiable, the decision maker could update the reputations by comparing the reports of the experts with the realized state. When the state is unknown, the updating is based only on the reports of the experts, keeping in mind their initial reputation.

In the second period ($t = 2$) the game is repeated, with the state of the world x_2 independent of x_1. In this model, everything apart from the type of the experts and their private signals is known by everyone.

2.1 Strategies and Solution Concept

The strategy profile for the players is $\left(\pi_{kt}^R\left(s_t^R\right), \pi_{kt}^L\left(s_t^L\right), a_t\left(l_t, r_t\right)\right)$, where $\pi_{kt}^R\left(s_t^R\right)$ is R's probability of reporting 1 when the signal is s_t^R, $\pi_{kt}^L\left(s_t^L\right)$ is probability of reporting 0 when the signal is s_t^L, and $a_t\left(l_t, r_t\right)$ is the action taken by the decision maker given l_t and r_t. It is important to note that the experts' strategies represent the probability that their report is the same as their potential bias.

Definition 1 *A Perfect Bayesian equilibrium is a strategy profile* $\pi_{kt}^R\left(s_t^R\right)$, $\pi_{kt}^L\left(s_t^L\right)$, $a_t\left(l_t, r_t\right)$ *such that (a) the experts's reports given their signals maximize their respective payoffs given the posterior reputational beliefs, (b) the decision maker's action maximizes his expected payoff given his posterior probability on the state of the world and (c) the posterior probabilities on the type of the expert and the state of the world are derived according to Bayes' rule.*

As the game is set over two periods the equilibrium outcomes will be determined by backward induction. In each stage game I will use the strategy profile without a time subscript for notational ease.

3 The Second Stage - No Reputational Concerns

In the last period R and L enter with reputations λ_2^R and λ_2^L. This is a cheap talk game where the experts' reports do not enter their payoff directly but indirectly through the influence they have on D's belief about the state of the world and consequently through D's action.

As the reports are costless, there always exist equilibria in which the decision maker does not infer anything from these reports so there is no incentive for the senders to send them anyway - this is a common feature of cheap talk games. These types of equilibria are called uninformative or babbling equilibria. By contrast the informative equilibria are those in which some information is transmitted by the experts to the decision maker.

For this particular game I will analyze both babbling and informative equilibria.

Babbling equilibrium is a situation in which each expert independent of her type randomizes with equal probability of reporting 0 or 1. In this case the decision maker will learn nothing from the messages and will continue to believe that the states are equally likely and the action is $a^* = \frac{1}{2}$, independent of the message. The experts also have no incentives to deviate from the uninformative actions.

An informative equilibrium is an equilibrium in which experts' reports are correlated with the state of the world for any l_2, r_2.

Proposition 1 *There exists an informative equilibrium where the decision maker's optimal action is $a_2^*(l_2, r_2) = \Gamma(l_2, r_2)$. The good experts' optimal strategies are $\pi_G^R(1) = 1$, $\pi_G^R(0) = 0$, $\pi_G^L(0) = 1$, and $\pi_G^L(1) = 0$. The bad experts' strategies are $\pi_B^i(s_2^i) = 1$, for any $i \in \{L, R\}$.*

The above equilibrium strategies reflect the fact that in the last period the good experts declare their signals while the bad experts' reports are consistent with their respective biases.

The idea behind this proposition is the fact that in an informative equilibrium both the messages sent by the players carry some information to the decision maker. Essentially, if the decision maker observes 1 from R (the message is informative) he will choose a higher action than if he had observed 0, thus the bad expert R will have a strict incentive to declare 1 while the good expert R will have a strict incentive to truthfully reveal her signal. If the decision maker observes 1 from L, he will increase his action while if he observes 0 he will decrease his action. Thus the bad expert L declares 0 while the good expert L declares her signal.

The optimal action of the decision maker for all possible reports is:

$$a_2^*(l_2, r_2) = \begin{cases} \frac{p^L\left(1-p^R\right)}{A} & \text{if } l_2 = 1, r_2 = 0 \\ \frac{\left(1-\lambda_2^L p^L\right)\left(1-p^R\right)}{1-\lambda_2^L A} & \text{if } l_2 = 0, r_2 = 0 \\ \frac{p^L\left(1-\left(1-p^R\right)\lambda_2^R\right)}{1-\lambda_2^R A} & \text{if } l_2 = 1, r_2 = 1 \\ \frac{\left(1-p^L\lambda_2^L\right)\left(1-\left(1-p^R\right)\lambda_2^R\right)}{2-\lambda_2^R-\lambda_2^L+\lambda_2^R\lambda_2^L A} & \text{if } l_2 = 0, r_2 = 1 \end{cases}$$

where $A \equiv \left(1-p^L\right)p^R + p^L\left(1-p^R\right)$ could be seen as an average precision of the players. Please see in the Appendix how the decision maker's optimal action is obtained.

3.1 Second Period Reputations - Some Insights

There are some straightforward results that shed some light on how experts' actions complement each other when career concerns are no longer present.

The experts enter the second period with reputations $\Lambda_2 \equiv \left(\lambda_2^L, \lambda_2^R\right)$, which they built in the first period. The effect of their reputations on the decision maker's action in the second period is utterly important to them as it determines an optimal course of action for both the experts in the first period.

First, if both the experts report 0 the decision maker's action $a_2^*(l_2 = 0, r_2 = 0)$

decreases with the reputation of L while it does not depend on the reputation of R.

This is due to the fact that the posterior belief that the true state is 0 increases with the reputation of L as biased reports are less likely to arrive from experts with better reputation. As R has no incentive for wrongly reporting 0 her reputation will not be at play in this case. In other words, R is good with probability 1.

Similarly, when both experts report 1, the same analysis as above stands. The action of the decision maker $a_2^*(l_2 = 1, r_2 = 1)$ is increasing in the reputation of R as a 1 report is more likely to be the truth, and it does not depend on the reputation of L.

Second, if the message of L is 0 and the message of R is 1 the optimal action $a_2^*(l_2 = 0, r_2 = 1)$ is increasing in the reputation of R and decreasing in the reputation of L. This is due to the fact that if L has a good reputation a 0 report is more likely to be the truth while if R has a good reputation a 1 report is more likely to be the truth.

A different question worth analyzing is how the individual reputational change affects experts' expected payoffs in the second period.

For a good type R the value of reputation acquired from the first period is her ex-ante expected payoff $-E\left[(x_2 - a_2^*)^2 | \Lambda_2\right]$ and it is calculated as follows:

$$v_R^G(\Lambda_2) = -\sum_{x_2}\sum_m\sum_n \Pr(x_2) \Pr\left[s_2^R = n | x_2\right] \Pr\left[l_2 = m | x_2\right] (x_2 - a_2(l_2 = m, r_2 = n))^2.$$

In the above expression $x_2, m, n \in \{0, 1\}$. Similarly for a good type L the value of her reputation is her ex-ante expected payoff in the second period.[1]

These expressions take into account that if the state of the world is drawn independently each time, a good expert entering the second period could face either a state 0 or 1 with equal probability while there is also uncertainty on the signal received by both experts given the state.

The bad experts however are biased towards 0 or 1 respectively, so irrespective of their signals, their expected payoffs feature these biases. As a result R's reputational value is

[1] $v_L^G = -\sum_{x_2}\sum_m\sum_n \Pr(x_2) \Pr\left[s_2^L = m | x_2\right] \Pr(r_2 = n | x_2) (x_2 - a_2(l_2 = m, r_2 = n))^2$

$v_R^B(\Lambda_2) = \sum_{x_2}\sum_m \Pr(x_2)\Pr[l_2 = m|x_2]\, a_2(l_2 = m, r_2 = 1).$

Similarly, L 's reputational value accounts for a bias towards 0.[2]

Proposition 2 *The second period ex-ante expected pay-off of an expert increases in her reputation irrespective of her type. It also (weakly) increases in the reputation of the counterpart irrespective of type.*

For proof and some discussions please see the Appendix.

The above proposition gives us a very important result which says that the experts, once they care about their future, try to acquire a good reputation in the first period.

The last part of the proposition accounts for the fact that a good expert prefers to be as close as possible to the state of the world so she would prefer to be paired with another good expert. The bad expert also prefers to have a counterpart of a good type as a bad counterpart will always report in the opposite direction of her preference.

Equilibrium Selection: In the first period I showed that there exist a babbling equilibrium in which no information is transmitted to the decision maker and an informative equilibrium in which the good experts disclose their signals while the bad ones the biases. Note that this is only an existence result and it does not addresses questions of uniqueness. I focus next on the described informative equilibrium as the babbling equilibrium in the second period does not induce reputational concerns in the first period.

4 First Stage Game

The first period game is similar with the second period game with the exception that the experts (R and L) have reputational concerns for the second period of the game. The prior probability of the experts being good is $\Lambda_1 \equiv \left(\lambda_1^L, \lambda_1^R\right)$.

[2] $v_B^L = -\sum_{x_2}\sum_n \Pr(x_2)\Pr[r_2 = n|x_2]\, a_2(l_2 = 0, r_2 = n)$

Experts' total payoff functions account for both current and future payoffs, taking into account their relative time preference. For notational ease I will represent the total payoffs in terms of the relative weight of the first period payoff i.e. $\mu^{ik} \equiv \frac{\mu_1^{ik}}{\mu_2^{ik}}$ with $i \in \{L, R\}$, $k \in \{G, B\}$. $\boldsymbol{\mu}^k$ represents the vector of relative weights of experts of type $k : \left(\mu^{Lk}, \mu^{Rk}\right)$.

Experts' total payoff is the sum of the first stage payoff weighted by the appropriate time preference and the second stage expected payoff (which I called in the previous section experts' value of reputation).

The good experts' total payoff is $\mu^{iG} u_G^i (l_1, r_1, s_1^i) + v_G^k (\Lambda_2)$ where $k \in \{G, B\}$. Their current payoff $u_G^i (l_1, r_1, s_1^i)$ is $-E \left(x_1 - a_1^* (l_1, r_1) | s_1^i\right)^2$ and captures the objective of the good experts to take an action as close as possible to the state of the world.[3]

The bad experts' total payoffs account for their preference for different states. As a result an R expert of bad type has a total payoff of $\mu^{RB} u_B^R (l_1, r_1) + v_B^R (\Lambda_2)$ where $u_B^R (l_1, r_1) = a_1 (l_1, r_1)$ while an L expert of bad type has a total payoff $\mu^{LB} u_B^L (l_1, r_1) + v_B^L (\Lambda_2)$ where $u_B^L (l_1, r_1) = -a_1 (l_1, r_1)$.

4.1 Reputation Formation

The experts enter the first stage game with some initial priors on their reputation Λ_1. After they send their reports, the decision maker updates his belief on their types. Λ_2 is the vector of posterior reputations.

In determining these posterior reputations the interaction between the experts' actions is captured by $\phi_k^L (l|r)$ and $\phi_k^R (r|l)$ which are the probabilities that a type $k \in \{G, B\}$ expert whether L or R sends a particular report given the counterpart's report. These probabilities take into account the fact that the state of the world is not verifiable.

[3] $u_G^R \left(l_1, r_1, s_1^R = 1\right) = -E \left(x_1 - a_1^* (l_1, r_1) | s_1^R = 1\right)^2 = -\frac{1}{2} p^R + p^R a_1 (l_1, r_1) - \frac{1}{2} a_1 (l_1, r_1)^2$
$u_G^L \left(l_1, r_1, s_1^L = 0\right) = -E \left(x_1 - a_1^* (l_1, r_1) | s_1^L = 0\right)^2 = -\frac{1}{2}(1-p^L) + (1-p^L) a_1 (l_1, r_1) - \frac{1}{2} a_1 (l_1, r_1)^2$

$$\phi_k^R(r_1|l_1) = \sum_{x_1=\{0,1\}} \phi_k^R(r_1|x_1)\Pr(x_1|l_1)$$

where $\phi_k^R(1|x_1) = \left[p^R \pi_k^R(x_1) + \left(1-p^R\right)\left(\pi_k^R(1-x_1)\right)\right]$.[4] Note that $\phi_k^R(r_1|x_1)$ denotes the probability that a type k expert R sends message r when the state is x_1. In the above expression $\Pr(x_1|l_1)$ represents the state probability given L's report and is calculated by Bayes' rule.[5]

The probability that a $k-$ type L sends message l when R sends message r is calculated in a similar fashion as $\phi_k^R(r_1|l_1)$.[6]

Thus, the posterior probability of an expert R to be of a good type is:

$$\lambda_2^R(l_1, r_1) = \frac{\lambda_1^R \phi_G^R(r_1|l_1)}{\lambda_1^R \phi_G^R(r_1|l_1) + \left(1-\lambda_1^R\right)\phi_B^R(r_1|l_1)}$$

while the posterior probability of an expert L to be of a good type is:

$$\lambda_2^L(l_1, r_1) = \frac{\lambda_1^L \phi_G^L(l_1|r_1)}{\lambda_1^L \phi_G^L(l_1|r_1) + \left(1-\lambda_1^L\right)\phi_B^L(l_1|r_1)}$$

Remark 1 *Experts' reputations are unaffected in the first period* ($\lambda_2^R(l_1, r_1) = \lambda_1^R$ *and* $\lambda_2^L(l_1, r_1) = \lambda_1^L$) *whenever* $\phi_G^R(r_1|l_1) = \phi_B^R(r_1|l_1)$ *and* $\phi_G^L(l_1|r_1) = \phi_B^L(l_1|r_1)$ *respectively.*

This occurs when either both L and R babble or report truthfully on their message. If so, either no information is disclosed to the decision maker or there is full disclosure.

Remark 2 *The posterior reputation* $\lambda_2^R(l_1, r_1)$ *decreases with* $\frac{\phi_B^R(r_1|l_1)}{\phi_G^R(r_1|l_1)}$. *Given L's report the higher the message that the bad type R expert is likely to send relative to the good type, the lower R's posterior reputation is after the reports are seen.*

[4] $\phi_k^L(l_1 = 1|x_1) = \left[p^L\left(1-\pi_k^L(x_1)\right) + \left(1-p^L\right)\left(1-\pi_k^L(1-x_1)\right)\right]$

[5] $\Pr(x_1|l_1) = \frac{\Pr(x_i)\Pr(l_1|x_i)}{\sum_{x_1=\{0,1\}} \Pr(x_i)\Pr(l_1|x_i)}$

[6] $\phi_k^L(l_1|r_1) = \sum_{x_1=\{0,1\}} \phi_k^L(l_1|x_1)\Pr(x_1|r_1)$

4.1.1 Posterior Belief on the State of the World

The decision maker updates not only her belief on the type of the experts but also on the state of the world. The posterior belief that the state is 1 when the messages are (l_1, r_1) is:

$$\Gamma(l_1, r_1) = \frac{\Pr(l_1, r_1|1)}{\Pr(l_1, r_1|1) + \Pr(l_1, r_1|0)}$$

with

$$\begin{aligned}\Pr(l_1, r_1|x_1) &= \left[\lambda_1^R \phi_G^R(r_1|x_1) + \left(1 - \lambda_1^R\right)\phi_B^R(r_1|x_1)\right] \\ &\quad \left[\lambda_1^L \phi_G^L(l_1|x_1) + \left(1 - \lambda_1^L\right)\phi_B^l(l_1|x_1)\right]\end{aligned}$$

for $x_1 \in \{0, 1\}$.

Remark 3 *If both experts babble, or $\phi_G^R(r_1|x) = \phi_B^R(r_1|x)$ and $\phi_G^L(l_1|x) = \phi_B^L(l_1|x)$, then the posterior belief is that the states of the world are equally likely.*

4.2 First Stage Equilibrium

Similar with the second period game, D does not observe the state and as a result his optimal action is his posterior belief about the state of the world.

$$a_1^*(l_1, r_1) = \Gamma(l_1, r_1)$$

The expected payoff of a 'good' expert R when L sends message l_1 and D believes that the state is 1 with probability $\Gamma(l_1, r_1)$ is $u_G^R(l_1, r_1)$ while the expected payoff of a 'bad' expert R is $u_B^R(\Gamma) = u_B^R(a_1^*) = a_1^*(l_1, r_1)$. Similarly $u_G^L(\Gamma)$ is the expected payoff of a good expert while $u_B^L(l_1, r_1) = u_B^L(a_1^*) = -a_1^*(l_1, r_1)$ is the expected payoff of a bad expert.

Again we could identify two types of strategies in this first period game: babbling strategies when no information is transmitted to the decision maker and informative strategies.

Definition 2 $\left(\pi_k^R, \pi_k^L, \Gamma, \lambda_2^R, \lambda_2^L\right)$ *is a babbling strategy profile if for* $c_1, c_2 \in [0,1]$: (1) $\pi_G^R(0) = \pi_B^R(0) = \pi_G^R(1) = \pi_B^R(1) = c_1$; (2) $\pi_G^L(0) = \pi_B^L(0) = \pi_G^L(1) = \pi_B^L(1) = c_2$; (3) $\lambda_2^R(l_1, r_1) = \lambda_1^R$, $\lambda_2^L(l_1, r_1) = \lambda_1^L$ *and* $\Gamma(l_1, r_1) = \frac{1}{2}$ *for any* l_1, r_1.

As this is a cheap talk game there always exists a babbling equilibrium and hence the following claim:

Claim 4 *Every babbling strategy profile is an equilibrium.*

This claim is true since if everyone babbles than the decision maker does not infer anything on the state of the world or the type of the experts, so the posteriors will be equal with the priors and no one will have a unilateral incentive to deviate.

Proposition 3 *An informative equilibrium* $\left(\pi_k^R, \pi_k^L, \Gamma, \lambda_2^R, \lambda_2^L\right)$ *satisfies the following properties:*

1. *When the good expert* R *observes signal* $s_1^R = 0$*, she always announces* 0 - $\pi_G^R(0) = 0$*; truthtelling is always optimal for* R *when her signal is* 0.

2. *When the good expert* L *observes signal* $s_1^L = 1$*, she always announces* 1 - $\pi_G^L(1) = 0$*; truthtelling is always optimal for* L *when her signal is* 1.

For a proof, please see the Appendix. This result tells us that the good agents always tell their signal when there is no reputational incentive for not doing so. Intuitively, if R's signal is 0 she never reports 1 as by doing so she only damages her reputation while her current gain decreases. The same argument applies to the good L agent who never reports 0 when her signal is 1.

In the first period the experts not only transmit information to the decision maker about the state of the world but also acquire reputation which will impact the decision maker's second period action. The following proposition summarizes the reputational incentives in equilibrium.

Proposition 4 *In equilibrium the reputations are such that :*

1. $\lambda_2^R(l_1,0) \geq \lambda_2^R(l_1,1)$ *for* $l_1 \in \{0,1\}$ *with one strict inequality;*

2. $\lambda_2^L(1,r_1) \geq \lambda_2^R(0,r_1)$ *for* $r_1 \in \{0,1\}$ *with one strict inequality;*

3. $\lambda_2^R(1,r_1) \geq \lambda_2^R(0,r_1)$ *iff* $\lambda_1^L \geq \bar{\lambda}_1^L$. *If instead,* $\lambda_1^L < \bar{\lambda}_1^L$ *then* $\lambda_2^R(1,r_1) < \lambda_2^R(0,r_1)$ *where:*
$$\bar{\lambda}_1^L = \frac{\pi_B^L(0) - \pi_B^L(1)}{\pi_G^L(0) + \pi_B^L(1) - \pi_B^L(0)}$$

4. $\lambda_2^L(l_1,0) \geq \lambda_2^L(l_1,1)$ *iff* $\lambda_1^R \geq \bar{\lambda}_1^R$ *If instead* $\lambda_1^R < \bar{\lambda}_1^R$, *then* $\lambda_2^L(l_1,0) < \lambda_2^L(l_1,1)$ *where*
$$\bar{\lambda}_1^R = \frac{\pi_B^R(1) - \pi_B^R(0)}{\pi_G^R(1) + \pi_B^R(0) - \pi_B^R(1)}$$

Reputational concerns make the experts to announce against their signal just for the purpose of showing that they are not biased. In particular, the R expert has always a strict incentive to announce 0 irrespective of the announcement of the L expert while the L expert has an incentive of announcing 1. This is an act of disavowing one's perceived bias.

However, there are also other incentives at play; these refer to the reports' effect on the experts' reputation in relation to the counterparts' reports. Morris (2001) showed that if the states were verifiable it would be more important for a good R expert to announce 0 when the state is 1 rather than 0. The reason for this is that by reporting 0 when the state is 1 the expert shows that she is of a good but it is possible that she did not observed the true state of the world as her signal is not fully precise.

When the states are not verifiable, both the reports have to be compared for the purpose of reputation formation. Now, the initial reputations of the experts plays a major role in determining the value of disavowing one's bias when the counterpart changes her report from 0 to 1. The decision maker now updates her beliefs on the state of the world and experts' type by looking at R's report in relation with a likelihood of a state of the world. This likelihood is captured in L's report. The more likely is L to misreport the state of the world the more L is perceived to be biased. As a result we could have situations in which a 0 report from R could be more important for reputation formation both when L announces 0 rather than 1 or the opposite.

Points 1 and 2 of the proposition can be reduced to the following: A bad expert R reports 1 more often than a good expert R as $\pi_B^R(1) \geq \pi_G^R(1)$ and $\pi_B^R(0) \geq \pi_G^R(0)$ with one strict inequality.[7] Similarly, a bad expert L reports 0 more often than a good expert L as $\pi_B^L(0) \geq \pi_G^L(0)$ and $\pi_B^L(1) \geq \pi_G^L(1)$ with one strict inequality.

The expression $\lambda_1^L \geq \bar{\lambda}_1^L$ in point 3 is equivalent with $\Pr(l_1 = 1 | x_1 = 1) \geq \Pr(l_1 = 1 | x_1 = 0)$, which means that L has a high probability of conveying the true state of the world.

4.2.1 Equilibrium Reputations - Further Insights

One of the major results conveyed in Morris (2001) is that when states are verifiable the incentive to report against one's bias depends on the state of the world. In particular, he finds that the reputation coming from announcing 0 is greater for an expert biased towards 1, when the state is 1 rather than 0. In this study however, as the states are not verifiable the decision maker compares both experts' reports and then makes a decision; in the next subsections I discuss how the equilibrium reputation of one expert changes with the counterpart's report.

4.2.2 Conforming as a form of separation

When the state of the world is not verifiable an expert decides what to report based on her counterpart's initial reputation. We see now that taking the same action does not necessarily mean that the experts agree with each other - it is just an information transmission mechanism regarding their level of trustworthiness. This follows from the following corollary (from Proposition 4, point 3, when $r_1 = 0$):

[7] This is due to $\frac{\phi_B^R(1|1)}{\phi_G^R(1|1)} \geq \frac{\phi_B^R(0|1)}{\phi_G^R(0|1)} = \frac{1-\phi_B^R(1|1)}{1-\phi_G^R(1|1)} \Leftrightarrow \phi_B^R(1|1) \geq \phi_G^R(1|1)$ and $\frac{\phi_B^R(1|0)}{\phi_G^R(1|0)} \geq \frac{\phi_B^R(0|0)}{\phi_G^R(0|0)} = \frac{1-\phi_B^R(1|0)}{1-\phi_G^R(1|0)} \Leftrightarrow \phi_B^R(1|0) \geq \phi_G^R(1|0)$

Corollary 1 $\lambda_2^R(1,0) < \lambda_2^R(0,0)$ *iff* $\Pr(l_1 = 0|x_1 = 0) < \Pr(l_1 = 0|x_1 = 1)$.

In particular when R says 0, D looks at L's report and forms beliefs on both experts' type and the state of the world. If the chance is high that L misreports the state a 0 report is more likely to come from a state 1 than from a state 0. Then, in contrast with the verifiable case it is more important for R to say 0 when L says 0 than when L says 1. This arises from the effect on R's future reputation from signaling: "Because you believe that my counterpart is most likely to say 0 wrongly due to her bias, my saying 0 should show you that I am not biased. Since I am as far away as possible from 1, which you figured out is probably closer to the true state of the world, it looks I am acting in good faith but I got an imprecise signal on the state, so I am of a good type." What might look like agreeing with the counterpart in fact is not necessarily true; "I do as you do so I can differentiate myself from you."

In this situation we still have that the experts report against their potential biases for reputational reasons, however the experts might also report against their bias in order to differentiate themselves from the highly likely biased counterpart. This is what I call "conformity as separation". Note that this effect is due to the combined effect of strategic interaction and unverifiability of the states. The unverifiability of state makes an expert resort to the counterpart's report in order to signal her type as there is no fixed reference point to compare against.

As an example think of a policy maker who is asking the opinion of two experts - one hired by the left party and one hired by the right party. When the initial situation is such that the left wing expert has a low reputation and she is likely to misreport on the state of the world the best decision of the right wing expert is to agree with the left expert not because she cares about left policies but as a form of building her reputation of being a fair expert.

A historical example relevant to this result is the 1972 Nixon's visit to China. At that time Nixon was perceived as having a strong anti-communist stance but by taking this action he gained popularity with the electorate. Note that Nixon took a great risk in visiting China as the American population was wary at the time of a possible diplomatic relationship with the People Republic of China (PRC). This was especially triggered by Taiwan being removed from the United Nations in favor of PRC against American opposition.

4.2.3 Sending different messages as a form of separation

Interpreting the model in the context of Morris (2001), if the states are verifiable the right biased expert's reputation coming from saying 1 when the state is 1 is always higher than when the state is 0; this is due to the fact that the risk of message 1 coming from a biased expert is lower in state 1.

However, when the states are not verifiable and the probability that L misreports the state is high, D believes that when he sees 0 from L it is more likely the state is 1 rather than 0. For R, this means that reporting 1 when L says 0 is much better than reporting 1 when L also says 1. First of all, when L says 0, R's probability of being biased is small given what D knows about L. On the other hand, when L says 1, D may place a greater probability on the state being 0, hence believing that R is likely to be biased.

Corollary 2 $\lambda_2^R(0,1) > \lambda_2^R(1,1)$ *iff* $\Pr(l_1 = 1|x_1 = 1) < \Pr(l_1 = 1|x_1 = 0)$

4.2.4 Counterpart with high reputation: no conformity as separation

If L has a high initial reputation, then $\lambda_2^R(1, r_1) \geq \lambda_2^R(0, r_1)$ where r_1 could be 1 or 0. This is due to an argument similar to Morris (2001): the reputation of an expert R that reports 1 is higher when the counterpart's report is 1 rather than 0 as counterpart's report is likely to reflect the true state of the world. In building up this reputation D realizes that the risk of a message 1 coming from a biased expert is lower if 1 is likely to be the true state (as per L's report).

Also an expert R gains a higher reputation from reporting 0 when L reports 1, compared to when L reports 0. The logic for this is that when 1 is likely to be the true state (captured in L's report) R makes a stronger argument of being unbiased by declaring 0. The case when both of the reports are 0 does not convey extra information about the biasedness of R. In this case as the L has a high reputation there is no conformity as separation effect.

A further point worth conveying is that if the states would be fully verifiable the model would collapse to a situation in which the two experts will not strategically interact. A possible extension of this model would be to have partial verifiability of the true states. A model of this type would nest both the current model and a model with two experts without strategic communication.

4.3 Informative Equilibrium with Truthtelling

A question worth analyzing is whether this game supports a *full truthtelling equilibrium* where both experts, irrespective of their type, report their signal i.e. an equilibrium where experts' strategies are: $\pi_G^R(1) = 1$, $\pi_G^R(0) = 0$; $\pi_B^R(1) = 1$, $\pi_B^R(0) = 0$; $\pi_G^L(1) = 0$, $\pi_G^L(0) = 1$; and $\pi_B^L(1) = 0$, $\pi_B^L(0) = 0$. However, it is easy to see that there does not exist such an equilibrium as the bad experts have incentives to deviate to their biased preferred action.

Claim 5 *There is no informative equilibrium with both R and L following full truthtelling strategies.*

This is due to the fact that if such an equilibrium exists the posterior reputations are equal with the priors. But this implies that there is no reputational cost for any of the bad experts of announcing their biases. As we are looking at an informative equilibrium, we know that the decision maker takes an action positively correlated with the reports. Thus regardless of their signals the bad experts always report their biases. But this is a contradiction to truthtelling. So, there is no full truthtelling equilibrium.

Let's look next under which conditions there is an *equilibrium in which the good experts report truthfully* on their signal.

Proposition 5 *For any* $\lambda_1^L \in (0,1)$ *and* $\lambda_1^R \in (0,1)$ *there exist* $\bar{\mu}^{RG}, \bar{\mu}^{LG} \in (0,1)$ *such that if* $\mu^{RG}\left(\Lambda_1, \boldsymbol{\mu}^B\right) > \bar{\mu}^{RG}$ *and* $\mu^{LG}\left(\Lambda_1, \boldsymbol{\mu}^B\right) > \bar{\mu}^{LG}$ *there exists a unique truthtelling equilibrium. The good R and L report their signal with probability* 1*, while the bad experts report their signal only when their signals coincide with their bias:* $\pi_B^R(1) = 1$ *and* $\pi_B^L(0) = 1$ *and* $\pi_B^R(0) \in (0,1]$ *and* $\pi_B^L(1) \in (0,1]$.

In order to prove this proposition we assume first that this equilibrium exists. In any informative equilibrium the bad experts report their biases more often than the good experts, thus the bad experts have to tell the truth when the signal received is their respective bias (considering that the good experts always tell the truth). However, when their signal is opposite their bias I compare their current benefit from lying with the reputation cost from telling the truth for determining their optimal strategies.

The equilibrium is unique due to the fact that bad experts' *net* current benefit from lying is a strictly decreasing function in the probability of them lying for all possible strategies of the counterpart. Intuitively, this comes from the fact that in the first period the decision maker is more likely to ignore a piece of advice highly tainted by the possibility of bias which implies a decreasing current benefit from lying; at the same time the future reputation costs are the highest when probability of lying is close to 1 as a report opposite the potential bias could come only from a good expert. Also, I account for the fact that expert's value of reputation increases with the first period counterpart's reputation. The monotonic properties of experts' net current benefit of lying imply that there is a unique strategy for any fixed counterpart strategy. Furthermore, this translates into unique equilibrium strategies for both bad experts.

Once I find the equilibrium strategies of the bad experts I verify when the good experts tell the truth in equilibrium. For having a truth-telling equilibrium one needs to look for appropriate time preference weights which will make the good experts tell the truth with probability 1. Please look in the appendix for a more detailed proof.

This particular equilibrium characterizes a situation in which the good experts follow truth-telling strategies as they do not care too much about their future payoffs. Further more, the bad experts semi-pool on the actions of the good experts as they would like to be perceived as good in the future. As a result in equilibrium the bad experts tell the truth with probability 1 when their signals are identical to their biases and randomize between telling the truth and lying when their signal is opposite their bias.

I also identify a new channel through which experts disavow their perceived bias: strategic interaction. One result of this is that the inclination of a bad expert to declare her signal is higher when her counterpart has the same report. This is true when the counterpart has a reputation below a threshold and is thus also likely to be bad. This is the conformity as separation effect described earlier applied to the actions of a bad expert. Table 1 summarizes conformity as separation for a bad R expert when $s_1^R = 0$:

$L:$	$l_1 = 1$		$l_1 = 0$
$R:$	$\lambda_2^R(1,0)$	$<$	$\lambda_2^R(0,0)$

Table 1: Conformity as separation for a bad R expert

Result 1 *In equilibrium conformity as separation reflects the fact that a bad R tells 0 which is the truth while L tells 0 (likely to be a biased report).*

One should also realize that the bad experts still prefer to report their bias more often than the good experts due to their different payoff functions so the truthtelling equilibrium is only a semi-pooling equilibrium between good and bad experts of L or R appartenance.

In this equilibrium the good experts tell the truth. However a question worth analyzing is what happens when the good experts' career concerns are getting more important. In particular I ask what is the effect of good experts' career concerns on the equilibrium strategies? The next section analyses the case where both good and bad experts may distort their reports for reputational reasons.

4.4 General Case Equilibrium

When the good experts' career concerns start becoming important they start to develop incentives to distort their reports. This is the case when the good experts' signal is their potential bias and the future benefit from lying is higher than the current benefit from telling the truth. In the career concerns literature this effect was described as political correctness. In this game however the experts could be biased in two opposing directions so both experts could disregard their signals in order to disavow their perceived bias. Going back to the political economy example we could have both liberal political correctness and conservative political correctness. The bad experts will still tell the truth with a positive probability for reputation building reasons.

In the next proposition I look at all possible equilibria: truthtelling equilibrium (analyzed previously), informative equilibrium when the experts (irrespective of their type) disavow their perceived biases with positive probability and non-informative equilibrium when no information is transmitted to the decision maker.

Proposition 6 *For any* $\lambda_1^L \in (0,1)$ *and* $\lambda_1^R \in (0,1)$ *there exist* $\bar{\mu}^{RG}, \bar{\mu}^{LG}, \underline{\mu}^{RG}, \underline{\mu}^{LG} \in (0,1)$ *such that:*

1. *if* $\mu^{RG}\left(\Lambda_1, \boldsymbol{\mu}^B\right) > \bar{\mu}^{RG}$ *and* $\mu^{LG}\left(\Lambda_1, \boldsymbol{\mu}^B\right) > \bar{\mu}^{LG}$ *there exist a truthtelling equilibrium;*

2. *if* $\underline{\mu}^{RG} < \mu^{RG}\left(\Lambda_1, \boldsymbol{\mu}^B\right) \leq \bar{\mu}^{RG}$ *or/and* $\underline{\mu}^{RG} < \mu^{LG}\left(\Lambda_1, \boldsymbol{\mu}^B\right) \leq \bar{\mu}^{LG}$ *there exists an informative equilibrium while;*

3. *if* $\mu_G^R\left(\Lambda_1, \boldsymbol{\mu}^B\right) \leq \underline{\mu}_G^R$ *or/and* $\mu_G^L\left(\Lambda_1, \boldsymbol{\mu}^B\right) \leq \underline{\mu}^{LG}$ *the equilibria of the games are non-informative.*

In an informative equilibrium, experts' tendency to report against their perceived biases is reflected in the fact that both of them could report either 0 or 1 with some positive probability. Furthermore, if one expert has a low reputation we observe the conformity as separation effect. This effect applied to the good expert R's actions when $s_1^R = 1$ is summarized in Table 2.

$L:$	$l_1 = 1$		$l_1 = 0$
$R:$	$\lambda_2^R(1,0)$	$<$	$\lambda_2^R(0,0)$

Table 2: Conformity as separation in the case of good experts

Result 2 *In equilibrium conformity as separation reflects a situation in which the good expert R reports 0 which is a biased report and L reports 0 - likely to be a biased report as well.*

5 Robustness Checks

This model allows for some special cases which could assist with further economic interpretations. In the next subsections I show that the conformity as separation effect is not only specific to this environment.

In the model I do not allow the experts to have initial reputations either 0 or 1 as this could trigger the decision maker not to consider both experts' reports. So the first question I would like to answer is what happens at these reputational limits.

In the first case I look at a situation in which one of the two experts is biased with probability 1. As the decision maker does not take into account the report of the biased expert, this reduces to a one potentially biased expert model without state verifiability. In this environment I show that the expert still prefers to report against her potential bias for reputational reasons. Moreover if I allow for the state priors to be different, the conformity as separation (with a slight different interpretation - a 0 R's report is compared with the prior on the state 0) is still preserved.

The second case is a situation where one expert is good with probability 1 while the other is still potentially biased.

5.1 Political Correctness as Anti-herding

If the decision maker knows L is biased with probability 1, the model reduces to one expert potentially biased toward 1 giving a report to D. The state of the world is $x \in \{0,1\}$ with $\Pr(x_1 = 0) = \tau$ and $\Pr(x_1 = 1) = 1 - \tau$, $\tau \in (0,1)$.

Similar with the original model D is not able to verify the state of the world. Expert's posterior reputation and the posterior belief on state are obtained by Bayesian updating given only the report provided by the expert. There is no comparison with a counterpart or a state.

An expert R of type $k \in (G,B)$ reports r_1 with probability $\phi_k^R(r_1)$. This probability takes into account the fact that the state of the world could be either 0 or 1. $\phi_k^R(r_1) = \phi_k^R(r_1|x=1)\Pr(x_1=1) + \phi_k^R(r_1|x_1=0)\Pr(x_1=0)$.

R's posterior reputation is:

$$\lambda_2^R(r_1) = \frac{\lambda_1^R \phi_G^R(r_1)}{\lambda_1^R \phi_G^R(r_1) + \left(1-\lambda_1^R\right)\phi_B^R(r_1)}$$

while posterior probability that the state of the world is 1 is:

$$\Gamma(r_1) = \frac{\Pr(r_1|1)\Pr(x=1)}{\Pr(r_1|1)\Pr(x=1) + \Pr(r_1|0)\Pr(x=0)}$$

In this extension the first result is that if a good expert gets a signal opposite her bias, she will report it truthfully. The logic behind this is the fact that if $s_1^R = 0$ there is no benefit from lying for a good R. There are also incentives to report against one's bias for reputational reasons. Furthermore, R's reporting against her potential bias decreases with the probability that the state is 0.

Proposition 7 *In equilibrium the reputations are such that* $\lambda_2^R(0) > \lambda_2^R(1)$ *and* $\frac{d\lambda_2^R(0)}{d\tau} < 0$.

The conformity as separation effect in this case is captured by $\frac{d\lambda_2^R(0)}{d\tau} < 0$. In particular the intensity of declaring 0 for the purpose of disavowing one's bias decreases with probability that the state of the world is 0. Basically, at low levels of τ, by doing 0 R says: "because the state is more likely to be 1 rather than 0, I report 0 and thus agree with a low prior on state 0, to show the decision maker that I am not biased as I am as far as possible from 1 which has a high probability of being the true state."

Since conformity as separation effect translates into increased probability of telling declaring against the potential bias I could make a more clear connection of my model with the herding literature. In particular, R is more likely to report 0 if the prior on the 0 state is low. This means that the expert contradicts public information by reporting 0 when the the prior on state 1 is high. This would be equivalent with the anti-herding idea developed by Levy (2004) when careerist experts contradict public information, applied to experts with possible misalignment of preferences.

This special case extends Morris (2001) by allowing the states of the world to be unverifiable. Morris' political correctness result is built however on the fact that the decision maker compares expert's report with a realized state when building expert's reputation. In this model as the state of the world of the world is uncertain, this comparison is not viable anymore and the decision maker has to make use of the public view on the state of the world. Similar to Morris, I find that the experts report against their possible bias for reputational reasons. However, in this uncertain environment there is a further incentive in place as shown in the proposition above; in order to build their reputation experts report also against the public prior on the state; furthermore declaring against one's possible bias is more intense when the public thinks the opposite. So this model depicts political correctness as an anti-herding result.

Morris's reasoning for the political correctness is based however on Loury (1994) who develops a syllogism for political correctness as a reputational distortion due to the inherent inclination of members of a community to adhere to communal values. People declare as their fellows as to not offend the community and remain in good standards with their peers. Failing to do so results in the "odds that the speaker is not in fact faithful to communal values as estimated by a listener otherwise uniformed about his views to increase." So, Morris adheres at least conceptually to Loury's argument, that political correctness is conformity to social norms. In this particular case however contrary to Morris I show that people act in a political correct manner not only to disavow their individual bias but also to show that they hold different views than their community. Thus political correctness is not herding but anti-herding.

5.2 Expert L is Unbiased with Probability $1 : \lambda_1^L = 1$

If L is unbiased with probability 1 but still is not fully informed on the state of the world, the decision maker will still take both reports into consideration when taking an action. However as L is a good type expert she will report her signal with probability 1.

R however still builds up her reputation in the first period so she has incentives to report against her potential bias In this case, the conformity as separation effect disappears as the likelihood that L misreports on the state is low $\left(p^L \geq \frac{1}{2}\right)$.

Proposition 8 *In equilibrium the reputations are such that* $\lambda_2^R(l_1, 0) \geq \lambda_2^R(l_1, 1)$ *for* $l_1 \in \{0, 1\}$ *with one strict inequality and* $\lambda_2^R(l_1 = 1, r_1) \geq \lambda_2^R(l_1 = 0, r_1)$ *for any* $r_1 \in \{0, 1\}$ *fixed.*

The second part is triggered by the fact that once L's signal precision is informative and L unbiased $\Pr(l_1 = x_1 | x_1) \geq \Pr(l_1 \neq x_1 | x_1)$ always.

Other extensions are also possible but I leave them to future work and I discuss them briefly in the conclusion below.

6 Conclusion and Future Work

This paper lies at the congruence of three bodies of research: the career concerns literature with uncertain misalignment of preferences between a decision maker and agent as in Morris (2001), the career concerns literature with uncertain level of expertise as in Prendergast and Stole (1996), Ottaviani and Sorensen (2001), Levy (2004), Dasgupta and Prat (2001) and cheap talk with multiple experts as in Austen-Smith (1993), Morgan and Krishna (1999) or McGee and Yang (2009).

This study is a career concerns model where two experts have certain ability to see the world but also uncertain misalignment of preferences with a decision maker.

In this environment in an informative equilibrium the experts have a tendency to report in opposite direction of their perceived possible bias for reputational reason. Also they strategically respond to the possible misreporting by the counterpart. As a result, when there is high probability that one expert is biased the other expert tends to offer the same report to show that she is of a good type but she committed an error (considering that she could only be biased in the other direction). I call this *conformity as separation*. This effect is due to the fact that the decision maker cannot verify the veracity of a speaker and furthermore there is no fixed reference point based on which he can infer anything on the type of the experts. As a result the decision maker relies on the other expert's report (which could be biased as well) in order to deduce her type.

In terms of equilibrium existence, the model allows for a rich set of equilibrium scenarios. Under some initial conditions there is a unique equilibrium in which good experts tell the truth while bad experts disclose their signals only partially. Also, depending on the degree of career concerns we can find equilibria in which good experts also disclose their signals only partially as they do not like to be perceived biased. A limiting case is a situation in which no information is transmitted to the decision maker.

In extensions to the model I also linked it to the career concerns models with anti-herding behavior. A possible extension left to future work is to have a model with two experts with the same potential bias. This situation is different than the current model as a bad expert could semi-pool on a good expert (of same appartenance) but at the same time any of the two experts (irrespective of type) could pool on the action of the counterpart. Finally, another possible extension is to allow partial state verifiability. A model of this type would nest both the current model and a model with no strategic interaction between experts.

References

[1] Austen-Smith, D., 1992, Explaining the Vote: Constituency Constraints on Sophisticated Voting, *American Journal of Political Science*, 36, 68–95.

[2] Austen-Smith, D., 1993, Interested Experts and Policy Advice: Multiple Referrals under Open Rule, *Games and Economic Behavior*, 5, 3–43.

[3] Avery, C.N., and J. A. Chevalier, 1999, Herding over Career, *Economic Letters*, 63, 327–333.

[4] Bernheim, B. D., 1994, A Theory of Conformity, *Journal of Political Economy*, 102, 841-877.

[5] Crawford, V. P., and J. Sobel, 1982, Strategic Information Transmission, *Econometrica*, 50, 1431-1451.

[6] Dasgupta, A., and A. Prat, 2008, Information Aggregation in Financial Markets with Career Concerns, *Journal of Economic Theory*, 143, 83-113.

[7] Gilligan, T., and K. Krehbiel, 1989, Asymmetric Information and Legislative Rules with a Heterogeneous Committee, *American Journal of Political Science*, 1459-490.

[8] Levy, G., 2004, Anti-Herding and Strategic Consultation, *European Economic Review*, 48, 503-525.

[9] Loury, G., 1994, Self-censorship in public discourse, *Rationality and Society,* 6, 428-61.

[10] Krishna, V., and J. Morgan, 2001, A Model of Expertise, *Quarterly Journal of Economics*, 116, 747-775.

[11] Krugman, P., 2012, The New Political Correctness, The New York Times Opinion Pages, New York, USA, 2012, May 26.

[12] Morris, S., 2001, Political Correctness, *Journal of Political Economy*, 109, 231-265.

[13] McGee, A., and H. Yang, 2013, Cheap Talk with Two Senders and Complementary Information, *Games and Economic Behavior*, 79, 181-191.

[14] Ottaviani, M., and P. Sorensen, 2001, Information Aggregation in Debate: Who Should Speak First, *Journal of Public Economics*, 81, 393-421.

[15] Prendergast, C., and L. Stole, 1996, Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning, *Journal of Political Economy*, 104, 1105-34.

[16] Scharfstein, D S., and J. C. Stein, 1990, Herd Behavior and Investment, *American Economic Review*, 80, 465-479.

[17] Sobel, J., 1985, A Theory of Credibility, *Review of. Economic Studies*, 52, 557–73.

[18] Trueman, B., 1994, Analyst Forecasts and Herding Behavior, *Review of Financial Studies*, 7, 97-124.

7 Appendix

7.1 Proof of Proposition 1

D believes that if $l_2 = 1$, L is good while if $l_2 = 0$ probability that L is of a good type is λ_2^L. Similarly, if $r_2 = 0$, R is good while if $r_2 = 1$ R is good with probability λ_2^R.

Based on these beliefs we could compute by Bayes' rule also probability of the state being 1 in period 2.

- $\Pr(x_2 = 1 | l_2 = 1, r_2 = 0) = \frac{p^L\left(1-p^R\right)}{(1-p^L)p^R + p^L(1-p^R)}$

  This is due to the fact that a 1 report from L happens with probability p^L while a 0 report from R happens with probability p^R. Note that, in the last period if the decision maker observes $r_2 = 0$ and $l_2 = 1$ he knows that both R and L are good.

- $\Pr(x_2 = 1 | l_2 = 0, r_2 = 0) = \frac{\left(1-\lambda_2^L p^L\right)\left(1-p^R\right)}{1-\lambda_2^L(p^L(1-p^R)+p^R(1-p^L))}$

  In calculating this probability we take into account that $l_2 = 0$ could be sent by both a bad and a good expert L while $r_2 = 0$ could be sent only by a a good R. The probability that a 0 report could come from a bias L is $1 - \lambda_2^L$ while the probability that $l_2 = 0$ is sent by a good expert is $\lambda_2^L\left(1 - p^L\right)$. Thus $\Pr(l_2 = 0 | x_2 = 1) = 1 - \lambda_2^L + \lambda_2^L\left(1 - p^L\right) = 1 - \lambda_2^L p^L$. We also know that $\Pr(r_2 = 0 | x_2 = 1) = 1 - p^R$.

  Hence, $\Pr(l_2 = 0, r_2 = 0 | x_2 = 1) = \left(1 - \lambda_2^L p^L\right)\left(1 - p^R\right)$.

  Similarly when the state is 0, $\Pr(l_2 = 0 | x_2 = 0) = 1 - \lambda_2^L + \lambda_2^L p^L$ while $\Pr(r_2 = 0 | x_2 = 0) = p^R$

  Thus, $\Pr(l_2 = 0, r_2 = 0 | x_2 = 0) = \left(1 - \lambda_2^L\left(1 - p^L\right)\right)p^R$. .

- $\Pr(x_2 = 1 | l_2 = 1, r_2 = 1) = \frac{p^L\left(1-\left(1-p^R\right)\lambda_2^R\right)}{1-\lambda_2^R[p^R(1-p^L)+p^L(1-p^R)]}$

  I used that $\Pr(l_2 = 1 | x_2 = 1) = p^L$ and

$\Pr(r_2 = 1|x_2 = 1) = 1 - \lambda_2^R + p^R \lambda_2^R = 1 - \left(1 - p^R\right)\lambda_2^R.$

Hence, $\Pr(l_2 = 1, r_2 = 1|x_2 = 1) = p^L\left(1 - \left(1 - p^R\right)\lambda_2^R\right).$

$\Pr(l_2 = 1, r_2 = 1|x_2 = 0) = \left(1 - \lambda_2^R p^R\right)\left(1 - p^L\right).$

- $\Pr(x_2 = 1|l_2 = 0, r_2 = 1) = \frac{\left(1-p^L\lambda_2^L\right)\left(1-\left(1-p^R\right)\lambda_2^R\right)}{2-\lambda_2^R-\lambda_2^L+[p^L(1-p^R)+p^R(1-p^L)]\lambda_2^R\lambda_2^L}$

  In this result we have possibility of bias from both experts.

  $\Pr(l_2 = 0, r_2 = 1|x_2 = 1) = \left(1 - p^L\lambda_2^L\right)\left(1 - \left(1 - p^R\right)\lambda_2^R\right).$

  $\Pr(l_2 = 0, r_2 = 1|x_2 = 0) = \left(1 - \left(1 - p^L\right)\lambda_2^L\right)\left(1 - p^R\lambda_2^R\right).$

  The denominator of the above probability is just

$$\Pr(l_2 = 0, r_2 = 1|x_2 = 1) + \Pr(l_2 = 0, r_2 = 1|x_2 = 0).$$

We used the following intermediate probabilities in the above computations:

$\Pr(l_2 = 1|x_2 = 1) = p^L\lambda_2^L$

$\Pr(l_2 = 1|x_2 = 0) = \left(1 - p^L\right)\lambda_2^L$

$\Pr(l_2 = 0|x_2 = 1) = 1 - p^L\lambda_2^L$

$\Pr(l_2 = 0|x_2 = 0) = 1 - \left(1 - p^L\right)\lambda_2^L$

$\Pr(r_2 = 1|x_2 = 1) = 1 - \left(1 - p^R\right)\lambda_2^R$

$\Pr(r_2 = 1|x_2 = 0) = 1 - p^R\lambda_2^R$

$\Pr(r_2 = 0|x_2 = 1) = \left(1 - p^R\right)\lambda_2^R$

$\Pr(r_2 = 0|x_2 = 0) = p^R\lambda_2^R$

In any informative equilibrium the reports are positively (without loss of generality) correlated with state of the world.

As the decision maker's payoff in the last period is $-E\left(x_2 - a_2\right)^2$, then for some messages (l_2, r_2) the optimal action of the principal is:

$$a_2^*(l_2, r_2) = \Pr(x_2 = 1|l_2, r_2)\,1 + \Pr(x_2 = 0|l_2, r_2)\,0 = \Pr(x_2 = 1|l_2, r_2)$$

An expert R of good type chooses r_2 such that she maximize her payoff.

We know that the signal precisions are such that $\Pr\left(s_2^R = 0|x_2 = 1\right) = 1 - p^R$ and $\Pr\left(s_2^R = 0|x_2 = 1\right) = p^R$. Thus if the R receives $s_2^R = 0$, then

$$-E\left[\left(x_2 - a_2^*\right)^2 |s_2^R = 0\right] =$$

$$\begin{cases} -\left(1-p^R\right)\left[1 - \Pr\left(x_2 = 1|l_2, r_2 = 0\right)\right]^2 - p^R\left[\Pr\left(x_2 = 1|l_2, r_2 = 0\right)\right]^2 & \text{if } r_2 = 0 \\ -\left(1-p^R\right)\left[1 - \Pr\left(x_2 = 1|l_2, r_2 = 1\right)\right]^2 - p^R\left[\Pr\left(x_2 = 1|l_2, r_2 = 1\right)\right]^2 & \text{if } r_2 = 1 \end{cases}$$

The net benefit from declaring 0 instead of 1 is:

$$\begin{aligned} &\left[\Pr\left(x_2 = 0|l_2, r_2 = 0\right) - \Pr\left(x_2 = 0|l_2, r_2 = 1\right)\right] \\ &\times\left[2p^R - \Pr\left(x_2 = 0|l_2, r_2 = 0\right) - \Pr\left(x_2 = 0|l_2, r_2 = 1\right)\right] \end{aligned}$$

This term is strictly positive as we are looking at an informative equilibrium.

Also, $p^R > \Pr\left(x_2 = 0|l_2, r_2 = 0\right)$ and $p^R > \Pr\left(x_2 = 0|l_2, r_2 = 1\right)$.

Note that the maximum $\Pr\left(x_2 = 0|l_2, r_2 = 0\right)$ is $\frac{p^L p^R}{p^L p^R + (1-p^L)(1-p^R)} < p^R$ and the maximum $\Pr\left(x_2 = 0|l_2, r_2 = 1\right)$ is $\frac{p^L\left(1-p^R\right)}{p^L(1-p^R)+(1-p^L)p^R} < p^R$ for $p^R, p^L > \frac{1}{2}$.

If R receives $s_2^R = 1$, then

$$-E\left[\left(x_2 - a_2^*\right)^2 |s_2^R = 1\right] =$$

$$\begin{cases} -p^R\left[1 - \Pr\left(x_2 = 1|l_2, r_2 = 0\right)\right]^2 - \left(1-p^R\right)\left[\Pr\left(x_2 = 1|l_2, r_2 = 0\right)\right]^2 & \text{if } r_2 = 0 \\ -p^R\left[1 - \Pr\left(x_2 = 1|l_2, r_2 = 1\right)\right]^2 - \left(1-p^R\right)\left[\Pr\left(x_2 = 1|l_2, r_2 = 1\right)\right]^2 & \text{if } r_2 = 1 \end{cases}$$

The net benefit from declaring 1 instead of 0 is:

$$\begin{aligned} &\left[\Pr\left(x_2 = 1|l_2, r_2 = 1\right) - \Pr\left(x_2 = 1|l_2, r_2 = 0\right)\right] \\ &\times\left[2p^R - \Pr\left(x_2 = 1|l_2, r_2 = 1\right) - \Pr\left(x_2 = 1|l_2, r_2 = 0\right)\right] \end{aligned}$$

This term is also strictly positive.

Thus, a good R experts always report her signals. By contrast, an expert R of bad type declares 1 irrespective of her signal as $r_1 = 1$ weakly increases the action of the decision maker.

We see that thus by declaring her signal a good R strictly increases her payoff while a bad R weakly increases her payoff.

By a similar argument, an expert L of good type declares her signal and an expert L of bad type declares her bias.

7.2 Proof of Proposition 2

First I acknowledge that the reputation acquired in the first period affects decision maker optimal action:

$$\frac{da_2^*\left(l_2=1,r_2=1\right)}{d\lambda_2^R}=\frac{\left(1-p^L\right)p^L\left(-1+2p^R\right)}{\left[1-\lambda_2^R A\right]^2}>0$$

$$\frac{da_2^*\left(l_2=0,r_2=1\right)}{d\lambda_2^R}=\frac{\left[1-\lambda_2^L\left(1-p^L\right)\right]\left(1-\lambda_2^L p^L\right)\left(2p^R-1\right)}{\left[2-\lambda_2^R-\lambda_2^L+\lambda_2^R\lambda_2^L A\right]^2}>0$$

$$\frac{da_2^*\left(l_2=0,r_2=0\right)}{d\lambda_2^L}=-\frac{\left(1-p^R\right)p^R\left[-1+2p^L\right]}{\left[1-\lambda_2^L A\right]^2}<0$$

$$\frac{da_2^*\left(l_2=0,r_2=1\right)}{d\lambda_2^L}=\frac{\left[1-\lambda_2\left(1-p^R\right)\right]\left(1-\lambda_2 p^R\right)\left(1-2p^L\right)}{\left[2-\lambda_2^R-\lambda_2^L+\lambda_2^R\lambda_2^L A\right]^2}<0$$

For a good type R her expected payoff at the beginning of period 2 given decision maker posterior belief on the experts reputation is:

$$v_R^G\left(\Lambda_2\right)=-E\left[\left(x_2-a_2^*\right)|\Lambda_2\right]$$

$$\begin{aligned}
E\left[\left(x_2-a_2^*\right)|\Lambda_2\right] &= \frac{1}{2}p^R\Pr(l_2=1|x_2=1)\left(1-\Pr\left(x_2=1|l_2=1,r_2=1\right)\right)^2+\\
\frac{1}{2}\left(1-p^R\right)\Pr(l_2 &= 1|x_2=0)\left(0-\Pr\left(x_2=1|l_2=1,r_2=1\right)\right)^2+\\
\frac{1}{2}\left(1-p^R\right)\Pr(l_2 &= 1|x_2=1)\left(1-\Pr\left(x_2=1|l_2=1,r_2=0\right)\right)^2\\
\frac{1}{2}p^R\Pr(l_2 &= 1|x_2=0)\left(0-\Pr\left(x_2=1|l_2=1,r_2=0\right)\right)^2+\\
\frac{1}{2}p^R\Pr(l_2 &= 0|x_2=1)\left(1-\Pr\left(x_2=1|l_2=0,r_2=1\right)\right)^2+\\
\frac{1}{2}\left(1-p^R\right)\Pr(l_2 &= 0|x_2=0)\left(0-\Pr\left(x_2=1|l_2=0,r_2=1\right)\right)^2+\\
\frac{1}{2}(1-p^R)\Pr(l_2 &= 0|x_2=1)\left(1-\Pr\left(x_2=1|l_2=0,r_2=0\right)\right)^2\\
\frac{1}{2}p^R\Pr(l_2 &= 0|x_2=0)\left(0-\Pr\left(x_2=1|l_2=0,r_2=0\right)\right)^2
\end{aligned}$$

In the above expression the terms that vary with λ_2^R are those with $r_2=1$ as the decision maker could not differentiate whether a 1 report comes from a biased or unbiased R expert. Next, I look only at these terms. Also, I denote $\Pr\left(x_2=1|l,r\right)$ as P_{lr}, for simplicity of notation .

$$\begin{aligned}
T^R &= \frac{1}{2}p^Rp^L\lambda_2^L\left(1-P_{11}\right)^2+\frac{1}{2}\left(1-p^R\right)\left(1-p^L\right)\lambda_2^LP_{11}^2+\\
&\quad \frac{1}{2}p^R\left(1-p^L\lambda_1^L\right)\left(1-P_{01}\right)^2+\frac{1}{2}\left(1-p^R\right)\left(1-\left(1-p^L\right)\lambda_2^L\right)P_{01}^2
\end{aligned}$$

Taking derivative with respect to λ_2^R I get:

$$\begin{aligned}
\frac{dT^R}{d\lambda_2^R} &= \left[-p^Rp^L\lambda_2^L+\left[p^Rp^L+\left(1-p^R\right)\left(1-p^L\right)\right]\lambda_2^LP_{11}\right]\frac{dP_{11}}{d\lambda_2^R}+\\
&\quad \left[-\left(1-p^L\lambda_1^L\right)p^R+\left[p^R\left(1-p^L\lambda_1^L\right)+\left(1-p^R\right)1-\left(1-p^L\right)\lambda_2^L\right]P_{01}\right]\frac{dP_{01}}{d\lambda_2^R}
\end{aligned}$$

If there was no possibility of bias P_{11} would have been $\frac{p^Rp^L}{\left[p^Rp^L+(1-p^R)(1-p^L)\right]}$.

However, in our case

$$P_{11} < \frac{p^R p^L}{[p^R p^L + (1-p^R)(1-p^L)]}$$

due to the fact that a decision maker could not differentiate between a 1 report from a bias or from unbiased player. As a result

$$-p^R p^L \lambda_2^L + \left[p^R p^L + \left(1-p^R\right)\left(1-p^L\right)\right] \lambda_2^L P_{11} < 0$$

By a similar argument

$$P_{01} < \frac{\left(1-p^L\lambda_1^L\right)p^R}{\left[p^R\left(1-p^L\lambda_1^L\right) + (1-p^R)\,1-(1-p^L)\,\lambda_2^L\right]}$$

as the decision maker has to account for the fact that a $r_2 = 1$ report could come from both a biased and unbiased R. Thus,

$$\left[-\left(1-p^L\lambda_1^L\right)p^R + \left[p^R\left(1-p^L\lambda_1^L\right) + \left(1-p^R\right)1-\left(1-p^L\right)\lambda_2^L\right]P_{01}\right] < 0$$

.

Further as $\frac{dP_{11}}{d\lambda_2^R} > 0$ and $\frac{dP_{01}}{d\lambda_2^R}$, then $\frac{dE\left[\left(x_2-a_2^*\right)|\lambda_2^L,\lambda_2^R\right]}{d\lambda_2^R} < 0$.

As $v_R^G(\Lambda_2) = -E\left[(x_2-a_2^*)\,|\Lambda_2\right]$ we can conclude that

$$\frac{dv_R^G(\Lambda_2)}{d\lambda_2^R} > 0.$$

Next, I look at how $v_R^G(\Lambda_2)$ changes with λ_2^L.

As the expert L is biased towards 0 for simplicity of calculations I express I will express $E\left[(x_2-a_2^*)\,|\Lambda_2\right]$ in terms of $\Pr(x=0|l_2,r_2)$ which I denote with Q_{lr}.

First, I look only at the terms which involve $l_2 = 0$ as Q_{0r} changes with λ_2^L.

$$\begin{aligned} T_1^L &= \frac{1}{2}p^R\left(1-p^L\lambda_2^L\right)Q_{01}^2 + \frac{1}{2}\left(1-p^R\right)\left(1-\left(1-p^L\right)\lambda_2^L\right)\left(1-Q_{01}\right)^2 + \\ & \frac{1}{2}\left(1-p^R\right)\left(1-p^L\lambda_2^L\right)Q_{00}^2 + \frac{1}{2}p^R\left(1-\left(1-p^L\right)\lambda_2^L\right)\left(1-Q_{00}\right)^2 \end{aligned}$$

Taking derivatives with respect to λ_2^L, I get $\frac{dT_1^L}{d\lambda_2^L}$ is

$$\begin{aligned}\frac{dT_1^L}{d\lambda_2^L} &= \left[p^R\left(1-p^L\lambda_2^L\right)Q_{01}+\left(1-p^R\right)\left(1-\left(1-p^L\right)\lambda_2^L\right)\left(Q_{01}-1\right)\right]\frac{dQ_{01}}{d\lambda_2^L}+\\&\left[\left(1-p^R\right)\left(1-p^L\lambda_2^L\right)Q_{00}+p^R\left(1-\left(1-p^L\right)\lambda_2^L\right)\left(Q_{00}-1\right)\right]\frac{dQ_{00}}{d\lambda_2^L}-\\&\frac{1}{2}p^Rp^LQ_{01}^2-\frac{1}{2}\left(1-p^R\right)\left(1-p^L\right)\left(1-Q_{01}\right)^2-\\&\frac{1}{2}\left(1-p^R\right)p^LQ_{00}^2-\frac{1}{2}p^R\left(1-p^L\right)\left(1-Q_{00}\right)^2\end{aligned}$$

Second, even though Q_{1r} does not change with λ_2^L, the overall $v_R^G\left(\lambda_2^L,\lambda_2^R\right)$ changes with λ_2^L when $l_2=1$ as Q_{1r} is multiplied by $\Pr(l_2=1|x_2)$.

Hence $\frac{dE\left[\left(x_2-a_2^*\right)|\Lambda_2\right]}{d\lambda_2^L}$ becomes:

$$\begin{aligned}&\left[p^R\left(1-p^L\lambda_2^L\right)Q_{01}+\left(1-p^R\right)\left(1-\left(1-p^L\right)\lambda_2^L\right)\left(Q_{01}-1\right)\right]\frac{dQ_{01}}{d\lambda_2^L}+\\&+\left[\left(1-p^R\right)\left(1-p^L\lambda_2^L\right)Q_{00}+p^R\left(1-\left(1-p^L\right)\lambda_2^L\right)\left(Q_{00}-1\right)\right]\frac{dQ_{00}}{d\lambda_2^L}+\\&\frac{1}{2}p^Rp^L\left(Q_{11}^2-Q_{01}^2\right)+\frac{1}{2}\left(1-p^R\right)\left(1-p^L\right)\left[\left(1-Q_{11}\right)^2-\left(1-Q_{01}\right)^2\right]+\\&\frac{1}{2}\left(1-p^R\right)p^L\left[Q_{10}^2-Q_{00}^2\right]+\frac{1}{2}p^R\left(1-p^L\right)\left[\left(1-Q_{10}\right)^2-\left(1-Q_{00}\right)^2\right]\end{aligned}$$

Now, $\frac{dQ_{01}}{d\lambda_2^L}$ and $\frac{dQ_{00}}{d\lambda_2^L}$ are positive as a 0 report from L is more trusted by the decision maker when L's reputation increases. Further more,

$$p^R\left(1-p^L\lambda_2^L\right)Q_{01}+\left(1-p^R\right)\left(1-\left(1-p^L\right)\lambda_2^L\right)\left(Q_{01}-1\right)<0$$

as

$$Q_{01}<\frac{\left(1-p^R\right)\left(1-\left(1-p^L\right)\lambda_2^L\right)}{p^R\left(1-p^L\lambda_2^L\right)+\left(1-p^R\right)\left(1-\left(1-p^L\right)\lambda_2^L\right)}$$

which would be $\Pr\left(x_2=0|l=0,r=1\right)$ in case R's report would be unbiased. I also used the fact that in the expression

$$Q_{01}=\frac{\left(1-p^R\lambda_2^R\right)\left(1-\left(1-p^L\right)\lambda_2^L\right)}{\left(1-\lambda_2^R+p^R\lambda_2^R\right)\left(1-p^L\lambda_2^L\right)+\left(1-p^R\right)\left(1-\left(1-p^L\right)\lambda_2^L\right)}$$

the term $\frac{1-p^R\lambda_2^R}{1-\lambda_2^R+p^R\lambda_2^R}$ decreases with λ_2^R for $p^R \geq \frac{1}{2}$.

Similarly,

$$\left[\left(1-p^R\right)\left(1-p^L\lambda_2^L\right)Q_{00}+p^R\left(1-\left(1-p^L\right)\lambda_2^L\right)(Q_{00}-1)\right]<0$$

Thus the first two terms in $\frac{dE\left[\left(x_2-a_2^*\right)|\Lambda_2\right]}{d\lambda_2^L}$ are negative.

The third term in $\frac{dE\left[\left(x_2-a_2^*\right)|\Lambda_2\right]}{d\lambda_2^L}$ is:

$\frac{1}{2}p^Rp^L\left(Q_{11}^2-Q_{01}^2\right)+\frac{1}{2}\left(1-p^R\right)\left(1-p^L\right)\left[(1-Q_{11})^2-(1-Q_{01})^2\right]$ and can be further written as :

$$\frac{1}{2}\left(Q_{11}-Q_{01}\right)\left[2p^Rp^L-\left(p^Rp^L+\left(1-p^R\right)\left(1-p^L\right)\right)(P_{11}+P_{01})\right]$$

Now

$$2p^Rp^L-\left(p^Rp^L+\left(1-p^R\right)\left(1-p^L\right)\right)(P_{11}+P_{01})>0$$

as

$$P_{11}+P_{01}<\frac{2p^Rp^L}{p^Rp^L+(1-p^R)(1-p^L)}$$

and maximum P_{11} is exactly $\frac{p^Lp^R}{p^Rp^L+(1-p^R)(1-p^L)}$ in the case when R's report would be unbiased. Furthermore the maximum P_{01} is $\frac{\left(1-p^L\right)p^R}{(1-p^L)p^R+p^L(1-p^R)}$ which is smaller than $\frac{p^Lp^R}{p^Rp^L+(1-p^R)(1-p^L)}$ as $p^L \geq \frac{1}{2}$; this is the case when both R and L would give unbiased reports.

Also, $Q_{11}-Q_{01} \leq 0$ as we are looking at informative equilibria where the state is correlated with the experts reports.

We can conclude that third term of $\frac{dE\left[\left(x_2-a_2^*\right)|\Lambda_2\right]}{d\lambda_2^L}$ is non-positive.

The last term in $\frac{dE\left[\left(x_2-a_2^*\right)|\lambda_2^L,\lambda_2^R\right]}{d\lambda_2^L}$ is

$$\frac{1}{2}\left(1-p^R\right)p^L\left[Q_{10}^2-Q_{00}^2\right]+\frac{1}{2}p^R\left(1-p^L\right)\left[(1-Q_{10})^2-(1-Q_{00})^2\right]$$

which can be written as

$$\frac{1}{2}\left(Q_{10}-Q_{00}\right)\left[2\left(1-p^{R}\right)p^{L}-\left(\left(1-p^{R}\right)p^{L}+p^{R}\left(1-p^{L}\right)\right)\left(P_{10}+P_{00}\right)\right]$$

which by similar argument is non-positive.

Putting everything together we can conclude that

$$\frac{dv_{R}^{G}\left(\Lambda_{2}\right)}{d\lambda_{2}^{L}}\geq 0.$$

Now, let's look at the value of reputation for a bad R expert.:

$$v_{R}^{B}\left(\Lambda_{2}\right)=E\left[a_{2}^{*}|\Lambda_{2}\right]$$

$$\begin{aligned}v_{R}^{B}\left(\Lambda_{2}\right) &= \frac{1}{2}\left[\Pr\left(l_{2}=1|x_{2}=0\right)+\Pr\left(l_{2}=1|x_{2}=1\right)\right]\Pr\left(x_{2}=1|l_{2}=1,r_{2}=1\right)+\\ &\quad \frac{1}{2}\left[\Pr\left(l_{2}=0|x_{2}=0\right)+\Pr\left(l_{2}=0|x_{2}=1\right)\right]\Pr\left(x_{2}=1|l_{2}=0,r_{2}=1\right)\end{aligned}$$

$$\begin{aligned}v_{R}^{B}\left(\Lambda_{2}\right) &= \frac{1}{2}\left[\left(1-p^{L}\right)\lambda_{2}^{L}+p^{L}\lambda_{2}^{L}\right]P_{11}+\frac{1}{2}\left[\left(1-p^{L}\lambda_{2}^{L}\right)+\left(1-\left(1-p^{L}\right)\lambda_{2}^{L}\right)\right]P_{01}\\ &= \frac{1}{2}\lambda_{2}^{L}P_{11}+\frac{1}{2}\left(1-\lambda_{2}^{L}\right)P_{01}\end{aligned}$$

It is easy to see that

$$\frac{dv_{R}^{B}\left(\Lambda_{2}\right)}{d\lambda_{2}^{R}}>0$$

as $\frac{dP_{11}}{d\lambda_{2}^{R}}>0$ and $\frac{dP_{01}}{d\lambda_{2}^{R}}>0$.

This is due to the fact that a bad expert might be believed more when she declares 1 if she has a higher reputation.

Also,

$$v_{R}^{B}\left(\Lambda_{2}\right)=\frac{1}{2}\lambda_{2}^{L}P_{11}+\frac{1}{2}\left(1-\lambda_{2}^{L}\right)P_{01}$$

$$\frac{dv_R^B(\Lambda_2)}{d\lambda_2^L} = \frac{1}{2}(P_{11} - P_{01}) - \frac{1}{2}\lambda_2^L \frac{dP_{01}}{\lambda_2^L} + \frac{1}{2}\frac{dP_{01}}{\lambda_2^L} = \frac{1}{2}(P_{11} - P_{01}) + \frac{1}{2}\frac{dP_{01}}{\lambda_2^L}\left(1 - \lambda_2^L\right)$$

As

$$P_{11} - P_{01} = \frac{\left[1 - \left(1 - p^R\right)\lambda_2^R\right]\left(1 - p^R\lambda_2^R\right)\left(2p^L - 1\right)}{\left[1 - \lambda_2^R A\right]\left[2 - \lambda_2^R - \lambda_2^L + A\lambda_2^R\lambda_2^L\right]}$$

and

$$\frac{dP_{01}}{\lambda_2^L}\left(1 - \lambda_2^L\right) = \frac{\left[1 - \lambda_2\left(1 - p^R\right)\right]\left(1 - p^R\lambda_2\right)\left(1 - 2p^L\right)\left(1 - \lambda_2^L\right)}{\left[2 - \lambda_2^R - \lambda_2^L + \lambda_2^R\lambda_2^L A\right]^2}$$

we can conclude after some simple calculations that

$$\frac{dv_R^B(\Lambda_2)}{d\lambda_2^L} > 0$$

The ex-ante expected payoff of a bad R expert is a weighted reputational average of posterior beliefs on the state when R reports 1 for sure while L could declare both 0 and 1.

When looking at the change imposed in $v_R^B(\Lambda_2)$ by a change in λ_2^L we see that there is a positive effect coming from the fact that both reporting 1 increases the chances of the state being 1. However there is a negative marginal effect of λ_2^L on the probability of the state being 1 when L reports 0. However, this negative effect is lower in absolute value than the positive effect and the overall effect of an increase of L's reputation on second period ex-ante expected payoff of R is positive.

This result is due to the fact that a bad R expert will prefer the L expert to be of a good type as her report will be as closed as possible to the state of the world while an L expert will report in the opposite direction of L's preference.

7.3 How experts' reputations are calculated

The reputation of the R is described by:

$$\lambda_2^R(l_1, r_1) = \frac{\lambda_1^R \phi_G^R(r_1|l_1)}{\lambda_1^R \phi_G^R(r_1|l_1) + \left(1 - \lambda_1^R\right) \phi_B^R(r_1|l_1)}$$

while the reputation of L is described by:

$$\lambda_2^L(l_1, r_1) = \frac{\lambda_1^L \phi_G^L(l_1|r_1)}{\lambda_1^L \phi_G^L(l_1|r_1) + \left(1 - \lambda_1^L\right) \phi_B^L(l_1|r_1)}$$

Now,

$$\phi_k^R(r_1|l_1) = \phi_k^R(r_1|x=1) \Pr(x=1|l_1) + \phi_k^R(r_1|x=0) \Pr(x=0|l_1)$$

and

$$\phi_k^L(l_1|r_1) = \phi_k^L(l_1|x=1) \Pr(x=1|r_1) + \phi_k^R(r_1|x=0) \Pr(x=0|r_1)$$

Where:

$$\phi_k^R(1|x_i) = \left[p^R \pi_k^R(x_i) + \left(1 - p^R\right) \left(\pi_k^R(1 - x_i)\right)\right]$$

$$\phi_k^L(1|x_i) = \left[p^L \left(1 - \pi_k^L(x_i)\right) + \left(1 - p^L\right) \left(1 - \pi_k^L(1 - x_i)\right)\right]$$

and

$$\phi_k^R(0|x_i) = \left[p^R \left(1 - \pi_k^R(x_i)\right) + \left(1 - p^R\right) \left(1 - \pi_k^R(1 - x_i)\right)\right]$$

$$\phi_k^L(0|x_i) = \left[p^L \left(\pi_k^L(x_i)\right) + \left(1 - p^L\right) \left(\pi_k^L(1 - x_i)\right)\right]$$

$$\Pr(x=1|l_1) = \frac{\Pr(l_1|x=1)}{\Pr(l_1|x=1) + \Pr(l_1|x=0)}$$

$$\Pr(l_1 = 1|x=1) = \begin{bmatrix} p^L \left(\lambda_1^L \left(1 - \pi_G^L(1)\right) + \left(1 - \lambda_1^L\right) \left(1 - \pi_B^L(1)\right)\right) + \\ \left(1 - p^L\right) \left(\lambda_1^L \left(1 - \pi_G^L(0)\right) + \left(1 - \lambda_1^L\right) \left(1 - \pi_B^L(0)\right)\right) \end{bmatrix}$$

$$\Pr(l_1 = 1|x=0) = \begin{bmatrix} \left(1 - p^L\right) \left(\lambda_1^L \left(1 - \pi_G^L(1)\right) + \left(1 - \lambda_1^L\right) \left(1 - \pi_B^L(1)\right)\right) + \\ p^L \left(\lambda_1^L \left(1 - \pi_G^L(0)\right) + \left(1 - \lambda_1^L\right) \left(1 - \pi_B^L(0)\right)\right) \end{bmatrix}$$

$$\Pr(l_1 = 0|x = 1) = \left[\begin{array}{c} p^L\left(\lambda_1^L \pi_G^L(1) + \left(1-\lambda_1^L\right)\pi_B^L(1)\right) + \\ \left(1-p^L\right)\left(\lambda_1^L \pi_G^L(0) + \left(1-\lambda_1^L\right)\pi_B^L(0)\right) \end{array}\right]$$

$$\Pr(l_1 = 0|x = 0) = \left[\begin{array}{c} p^L\left(\lambda_1^L \pi_G^L(0) + \left(1-\lambda_1^L\right)\pi_B^L(0)\right) + \\ \left(1-p^L\right)\left(\lambda_1^L \pi_G^L(1) + \left(1-\lambda_1^L\right)\pi_B^L(1)\right) \end{array}\right]$$

$$\Pr(r_1 = 1|x = 1) = \left[\begin{array}{c} p^R\left(\lambda_1^R \pi_G^R(1) + \left(1-\lambda_1^R\right)\pi_B^R(1)\right) \\ +\left(1-p^R\right)\left(\lambda_1^R \pi_G^R(0) + \left(1-\lambda_1^R\right)\pi_B^R(0)\right) \end{array}\right]$$

$$\Pr(r_1 = 1|x = 0) = \left[\begin{array}{c} p^R\left(\lambda_1^R \pi_G^R(0) + \left(1-\lambda_1^R\right)\pi_B^R(0)\right) \\ +\left(1-p^R\right)\left(\lambda_1^R \pi_G^R(1) + \left(1-\lambda_1^R\right)\pi_B^R(1)\right) \end{array}\right]$$

$$\Pr(r_1 = 0|x = 1) = \left[\begin{array}{c} p^R\left(\lambda_1^R\left(1-\pi_G^R(1)\right) + \left(1-\lambda_1^R\right)\left(1-\pi_B^R(1)\right)\right) \\ +\left(1-p^R\right)\left(\lambda_1^R\left(1-\pi_G^R(0)\right) + \left(1-\lambda_1^R\right)\left(1-\pi_B^R(0)\right)\right) \end{array}\right]$$

$$\Pr(r_1 = 0|x = 0) = \left[\begin{array}{c} p^R\left(\lambda_1^R\left(1-\pi_G^R(0)\right) + \left(1-\lambda_1^R\right)\left(1-\pi_B^R(0)\right)\right) \\ +\left(1-p^R\right)\left(\lambda_1^R\left(1-\pi_G^R(1)\right) + \left(1-\lambda_1^R\right)\left(1-\pi_B^R(1)\right)\right) \end{array}\right]$$

7.4 Decision Maker's Optimal Decision in Equilibrium

The decision maker's optimal action is the belief that the state is 1 when the messages (l_1, r_1) are announced:

$$a_1^*(l_1, r_1) = \Gamma(1|l_1, r_1) = \frac{\Pr(l_1, r_1|1)}{\Pr(l_1, r_1|1) + \Pr(l_1, r_1|0)}$$

$$\Pr(l_1, r_1|x_1) = \left[\lambda_1^R \phi_G^R(r_1|x_1) + \left(1-\lambda_1^R\right)\phi_B^R(r_1|x_1)\right]\left[\lambda_1^L \phi_G^L(l_1|x_1) + \left(1-\lambda_1^L\right)\phi_B^l(l_1|x_1)\right]$$

7.5 Proof of Claim 4

The messages sent by the experts do not influence D's belief about the state of the world which is $\Gamma(L_1, r_1) = \frac{1}{2}$ for any $l_1, r_1 \subset \{0, 1\}$ nor his optimal action $a_1^*(l_1, r_1) = \frac{1}{2}$ for any $l_1, r_1 \in \{0, 1\}$. Thus the experts are indifferent between any strategies used including these uninformative ones, thus they do not have any incentive to deviate.

As a result these strategies which convey no information determine D's belief and action.

I turn next to the more interesting case when some information is transmitted to the decision maker.

7.6 Proof of Claim 5

Assume that such an equilibrium existed then $\lambda_2^R(l_1, r_1) = \lambda_1^R$ and $\lambda_2^L(l_1, r_1) = \lambda_1^L$. But this implies that there is no reputational cost for R of announcing 1. On the other hand D will take a higher action after R sending a 1 message. Because a biased R prefers a higher action there is a strict incentive to send $r_1 = 1$. Thus regardless of the signal the bad R will send message 1: $\pi_B^R(1) = \pi_B^R(0) = 1$ which is a contradiction. Similarly when there is no reputational cost for reporting 0 a bad L reports it irrespective of his signal - again a contradiction.

7.7 Proof of Proposition 3

Proof 1. If $s_1^R = 0$ then $\Pi_G^R(s_1^R, l_1) < 0$ since announcing 1 will increase the action of the decision maker irrespective of l_1 and the good R will be further from his signal. Thus the good R will never have an incentive to announce 1 when the signal is 0 and $\pi_G^R(0) = 0$. ■

Proof 2. If $s_1^L = 1$ then $\Pi_G^R(s_1^R, l_1) < 0$ since announcing announcing 0 will decrease the action of the decision maker irrespective of r_1 and the good R will be farther from his signal. Thus the good L will never have an incentive of announcing 0 when the signal is 1.Thus $\pi_G^L(1) = 0$ ■

7.8 Proof of Proposition 4

Proof 1. I will prove the first point by contradiction.

Suppose not and $\lambda_2^R(l_1, 1) > \lambda_2^R(l_1, 0)$; in this situation R has a both a higher reputation by declaring 1 and a higher current payoff for any $s^R = \{0, 1\}$; thus the biased R will always say 1 and $\pi_B^R(0) = \pi_B^R(1) = 1$ resulting in $\phi_B^R(0|l_1) = \phi_B^R(1|l_1)$. Then,

$$\lambda_2^R(l_1, r_1) = \frac{1}{1 + \frac{1-\lambda_1^R}{\lambda_1^R} \frac{1}{\phi_G^R(r_1|l_1)}}$$

Now, in order to have $\lambda_2^R(l_1, 1) > \lambda_2^R(l_1, 0)$ then $\phi_G^R(0|l_1) < \phi_G^R(1|l_1)$ must be satisfied. However this is not possible as a 0 report from R implies that R is of a good type, and thus $\phi_G^R(0|l_1) > \phi_G^R(1|l_1)$ always.

Hence, $\lambda_2^R(l_1, 0) \geq \lambda_2^R(l_1, 1)$ .The one strict inequality comes from the fact that if $\lambda_2^R(0,0) = \lambda_2^R(0,1)$ and $\lambda_2^R(1,0) = \lambda_2^R(1,1)$ then the bad R will have a strict incentive to choose 1 which leads to a babbling equilibrium ■

Proof 2. I look know at what is the effect on R's reputation of L changing her report from 0 to 1.

Now,

$\lambda_2^R(l_1, r_1) = \frac{\lambda_1^R \phi_G^R(r_1|l_1)}{\lambda_1^R \phi_G^R(r_1|l_1) + \left(1-\lambda_1^R\right)\phi_B^R(r_1|l_1)}$ where

$\phi_k^R(r_1|l_1) = \phi_k^R(r_1|x=0) + \left[\phi_k^R(r_1|x=1) - \phi_k^R(r_1|x=0)\right] \Pr(x=1|l_1)$

I will denote with $A_k = \phi_k^R(r_1|x=0) \geq 0$ and $B_k = \left[\phi_k^R(r_1|x=1) - \phi_k^R(r_1|x=0)\right]$ and $X = \Pr(x=1|l_1)$

$$\lambda_2^R(X) = \frac{\lambda_1^R [A_G + B_G X]}{\lambda_1^R [A_G + B_G X] + \left(1-\lambda_1^R\right)[A_B + B_B X]}$$

or

$$\lambda_2^R(X) = \frac{1}{1 + \frac{1-\lambda_1^R}{\lambda_1^R} \frac{A_B + B_B X}{A_G + B_G X}} = \frac{1}{1 + \frac{1-\lambda_1^R}{\lambda_1^R} f(X)}$$

thus

$$\frac{d\lambda_2^R(X)}{dX} = -\frac{1}{\left(1 + \frac{1-\lambda_1^R}{\lambda_1^R}\frac{A_B + B_B X}{A_G + B_G X}\right)^2} \frac{1-\lambda_1^R}{\lambda_1^R} \frac{df(X)}{dx}$$

$\frac{df(X)}{dx} = \frac{B_B}{A_G + B_G X} - \frac{A_B + B_B X}{(A_G + B_G X)^2} B_G = \frac{B_B A_G + B_B B_G X - A_B B_G - B_B B_G X}{(A_G + B_G X)^2} = \frac{B_B A_G - A_B B_G}{(A_G + B_G X)^2}$

which is positive if $B_B A_G - A_B B_G \geq 0$

Returning to the original notations this means $\frac{\phi_B^R(r_1|x=0)}{\phi_G^R(r_1|x=0)} \geq \frac{\phi_B^R(r_1|x=1)}{\phi_G^R(r_1|x=1)}$. However this is always true as long as $\pi_B^R(1) \geq \pi_G^R(1)$ and $\pi_B^R(0) \geq \pi_G^R(0)$ which is implied by point 1.

Let's look now at $\Pr(x=1|l_1)$ when l_1 changes from 1 to 0.

I use $\Pr(x=1|l_1=0) = \frac{\Pr(l_1=0|x=1)}{\Pr(l_1=0|x=1)+\Pr(l_1=0|x=0)} = \frac{1-\Pr(l_1=1|x=1)}{2-\Pr(l_1=1|x=1)-\Pr(l_1=1|x=0)}$ and $\Pr(x=1|l_1=1) = \frac{\Pr(l_1=1|x=1)}{\Pr(l_1=1|x=1)+\Pr(l_1=1|x=0)}$.

Further $\Pr(x=1|l_1=1) \geq \Pr(x=1|l_1=0)$ can be written as $\Pr(l_1=1|x_1=1) \geq \Pr(l_1=1|x_1=0)$ or $\Pr(l_1=0|x_1=0) \geq \Pr(l_1=0|x_1=1)$ as the states are equally likely.

By direct substitution I find the cut-off point :

$$\bar{\lambda}_1^L = \frac{\pi_B^L(0) - \pi_B^L(1)}{\pi_G^L(0) + \pi_B^L(1) - \pi_B^L(0)}$$

such that if $\lambda_1^L \geq \bar{\lambda}_1^L$ iff $\Pr(x=1|l_1=1) \geq \Pr(x=1|l_1=0)$ or $\Pr(l_1=1|x_1=1) \geq \Pr(l_1=1|x_1=0)$ or $\Pr(l_1=0|x_1=0) \geq \Pr(l_1=0|x_1=1)$

Iff $\lambda_1^L < \bar{\lambda}_1^L$ then $\Pr(x=1|l_1=1) < \Pr(x=1|l_1=0)$ or $\Pr(l_1=1|x_1=1) < \Pr(l_1=1|x_1=0)$ or $\Pr(l_1=0|x_1=0) < \Pr(l_1=0|x_1=1)$. ■

7.9 Proof of Proposition 5

Let's assume that this equilibrium exists: $\pi_G^R(1) = 1$; $\pi_G^R(0) = 0$ and $\pi_G^L(1) = 0$; $\pi_G^L(0) = 1$. As we have seen in the claim above it cannot be the case that the bad expert also tells the truth always. In any informative equilibrium we know that the posterior reputation of an R expert after announcing 0 must be higher irrespective of the L's report. Thus $\lambda_2^R(1,0) \geq \lambda_2^R(1,1)$ and $\lambda_2^R(0,0) \geq \lambda_2^R(0,1)$ with a strict inequality, translates into $\pi_B^R(1) \geq \pi_G^R(1)$ and $\pi_B^R(0) \geq \pi_G^R(0)$ with one strict inequality. But if good R tells the truth in equilibrium this implies $\pi_B^R(1) = 1$ and $\pi_B^R(0) > 0$.

Next we look for the equilibrium strategies of the bad experts. The expected current benefit from lying of a bad R when her private signal is 0 is:

$$\Pi_B^R\left(l_1 = i, s_1^R = 0\right) = \mu_B^R(a_1^*(i,1)) - \mu_B^R(a_1^*(i,0))$$

While her reputation cost of lying when observing signal 0 is:

$$\Pi R_B^R\left(l_1 = i, s_1^R = 0\right) = v_R^B(l_1 = i, r_1 = 0) - v_R^B(l_1 = i, r_1 = 1)$$

Her equilibrium strategy $\pi_B^R(0)$ is determined by the indifference condition between the current benefit versus the future reputational costs taking into account what L does:

$\sum_{i=0}^{1}\left[\left(1-p^R\right)\Pr(l_1 = i|x_1 = 1) + p^R \Pr(l_1 = i|x_1 = 0)\right]\Pi_B^R\left(l_1 = i, s_1^R = 0\right) =$
$\sum_{i=0}^{1}\left[\left(1-p^R\right)\Pr(l_1 = i|x_1 = 1) + p^R \Pr(l_1 = i|x_1 = 0)\right]\Pi R_B^R\left(l_1 = i, s_1^R = 0\right)$

In the same time a bad L equilibrium strategy $\pi_B^L(1)$ is determined by the indifference condition between L's expected current benefit and her expected future cost:

$\sum_{i=0}^{1}\left(1-p^L\right)\Pr(r_1 = i|x_1 = 0) + p^L \Pr(r_1 = i|x_1 = 1)\,\Pi_B^L\left(r_1 = i, s_1^L = 1\right) =$
$\sum_{i=0}^{1}\left(1-p^L\right)\Pr(r_1 = i|x_1 = 0) + p^L \Pr(r_1 = i|x_1 = 1)\,\Pi R_B^L\left(r_1 = i, s_1^L = 1\right)$

If RB's current benefits from lying are greater or equal to her future reputation costs for all $\pi_B^R(0) \in (0,1]$ then the optimum strategy is $\pi_B^R(0) = 1$. Similarly if LB's current benefits from lying are greater or equal to her future reputation costs for all $\pi_B^L(1) \in (0,1]$ then the optimum strategy is $\pi_B^L(1) = 1$.

The optimal values $\pi_B^{*R}(0)$ and $\pi_B^{*L}(1)$ are determined simultaneous.

$\pi_B^{*R}(0)$ and $\pi_B^{*L}(1)$ are *unique* as the RB's overall net current benefit from lying denoted as NBL_{RB}:

$\sum_{i=0}^{1}\left[\left(1-p^R\right)\Pr(l_1 = i|x_1 = 1) + p^R \Pr(l_1 = i|x_1 = 0)\right]\Pi_B^R\left(l_1 = i, s_1^R = 0\right) -$
$\sum_{i=0}^{1}\left[\left(1-p^R\right)\Pr(l_1 = i|x_1 = 1) + p^R \Pr(l_1 = i|x_1 = 0)\right]\Pi R_B^R\left(l_1 = i, s_1^R = 0\right)$ strictly decreases with $\pi_B^R(0)$ keeping $\pi_B^L(1)$ fixed.

This is due to the fact that the first term in NBL_{RB} is strictly decreasing in

$\pi_B^R(0)$ as $\frac{d\Pi_B^R\left(l_1=i,s_1^R=0\right)}{\pi_B^R(0)} < 0$ while $\Pi R_B^R\left(l_1=i,s_1^R=0\right)$ is

strictly increasing in $\pi_B^R(0)$.

While the first derivative is straightforward, the second result is due to the fact that

$$\frac{d\Pi R_B^R\left(l_1=i,s_1^R=0\right)}{\pi_B^R(0)} = \underbrace{\frac{dv_R^B(l_1,r_1=0)}{d\lambda_2^R(l_1,r_1=0)}}_{>0}\underbrace{\frac{d\lambda_2^R(l_1,r_1=0)}{d\pi_B^R(0)}}_{<0} - \underbrace{\frac{dv_R^B(l_1,r_1=1)}{d\lambda_2^R(l_1,r_1=1)}}_{>0}\underbrace{\frac{d\lambda_2^R(l_1,r_1=1)}{d\pi_B^R(0)}}_{>0} +$$
$$+\underbrace{\frac{dv_R^B(l_1,r_1=0)}{d\lambda_2^L(l_1,r_1=0)}}_{>0}\underbrace{\frac{d\lambda_2^L(l_1,r_1=0)}{d\pi_B^R(0)}}_{<0} - \underbrace{\frac{dv_R^B(l_1,r_1=1)}{d\lambda_2^L(l_1,r_1=1)}}_{>0}\underbrace{\frac{d\lambda_2^L(l_1,r_1=1)}{d\pi_B^R(0)}}_{>0}.$$

It's important to see that $\frac{d\lambda_2^L(l_1,r_1=1)}{d\pi_B^R(0)} > 0$ as $\frac{d\lambda_2^L(l_1,r_1)}{d\Pr(x=0|r)} > 0$ and $\frac{d\Pr(x_1=0|r_1=1)}{d\pi_B^R(0)} = \frac{\left(1-\lambda_1^R\right)\left(2p^R-1\right)}{\left(1+\left(1-\lambda_1^R\right)\pi_B^R(0)\right)^2} > 0$; thus $\frac{dv_R^B(l_1,r_1=1)}{d\lambda_2^R(l_1,r_1=1)}\frac{d\lambda_2^R(l_1,r_1=1)}{d\pi_B^R(0)} > 0$

Also, $\frac{d\lambda_2^R(l_1,r_1=0)}{d\pi_B^R(0)} < 0$ as as $\frac{d\lambda_2^L(l_1,r_1)}{d\Pr(x=0|r)} > 0$ and $\frac{d\Pr(x_1=0|r_1=0)}{d\pi_B^R(0)} < 0$.

Thus $\frac{dv_R^B(l_1,r_1=0)}{d\lambda_2^R(l_1,r_1=0)}\frac{d\lambda_2^R(l_1,r_1=0)}{d\pi_B^R(0)} < 0$.

I follow a similar procedure in order to show $\frac{dv_R^B(l_1,r_1=0)}{d\lambda_2^L(l_1,r_1=0)}\frac{d\lambda_2^L(l_1,r_1=0)}{d\pi_B^R(0)} < 0$ and $\frac{dv_R^B(l_1,r_1=1)}{d\lambda_2^L(l_1,r_1=1)}\frac{d\lambda_2^L(l_1,r_1=1)}{d\pi_B^R(0)} > 0$

Similarly, keeping $\pi_B^R(0)$ fixed LB's net current benefit from lying, denoted as (NBL_{LB}):

$\sum_{i=0}^{1}\left(1-p^L\right)\Pr(r_1=i|x_1=0) + p^L\Pr(r_1=i|x_1=1)\,\Pi_B^L\left(r_1=i,s_1^L=1\right) -$

$\sum_{i=0}^{1}\left(1-p^L\right)\Pr(r_1=i|x_1=0)+p^L\Pr(r_1=i|x_1=1)\,\Pi R_B^L\left(r_1=i,s_1^L=1\right)$ also strictly decreases with $\pi_B^L(1)$.keeping $\pi_B^R(0)$ fixed.

The monotonic properties of these functions imply that $\pi_B^R(0)$ and $\pi_B^L(1)$ strategies - solutions to the indifference conditions, are unique for a fixed counterpart strategy. Furthermore, this translates into unique solutions to the above system of two (different) equations with two unknowns. $\pi_B^{*R}(0)$ and $\pi_B^{*L}(1)$ are parameter $(\lambda_1^R, \lambda_1^L)$ sensitive.

Further, I look under what conditions the good experts optimally decides to tell the truth. We saw in Proposition 2 that a good R always tells the truth when she receives signal 0 while a good L tells the truth when the signal received is 1. This is due to the fact that they get an improvement in their reputation with no additional current cost. The problem arises when R's private signal is 1 and when L's private signal is 0. For having a truthtelling equilibrium i.e. $\pi_G^R(1) = 1$ and $\pi_G^L(0) = 1$ in which no political correctness takes place then the experts' net current benefits from telling the truth have to be positive for both the good experts

R's current benefit from telling the truth when $s_1^R = 1$ is

$$\Pi_G^R\left(l_1 = i, s_1^R = 1\right) = \mu^{RG}\left[u_G^R\left(i, 1, s_1^R = 1\right) - u_G^R\left(i, 0, s_1^R = 1\right)\right]$$

while her reputation cost from reporting 1 is

$$\Pi R_G^R\left(l_1 = i, s_1^R = 1\right) = v_R^G\left(l_1 = i, r_1 = 0\right) - v_R^G\left(l_1 = i, r_1 = 1\right)$$

L's current benefit from telling the truth when $s_1^L = 0$ is

$$\Pi_G^L\left(r_1 = i, s_1^L = 0\right) = \mu^{LG}\left[u_G^L\left(i, 0, s_1^L = 0\right) - u_G^R\left(i, 1, s_1^L = 0\right)\right]$$

while her reputation cost from reporting $l_1 = 0$ is

$$\Pi R_G^L\left(r_1 = i, s_1^L = 0\right) = v_R^G\left(l_1 = 1, r_1 = i\right) - v_R^G\left(l_1 = 0, r_1 = i\right)$$

It is necessary and sufficient that the current gain of telling the truth is greater or equal with the future reputation cost of telling the truth (for both R and L). For any parameter $(\lambda_1^R, \lambda_1^L)$ we can find thresholds for the time preference parameters $\bar{\mu}^{RG}$ and $\bar{\mu}^{LG}$ such that $\mu^{RG} > \bar{\mu}^{RG}$ and $\mu^{LG} > \bar{\mu}^{LG}$ the truthtelling inequalities hold and there exists an equilibrium in which the good experts always tell the truth. $\bar{\mu}^{RG}$ and $\bar{\mu}^{LG}$ are found as solutions to the system of equations:

$$\sum_{i=0}^{1}\left[\left(1 - p^R\right)\Pr\left(l_1 = i | x_1 = 0\right) + p^R \Pr\left(l_1 = i | x_1 = 1\right)\right]\Pi_G^R\left(l_1 = i, s_1^R = 1\right) =$$
$$\sum_{i=0}^{1}\left[\left(1 - p^R\right)\Pr\left(l_1 = i | x_1 = 0\right) + p^R \Pr\left(l_1 = i | x_1 = 1\right)\right]\Pi R_B^R\left(l_1 = i, s_1^R = 1\right);$$
$$\sum_{i=0}^{1}\left(1 - p^L\right)\Pr\left(r_1 = i | x_1 = 1\right) + p^L \Pr\left(r_1 = i | x_1 = 0\right)\Pi_G^L\left(r_1 = i, s_1^L = 0\right) =$$
$$\sum_{i=0}^{1}\left(1 - p^L\right)\Pr\left(r_1 = i | x_1 = 1\right) + p^L \Pr\left(r_1 = i | x_1 = 0\right)\Pi R_G^L\left(r_1 = i, s_1^L = 0\right).$$

7.10 Proof of Proposition 6

To determine the equilibrium existence in the general case I follow the same procedure as in the truthtelling equilibrium with the difference that I capture both the bad experts' discipline effect but also the good experts' political correctness.

The good experts report always truthfully when their signal is opposite their possible bias. However a different situation arrises when they get a signal similar to their bias. Let's look first at the good R optimal strategy when $s_1^R = 1$.

R's current benefit from telling the truth when $s_1^R = 1$ is

$$\Pi_G^R\left(l_1 = i, s_1^R = 1\right) = \mu^{RG}\left[\hat{u}_G^R\left(i, 1, s_1^R = 1\right) - \hat{u}_G^R\left(i, 0, s_1^R = 1\right)\right]$$

while her reputation cost from reporting 1 is

$$\Pi R_G^R\left(l_1 = i, s_1^R = 1\right) = v_R^G\left(l_1 = i, r_1 = 1\right) - v_R^G\left(l_1 = i, r_1 = 0\right)$$

So, there always exists a threshold $\bar{\mu}^{RG} \in \mathbb{R}^+$ such that:
$\sum_{i=0}^{1}\left[\left(1 - p^R\right)\Pr\left(l_1 = i | x_1 = 0\right) + p^R \Pr\left(l_1 = i | x_1 = 1\right)\right]\Pi_G^R\left(l_1 = i, s_1^R = 1\right) =$
$\sum_{i=0}^{1}\left[\left(1 - p^R\right)\Pr\left(l_1 = i | x_1 = 0\right) + p^R \Pr\left(l_1 = i | x_1 = 1\right)\right]\Pi R_G^R\left(l_1 = i, s_1^R = 1\right)$ since μ^{RG} is just a linear operator. This means that an R expert of good type is indifferent between reporting her signal or not. As her current payoff is strictly increasing in μ^{RG}, if $\mu^{RG} > \bar{\mu}^{RG}$ R always tells the truth as the current benefit is higher than future reputation costs for all $\pi_G^R(1) \in [0, 1]$ thus $\pi_G^{*R}(1) = 1$. If $\mu^{RG} \leq \bar{\mu}^{RG}$ the good R does not always tell the truth. In order to find good R equilibrium strategy we fix μ^{RG} to a value below the threshold and we calculate $\pi_G^{*R}(1) \in [0, 1)$ which makes a good R indifferent between telling the truth and not.

The same reasoning applies to the L expert of a good type:

If a good expert L receives signal $s_1^L = 0$ there always exists a threshold $\bar{\mu}^{LG} \in \mathbb{R}^+$ such that:
$\sum_{i=0}^{1}\left[\left(1 - p^L\right)\Pr\left(l_1 = i | x_1 = 1\right) + p^L \Pr\left(r_1 = i | x_1 = 0\right)\right]\Pi_G^L\left(r_1 = i, s_1^L = 0\right) =$

$$\sum_{i=0}^{1}\left[\left(1-p^{L}\right)\Pr\left(l_{1}=i|x_{1}=1\right)+p^{L}\Pr\left(r_{1}=i|x_{1}=0\right)\right]\Pi R_{G}^{L}\left(r_{1}=i,s_{1}^{1}=0\right)$$

which makes the good L expert indifferent between reporting her signal or not. As her current payoff is strictly increasing in μ^{LG}, if $\mu^{LG}>\bar{\mu}^{LG}$ there is always truthtelling in equilibrium and $\pi_{G}^{*L}(0)=1$. If $\mu^{LG}\leq\bar{\mu}^{LG}$ the good L does not always tell the truth. $\pi_{G}^{*L}(0)\in[0,1)$ is determined by the above indifference condition.

As before the bad expert equilibrium strategies are determined at the indifference conditions between current benefits versus future reputation costs.

Bad R's equilibrium strategy when her signal is 0, $\pi_{B}^{*R}(0)$ is determined by:

$$\sum_{i=0}^{1}\left[\left(1-p^{R}\right)\Pr\left(l_{1}=i|x_{1}=1\right)+p^{R}\Pr\left(l_{1}=i|x_{1}=0\right)\right]\Pi_{B}^{R}\left(l_{1}=i,s_{1}^{R}=0\right)=$$
$$\sum_{i=0}^{1}\left[\left(1-p^{R}\right)\Pr\left(l_{1}=i|x_{1}=1\right)+p^{R}\Pr\left(l_{1}=i|x_{1}=0\right)\right]\Pi R_{B}^{R}\left(l_{1}=i,s_{1}^{R}=0\right).$$

while $\pi_{B}^{*R}(1)$ is determined by:

$$\sum_{i=0}^{1}\left[\left(1-p^{R}\right)\Pr\left(l_{1}=i|x_{1}=0\right)+p^{R}\Pr\left(l_{1}=i|x_{1}=1\right)\right]\Pi_{B}^{R}\left(l_{1}=i,s_{1}^{R}=1\right)=$$
$$\sum_{i=0}^{1}\left[\left(1-p^{R}\right)\Pr\left(l_{1}=i|x_{1}=0\right)+p^{R}\Pr\left(l_{1}=i|x_{1}=1\right)\right]\Pi R_{B}^{R}\left(l_{1}=i,s_{1}^{R}=1\right).$$

When the bad expert L receives signal 0 the equilibrium strategy $\pi_{B}^{*L}(0)$ is determined by:

$$\sum_{i=0}^{1}\left[\left(1-p^{L}\right)\Pr\left(l_{1}=i|x_{1}=1\right)+p^{L}\Pr\left(r_{1}=i|x_{1}=0\right)\right]\Pi_{B}^{L}\left(r_{1}=i,s_{1}^{L}=0\right)=$$
$$\sum_{i=0}^{1}\left[\left(1-p^{L}\right)\Pr\left(l_{1}=i|x_{1}=1\right)+p^{L}\Pr\left(r_{1}=i|x_{1}=0\right)\right]\Pi R_{B}^{L}\left(r_{1}=i,s_{1}^{1}=0\right).$$

and $\pi_{B}^{*L}(1)$ is determined by:

$$\sum_{i=0}^{1}\left[\left(1-p^{L}\right)\Pr\left(l_{1}=i|x_{1}=0\right)+p^{L}\Pr\left(r_{1}=i|x_{1}=1\right)\right]\Pi_{B}^{L}\left(r_{1}=i,s_{1}^{L}=1\right)=$$
$$\sum_{i=0}^{1}\left[\left(1-p^{L}\right)\Pr\left(l_{1}=i|x_{1}=0\right)+p^{L}\Pr\left(r_{1}=i|x_{1}=1\right)\right]\Pi R_{B}^{L}\left(r_{1}=i,s_{1}^{1}=1\right).$$

The equilibrium strategies will be determined simultaneously by the above six indifference conditions..

The non-informative equilibrium arrises in the situation in which political correctness takes full hold of good experts behavior. As a result no expert ever declares her possible bias. The lower weight bounds $\underline{\mu}^{RG}$ and $\underline{\mu}^{LG}$ which trigger this type of non-informative equilibrium are given by the indifference conditions:

$$\sum_{i=0}^{1}\left[\left(1-p^{R}\right)\Pr\left(l_{1}=i|x_{1}=0\right)+p^{R}\Pr\left(l_{1}=i|x_{1}=1\right)\right]\Pi_{G}^{R}\left(l_{1}=i,s_{1}^{R}=1\right)=$$

$\sum_{i=0}^{1}\left[\left(1-p^{R}\right)\Pr\left(l_{1}=i|x_{1}=0\right)+p^{R}\Pr\left(l_{1}=i|x_{1}=1\right)\right]\Pi R_{G}^{R}\left(l_{1}=i,s_{1}^{R}=1\right)$
evaluated at babbling strategies $\pi_{G}^{R}(0)=\pi_{G}^{R}(1)=\pi_{B}^{R}(1)=\pi_{B}^{R}(0)=\frac{1}{2}$ and
$\sum_{i=0}^{1}\left[\left(1-p^{L}\right)\Pr\left(l_{1}=i|x_{1}=1\right)+p^{L}\Pr\left(r_{1}=i|x_{1}=0\right)\right]\Pi_{G}^{L}\left(r_{1}=i,s_{1}^{L}=0\right)=$
$\sum_{i=0}^{1}\left[\left(1-p^{L}\right)\Pr\left(l_{1}=i|x_{1}=1\right)+p^{L}\Pr\left(r_{1}=i|x_{1}=0\right)\right]\Pi R_{G}^{L}\left(r_{1}=i,s_{1}^{1}=0\right)$
evaluated at babbling strategies $\pi_{G}^{L}(0)=\pi_{G}^{L}(1)=\pi_{B}^{L}(1)=\pi_{B}^{L}(0)=\frac{1}{2}$.

7.11 Proof of Proposition 7

Proof 1. The first proof follows the same logic of the proof in the original case

I will prove the first point by contradiction.

Suppose not and $\lambda_{2}^{R}(1)>\lambda_{2}^{R}(0)$; in this situation a bad R has a both a higher reputation by declaring 1 and a higher current payoff for any $s_{1}^{R}=\{0,1\}$; thus the biased R will always say 1 and $\pi_{B}^{R}(0)=\pi_{B}^{R}(1)=1$ resulting in $\phi_{B}^{R}(0)=\phi_{B}^{R}(1)=1$. Then,

$$\lambda_{2}^{R}(r_{1})=\frac{1}{1+\frac{1-\lambda_{1}^{R}}{\lambda_{1}^{R}}\frac{1}{\phi_{G}^{R}(r_{1})}}$$

Now, in order to have $\lambda_{2}^{R}(1)>\lambda_{2}^{R}(0)$ then $\phi_{G}^{R}(0)<\phi_{G}^{R}(1)$ must be satisfied. However this is not possible as a 0 report from R implies that R is of a good type, and thus $\phi_{G}^{R}(0)>\phi_{G}^{R}(1)$ always.

Hence, $\lambda_{2}^{R}(0)>\lambda_{2}^{R}(1)$. ■

Proof 2. Now,

$$\lambda_{2}^{R}(r_{1})=\frac{\lambda_{1}^{R}\phi_{G}^{R}(r_{1})}{\lambda_{1}^{R}\phi_{G}^{R}(r_{1})+\left(1-\lambda_{1}^{R}\right)\phi_{B}^{R}(r_{1})}$$

where

$$\phi_{k}^{R}(r_{1})=\phi_{k}^{R}(r_{1}|x=0)+\left[\phi_{k}^{R}(r_{1}|x=1)-\phi_{k}^{R}(r_{1}|x=0)\right](1-\tau)$$

As long as $\frac{\phi_{B}^{R}(r_{1}|x=0)}{\phi_{G}^{R}(r_{1}|x=0)}\geq\frac{\phi_{B}^{R}(r_{1}|x=1)}{\phi_{G}^{R}(r_{1}|x=1)}$, which is implied by Proposition 7.1, $\frac{d\lambda_{2}^{R}(r_{1})}{d\tau}<0$.

■

7.12 Proof of Proposition 8

Proof 2. Now,

$$\lambda_2^R(l_1, r_1) = \frac{\lambda_1^R \phi_G^R(r_1|l_1)}{\lambda_1^R \phi_G^R(r_1|l_1) + \left(1 - \lambda_1^R\right) \phi_B^R(r_1|l_1)}$$

where

$$\phi_k^R(r_1) = \phi_k^R(r_1|x=0) + \left[\phi_k^R(r_1|x=1) - \phi_k^R(r_1|x=0)\right] \Pr(x_1 = 1|l_1)$$

As long as $\frac{\phi_B^R(r_1|l_1=0)}{\phi_G^R(r_1|l_1=0)} \geq \frac{\phi_B^R(r_1|l_1=1)}{\phi_G^R(r_1|l_1=1)}$, which is implied by Proposition 8.1, $\lambda_2^R(1, r_1) \geq \lambda_2^R(0, r_1)$ if $\Pr(x_1 = 1|l_1 = 1) \geq \Pr(x_1 = 1|l_1 = 0)$. But as L is not biased

$$\Pr(x_1 = 1|l_1 = 1) = \frac{p^L(1-\tau)}{p^L(1-\tau) + (1-p^L)\tau}$$

$$\Pr(x_1 = 1|l_1 = 0) = \frac{\left(1-p^L\right)(1-\tau)}{(1-p^L)(1-\tau) + p^L\tau}$$

if we also allow for a prior on the state - $\Pr(x_1 = 0) = \tau$. But if $p^L \geq \frac{1}{2}$ then $\Pr(x_1 = 1|l_1 = 1) \geq \Pr(x_1 = 1|l_1 = 0)$ always and thus conformity as separation disappears. ■

Chapter 2

Conforming to Stand Out: Comparative Statics

1 Comparative Statics

In this chapter I look at how the experts' optimal strategies change with their own and counterpart's *time preferences, initial reputations and signal precisions.*

The change with respect to the time preference parameters is derived analytically but I also provide a numerical example.

The effect of a change in initial reputations and signal precisions are explained with numerical examples.

First, I look at the implication for the truthtelling equilibrium. As the good experts always tell the truth in a truthtelling equilibrium, these implications are derived for to the bad expert equilibrium behavior.

1.1 Bad Experts' Optimal Strategies: Comparative Statics

First I analyze the effect of a *change in experts' time preferences* on the bad experts' optimal strategies. The biased experts' optimal strategies when their signals are opposite their potential biases are determined by the indifference conditions between telling the truth versus lying. The following result is obtained by applying multivariate implicit function theorem on these indifference conditions.

Result 3 *The change of the optimal strategies with the parameters of time preference is*

$$\frac{d\pi_B^i\left(1-b_i\right)}{d\mu_B^i} > 0 \text{ and } \frac{d\pi_B^i\left(1-b_i\right)}{d\mu_B^{\bar{i}}} > 0$$

The first inequality conveys the fact that if a bad expert values the future less she lies more and vice-versa - she lies less if she values the future more. Furthermore this translates in an overall discipline effect for the bad experts: an bad expert lie less as she cares more about the future, while the counterpart also lies less as she cannot afford to loose further reputation in the future.

I look next at numerical example with initial parameters: $\lambda_1^R = \lambda_1^L = 0.6$, $p^R = p^L = 0.75$. I allow the bad L expert to vary his time preference parameter μ_1^{LB} from 0.1 to 0.25.

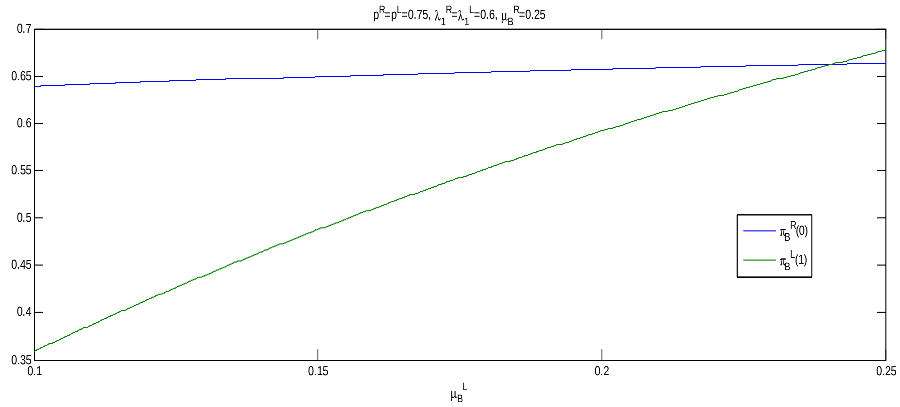


Fig. 1: Equilibrium strategies at different time preferences

Fig. 1 plots the optimal strategies of the bad experts at these particular parameter values. We observe that once L values the future more (at lower values of μ_1^{LB}) she lie less in equilibrium and so does her counterpart R. This result has an important applicability in the sense that a decision maker could make two experts (which he suspects of being of bad type) tell the truth in equilibrium not necessarily by disciplining both of them but by disciplining just one.

I also look next (through a numerical example) at how the optimal strategies of the bad experts change with the initial reputation of L. Initially, I allow for L's initial reputation to vary monotonically from 0 to 1 while R's reputation is fixed at 0.6. The other parameters take values: $p^R = p^L = 0.75$, $\mu^{RB} = \mu^{LB} = 0.25$.

If *L's initial reputation λ_1^L increases*, (Fig. 2a) a bad L lies more in equilibrium as a higher reputation makes her more believable in the eyes of the decision maker. As a result more 0 reports from L will be seen in equilibrium. At the same time a bad R lies less in equilibrium as L's reputation increases; this is due to R's realization that D takes both reports into account when he decides on his action. A report of 1 from R might not have much weight in her current payoff but a report of 0 significantly increases her next period payoff. Thus it is better for R to lie less once L's reputation increases.

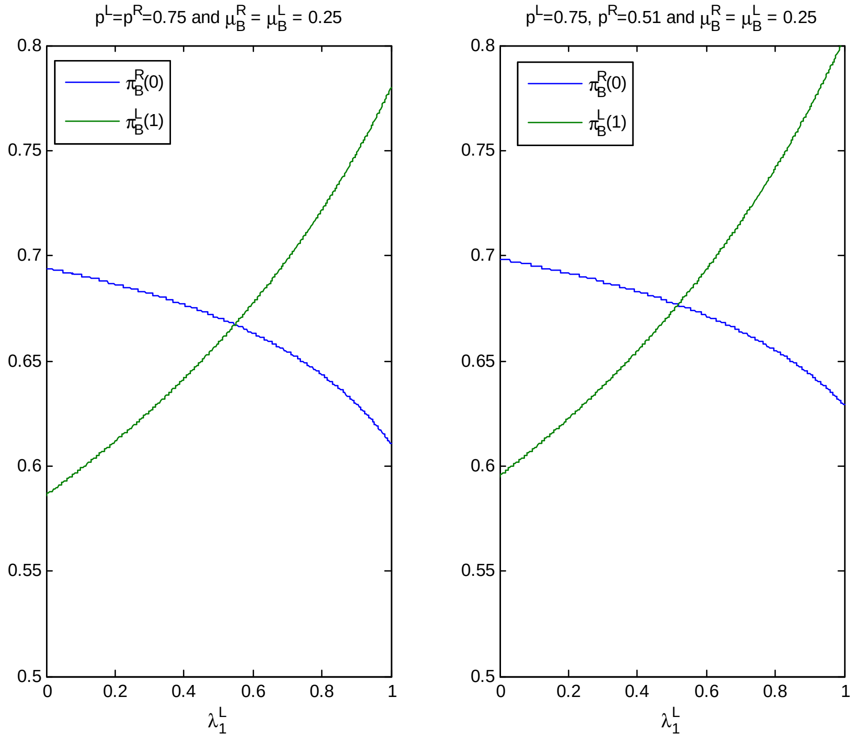


Fig. 2: Equilibrium strategies at different initial reputation levels

Next, I change the signal precision for R monotonically from $p^R = 0.51$ to 0.99 We see in Fig. 3. that a bad R expert now lies more as she realizes that the decision maker accounts for her less precise signal when observing a 1 report from her. This in return implies a lesser ability of D to differentiate between a bad and good expert and thus R lies more. If R has a lower precision than L, a bad L also lies more in equilibrium as she accounts for the fact that the decision maker understands that a higher precision of L's signal implies a more accurate report.

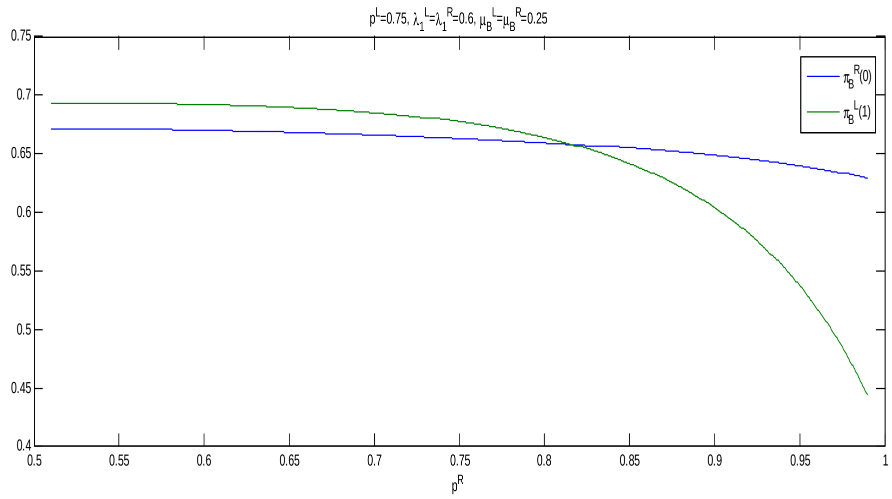


Fig. 3: Equilibrium strategies at different precision level

Thus, lower signal precision translates in higher inability of the decision maker to differentiate between a good expert that tells the truth versus a bad expert that lies; thus a bad experts is able to lie more.

In the above discussion we assumed that the good experts tell the truth in equilibrium. However, the *good experts* also have incentives to misreport their signals when the signals are their perceived biases. Next I look at the implications of changes of parameters for the good experts' behavior in equilibrium.

1.2 Good Experts Optimal Strategies: Comparative Statics

In Chapter 1 I found and characterized the equilibrium of the model. The truthtelling equilibrium was determined by computing the optimal weights $\bar{\mu}^{RG}$ and $\bar{\mu}^{LG}$ which made the good experts tell the truth with probability 1 when their signals were their potential biases. Once the good experts value the future more they start distorting their reports for fear of being identified as biased. Thus for $\mu^{RG} \leq \bar{\mu}^{RG}$ or/and $\mu^{LG} \leq \bar{\mu}^{LG}$ the equilibrium will be just informative.

Next, I look at the effect of a *change in experts' time preferences* on the good experts' optimal strategies. These effect is capture trough the change in $\bar{\mu}^{RG}$ and $\bar{\mu}^{LG}$.

In Fig. 4 I present the optimal thresholds above which the good experts disclose truthfully their signal. For values below these weights the good experts lie with some positive probability when the signal is their possible bias. They are calculated at initial parameters: $\lambda_1^R = \lambda_1^L = 0.6$, $p^R = p^L = 0.75$, $\mu^{RB} = \frac{1}{4}$ and μ^{LB} varies from 0.1 to 0.25.

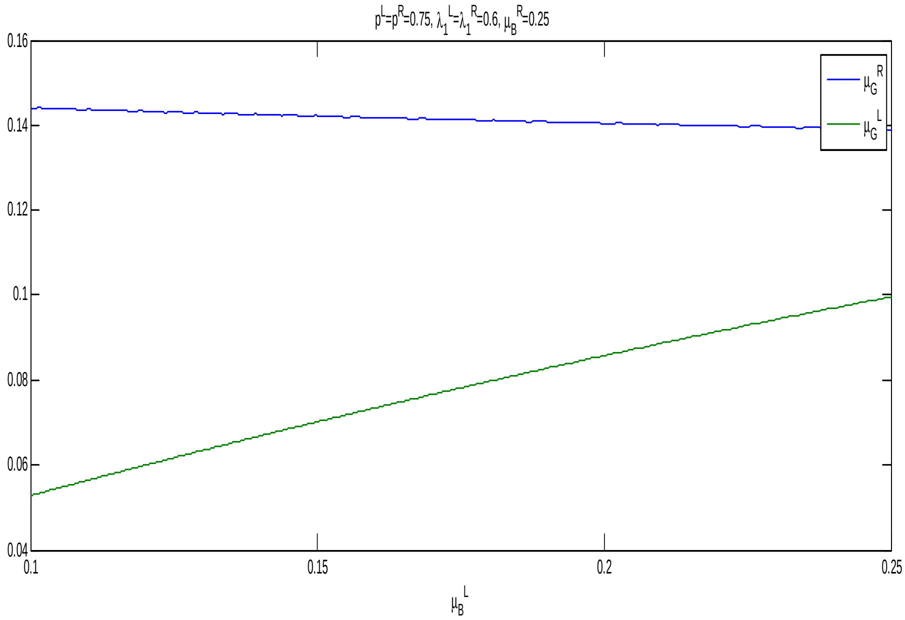


Fig. 4: Optimal time preference parameters for the good experts for μ_B^L

We see that if a bad L puts a weight of 0.1 on the present than a good L needs to put a weight of approximately 0.05 or more on the present in order for a truthtelling equilibrium to exist. However when the bad L starts valuing the present more (i.e. μ^{LB} increases), the optimal threshold $\bar{\mu}^{LG}$ for a truthtelling equilibrium increases as well. Which means that if the good L values the present at the same value as before at $\bar{\mu}^{LG} = 0.05$ the equilibrium is not a truthtelling one anymore, so the good L expert will lie with positive probability when her signal is 0 - her potential bias.

This result is due to fact that the decision maker sees more 0 in equilibrium which he assumes to be from a biased expert as μ^{LB} increases. The good L by reporting 1 shows that she is not biased so there is an extra incentive to report the potential bias so $\pi_G^L(0)$ (probability of lying of a good expert) increases with μ^{LB}.

We see however that $\bar{\mu}^{RG}$ decreases when the bad expert values the present more. The good R tells the truth at lower values of μ^{RG}, so the incentive to lie for reputational reason $(\pi_G^R(1))$ decreases with μ^{LB}. This result is could be due to the fact that the decision maker realizes that the bad L lies more when he values the present more, so he puts more weight on R's report.

In Fig. 5 I represent the optimal thresholds above which the good experts disclose truthfully their signal when *the initial reputation of* L changes. They are calculated at initial parameters: $\lambda_1^R = 0.6$, $p^R = p^L = 0.75$, $\mu^{RB} = \frac{1}{4}$ and $\mu^{LB} = \frac{1}{10}$ while λ_1^L varies from 0 to 1.

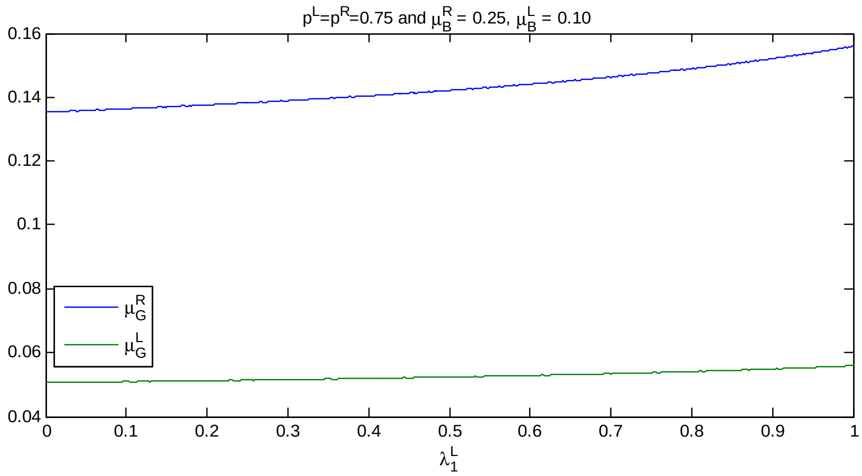


Fig. 5: Optimal time preference parameters for the good experts at different λ_1^L

We see that if the bad L initial reputation increases both $\bar{\mu}^{LG}$ and $\bar{\mu}^{RG}$ increase and thus we observe an extra incentives of lying for the good experts: $\pi_G^L(0)$ and $\pi_G^R(1)$ increase with with λ_1^L.

The reasoning for these results is the fact that if the initial reputation of L increases, then a bad R lies less (when signal 1). As a result, it is harder for the decision maker to differentiate between a bad R and a good expert R. In this case, it is easier for a good R to report 0 when in fact the signal received was 1. This could summarize into the argument that a higher initial reputation for L implies more signal distortion for the good R. So the good R has an extra incentive to lie for reputational reasons i.e. $\pi_G^R(1)$ increases with λ_1^L.

Also if λ_1^L increases, a bad L reports 0 more, so more zeros will be observed in equilibrium. The good L by reporting 1 shows that he is not biased so there is an extra incentive to report against the potential bias so $\pi_G^L(0)$ increases with λ_1^L as well.

Fig. 6 presents the optimal thresholds above which the good experts disclose truthfully their signal when *the signal precision of* R changes. They are calculated at initial parameters: $\lambda_1^R = 0.6$, $p^L = 0.75$, $\mu^{RB} = \mu^{LB} = \frac{1}{4}$ and $\mu^{LB} = \frac{1}{4}$ while p^R varies from 0.51 to 0.99.

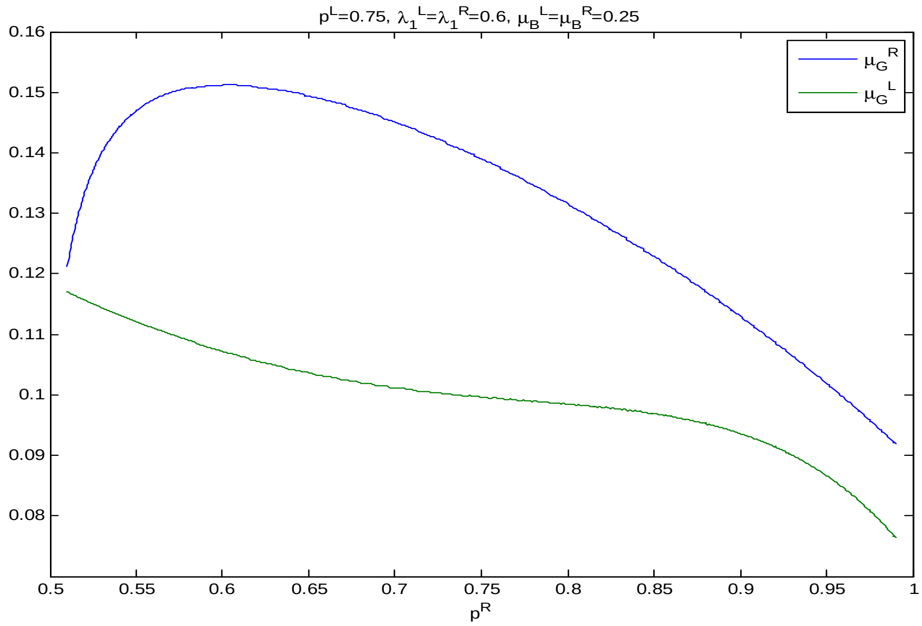


Fig. 6: Optimal time preference parameters for the good experts at different p^R

When R's signal precision increases $\bar{\mu}^{RG}$ increases in p^R, so a good R distorts her signal (i.e. $\pi_G^R(1)$ increases in p^R). After a threshold $\bar{\mu}^{RG}$ decreases with p^R as R does not need to build-up her future reputation anymore as she has a good signal which is known by the decision maker and allows her to focus on the current payoff.

We see that when R's signal precision increases a good L realizes she needs to counteract R's higher reputation triggered by better signal - so $\pi_G^L(0)$ decreases with p^R and $\bar{\mu}^{LG}$ decreases with p^R. Also there is an inflexion point at approximately $p^R = 0.75$: for values of $p^R < p^L = 0.75$, $\bar{\mu}^{LG}$ is convex while afterwards it becomes concave. This captures the fact that the rate of change of signal distortion increases with the counterpart's precision for $p^R < p^L$; once R's signal becomes more precise i.e. $p^R > p^L$ this rate of change decreases with p^R.

2 Conformity as Separation Effect

In the previous chapter we saw that conformity as separation represents an action to disavow one's bias when the counterpart has a reputation below a particular threshold.

For $p^R = p^L = 0.75$, $\mu^{RB} = \mu^{LB} = \frac{1}{4}$, $\lambda_1^R = 0.6$ and $\lambda_1^L \in (0,1)$, I calculate the reputational threshold $\bar{\lambda}_1^L$ based on which R conforms to separate in a truthtelling equilibrium.

Fig. 7 plots L's actual initial reputation against the optimal cut-off point $\bar{\lambda}_1^L$ evaluated at equilibrium strategies. At low levels of L's initial reputation (when $\lambda_1^L < \bar{\lambda}_1^L$) R gets a higher reputation by reporting 0 when L reports 0 rather than when L reports 1 - this is *the conformity as separation effect*

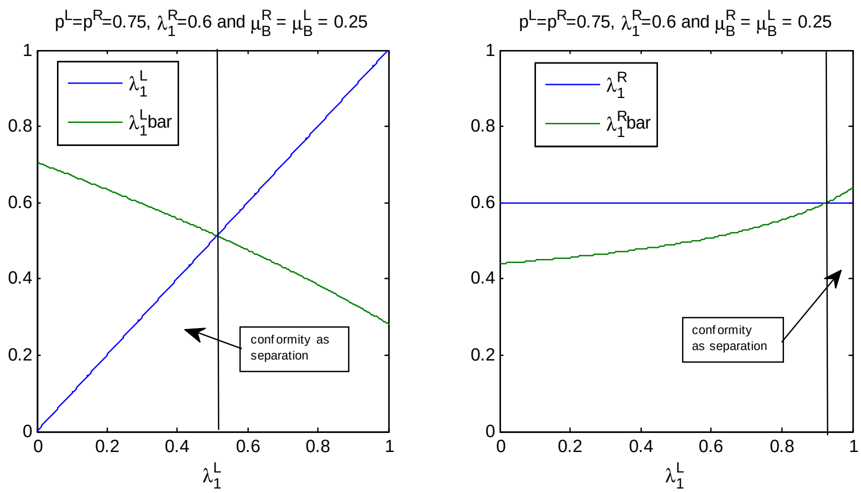


Fig. 7: Conformity as Separation

This effect, however has be analyzed in the context that 0 reports from L are not seen often by the decision maker - as an L with low initial reputation prefers to tell 1 to increase her future reputation.

For the same parameter values (and thus, the same equilibrium strategies) I look for the same effect in L's behavior. R's actual initial reputation is $\lambda_1^R = 0.6$ while the cut-off optimal $\bar{\lambda}_1^R$ varies with λ_1^L as well.

As R's initial reputation is high in comparison with most of L's initial reputation values, L's conformity as separation behavior can be seen only at high levels of λ_1^L.

If $\lambda_1^L = 0.6$ and $\lambda_1^R = 0.6$ the cut-off points $\bar{\lambda}_1^R = \bar{\lambda}_1^L$. At this particular level of experts' initial reputations there is no conformity as separation effect in equilibrium.

3 Welfare Analysis

By increasing the reputation of one expert, we observe two opposite effects on the equilibrium strategies of the bad experts. First, the expert whose reputation increases lies more; however, her counterpart lies less. Hence, an obvious question is what is the predominant effect on the decision maker's payoff. In Figure 8 I plot D's ex-ante first stage expected payoff at parameter values $\lambda_1^R = 0.6$, $p^R = p^L = 0.75$, $\mu^{RB} = \frac{1}{4}$ and $\mu^{LB} = \frac{1}{10}$.

Even though there is a trade-off between one expert being disciplined to tell the truth effect versus the other one lying more, we see that there is a positive overall effect on D's expected payoff in period 1. Hence, by increasing one expert's reputation the positive discipline effect on the counterpart's action overcomes the negative effect of the lying by the first expert.

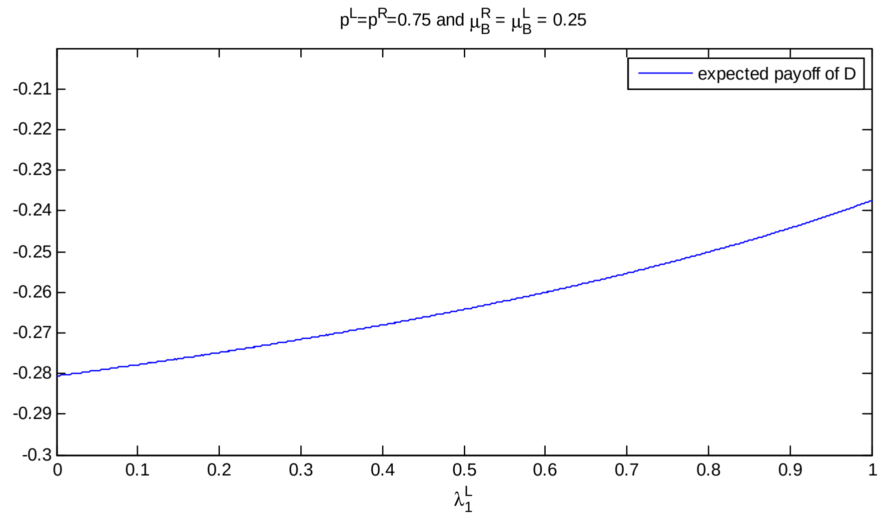


Fig. 8: Decision Maker's Expected Payoff

If we also increase λ_1^R there is an upward shift of D's expected payoff; thus the decision maker would prefer to have both of the experts with reputations as high as possible.

Notice that I look only at the first period decision maker's payoff as the second stage payoff is not affected by experts' career concerns.

In general due to the career concerns of the experts we also see a different trade-off taking place in equilibrium: the bad experts declare their signals with some positive probability, while the good experts misreport their signal with some positive probability. The overall effect on the decision maker's payoff could be either positive or negative and it depends on experts' time preferences.

4 Conclusion

This model attempts to shed light in the efficiency of communication when advice is provided by two experts who care about their future advancement, and have uncertain opposing biases while the decision maker is unable to verifiability the state of the world.

In a truthtelling equilibrium where the good experts are always telling the truth, a decision maker could discipline both experts to disclose their signals, in case he suspects that they are of a bad type by offering making only one expert to value the future more and not to both. Career concerns in this case makes the expert that did not receive any incentive to tell the truth as otherwise she will get a lower pay in the future since the counterpart will become even more trustworthy than her.

Another important result comes from experts' ability to observe the true state. If the precision of their signal drops bad experts are more likely to lie in equilibrium as the decision maker is unable to differentiate between a good expert that got a wrong signal and a bad expert that lies. Also, if the signal of one expert increases, the counterpart (in case she is of a good type) will not distort her signal only as long as her signal precision is higher than her competitor's This perspective should also be taken into account by a decision maker that ranks different experts - for example financial or economic advisers - when the environment is volatile or goes through adjustments.

Another negative effect occurs in an informative equilibrium when a decision maker is able to pick one expert with a higher reputation, then the counterpart (if she is of a good type) tends to misreport her signal in order not to appear biased. As a result, biased information might be transmitted to the decision maker. On the positive side, if the counterpart is of a bad type, she is likely to tell her signal more often than before.

As this model is build on Morris (2001) a pertinent question is whether the decision maker is better off by asking advice from two experts which are from opposing sides. While the results found by Morris still persist: disavowing own bias for reputational reasons - which translates in discipline of bad experts and perverse incentive for lying of good experts, this study brings into attention a further type of distortion due to the strategic interaction of experts - conformity as separation. However, the overall effect of adding of one more expert is not clear, and it depends on the level of career concerns of each expert. While there is a further discipline effect on the bad experts, the negative effects on the good experts persists as they are incentivized to lie even more for reputation reasons.

There are papers related to this model which analyses communication by multiple experts, for example: Austen-Smith (1990, 1993), Krishna and Morgan (2001 a,b), Gilligan and Krehbiel (1989), Wolinsky (2002). These papers build on the seminal paper of Crawford and Sobel (1982) by adding one more expert to the communication. The expert advice could be simultaneous or sequential, experts could have similar biases or not. In these papers however the biases of the experts are common knowledge. Li (2010) adds to the literature by looking at the efficiency of two expert communication when there is asymmetric information about the biases of the experts.

My study contributes to the literature on communication by multiple experts to an uninformed policy maker by analyzing the information transmission by two experts motivated by career concerns. The experts are imperfectly informed and there is uncertainty about their biases. As in the benchmark model - Morris (2001), the factors that can induce different results in the welfare of the decision maker are the career concerns of the experts. The overall effect is determined by the trade-offs implied by the manner in which the experts value the future.

References

[1] Austen-Smith, D., 1990, Information Transmission in Debate, American Journal of Political Science 34, 124-152.

[2] Austen-Smith, D., 1993, Interested Experts and Policy Advice: Multiple Referrals under Open Rule, *Games and Economic Behavior*, 5, 3–43.

[3] Crawford, V. P., and J. Sobel, 1982, Strategic Information Transmission, *Econometrica*, 50, 1431-1451.

[4] Gilligan, T., and K. Krehbiel, 1989, Asymmetric Information and Legislative Rules with a Heterogeneous Committee, *American Journal of Political Science*, 1459-490.

[5] Krishna, V., and J. Morgan, 2001 a, A Model of Expertise, *Quarterly Journal of Economics*, 116, 747-775.

[6] Krishna, V., and J. Morgan 2001 b, The Art of Conversation: Eliciting Information from Experts through Multi-Stage Communication, *American Political Science Review,*95, 435-52.

[7] Li, M., 2010, Advice from Multiple Experts: A Comparison of Simultaneous, Sequential, and Hierarchical Communication, *The B.E. Journal of Theoretical Economics*

[8] Morris, S., 2001, Political Correctness, *Journal of Political Economy*, 109, 231-265.

[9] Wolinsky, A., 2002, Eliciting Information from Multiple Experts, *Games and Economic Behavior*, 41, 141-60.

5 Appendix

While proving uniqueness of the truthtelling equilibrium we saw that the net benefit of lying for one expert decreases in her probability of lying in case she is biased, keeping the counterpart's strategy fixed. $\frac{dNBL_{RB}}{\pi_B^R(0)} < 0$ for any $\pi_B^L(1)$ fixed and similarly $\frac{dNBL_{LB}}{\pi_B^L(1)} < 0$ for any $\pi_B^R(0)$ fixed. This result says that the net benefit of lying for each expert is a monotonic function in own strategy.

In order to be able to look at the comparative statics of the optimal strategies with respect to the parameters of the model, I need to ask first how the net benefit of lying changes with the counterpart probability of lying keeping own strategy fixed$\frac{dNBL_{RB}}{\pi_B^L(1)}$ for any $\pi_B^R(0)$.

I prove next the following intermediary result:

Result 4 *Keeping own strategy fixed, $\pi_B^R(0)$, a bad R net benefit of lying is increasing in the probability of lying of the counterpart (in case she is biased) $\frac{dNBL_{RB}}{\pi_B^L(1)} > 0$ for any $\pi_B^R(0)$.*

As a reminder, NBL_{RB} is:

$$\sum_{i=0}^{1}\left[\left(1-p^R\right)\Pr\left(l_1=i|x_1=1\right)+p^R\Pr\left(l_1=i|x_1=0\right)\right]\Pi_B^R\left(l_1=i,s_1^R=0\right)-$$
$$\sum_{i=0}^{1}\left[\left(1-p^R\right)\Pr\left(l_1=i|x_1=1\right)+p^R\Pr\left(l_1=i|x_1=0\right)\right]\Pi R_B^R\left(l_1=i,s_1^R=0\right)$$

First, I look at the change of the bad R current benefit from lying - the first term of NBL_{RB} - with respect to $\pi_B^L(1)$. The bad R benefit from lying I denote with BL_{RB}. The first term of BL_{RB} (i.e. $l_1=0$) differentiated with respect to $\pi_B^L(1)$ is:

$$\frac{d\left[\left(1-p^R\right)\Pr(l_1=0|x_1=1)+p^R\Pr(l_1=0|x_1=0)\right]}{d\pi_B^L(1)}\Pi_B^R\left(l_1=0,s_1^R=0\right)+$$
$$\left[\left(1-p^R\right)\Pr\left(l_1=0|x_1=1\right)+p^R\Pr\left(l_1=0|x_1=0\right)\right]\frac{d\Pi_B^R\left(l_1=0,s_1^R=0\right)}{d\pi_B^L(1)}.$$

This is positive as:

$$\frac{d\left[\left(1-p^R\right)\Pr(l_1=0|x_1=1)+p^R\Pr(l_1=0|x_1=0)\right]}{d\pi_B^L(1)}=\left(1-p^R\right)p^L\left(1-\lambda_1^L\right)+\left(1-p^L\right)\lambda_1^L>0.$$

The term

$$\frac{d\Pi_B^R\left(l_1=0,s_1^R=0\right)}{d\pi_B^L(1)}=\frac{d\left[\mu_B^R(a_1(0,1))-\mu_B^R(a_1(0,0))\right]}{d\pi_B^L(1)}$$

is positive as well as:

$$\frac{da_1(0,1)}{d\pi_B^L(1)} = -\frac{1}{(\Pr(0,1|1)+\Pr(0,1|0))^2}AB > 0$$

$$\begin{aligned} A &= \left(1-\lambda_1^L\right)\left[\lambda_1^R\phi_G^R(1|1)+\left(1-\lambda_1^R\right)\phi_B^R(1|1)\right] \\ &\quad \left[\lambda_1^R\phi_G^R(1|0)+\left(1-\lambda_1^R\right)\phi_B^R(1|0)\right] \end{aligned}$$

$$\begin{aligned} B &= \left(1-p^L\right)\left[\lambda_1^L\phi_G^L(0|1)+\left(1-\lambda_1^L\right)\phi_B^L(0|1)\right]- \\ &\quad p^L\left[\lambda_1^L\phi_G^L(0|0)+\left(1-\lambda_1^L\right)\phi_B^L(0|0)\right] \end{aligned}$$

.

A is positive while$B$ simplifies at $\left(1-2p^L\right)<0$.

In finding this result I used that

$$\Pr(0,1|0) = \left[\lambda_1^R\phi_G^R(1|0)+\left(1-\lambda_1^R\right)\phi_B^R(1|0)\right]\left[\lambda_1^L\phi_G^L(0|0)+\left(1-\lambda_1^L\right)\phi_B^L(0|0)\right]$$

and

$$\Pr(0,1|1) = \left[\lambda_1^R\phi_G^R(1|1)+\left(1-\lambda_1^R\right)\phi_B^R(1|1)\right]\left[\lambda_1^L\phi_G^L(0|1)+\left(1-\lambda_1^L\right)\phi_B^L(1|1)\right]$$

Similarly,

$$\frac{da_1(0,0)}{d\pi_B^L(1)} = -\frac{1}{(\Pr(0,0|1)+\Pr(0,0|0))^2}AB < 0$$

$$\begin{aligned} A &= \left(1-\lambda_1^L\right)\left[\lambda_1^R\phi_G^R(0|1)+\left(1-\lambda_1^R\right)\phi_B^R(0|1)\right] \\ &\quad \left[\lambda_1^R\phi_G^R(0|0)+\left(1-\lambda_1^R\right)\phi_B^R(0|0)\right] \end{aligned}$$

$$\begin{aligned} B &= p^L \left[\lambda_1^L \phi_G^L (0|1) + \left(1-\lambda_1^L\right) \phi_B^L (0|1)\right] - \\ &\quad \left(1-p^L\right) \left[\lambda_1^L \phi_G^L (0|0) + \left(1-\lambda_1^L\right) \phi_B^L (0|0)\right] \end{aligned}$$

A is positive and B simplifies at $\left(2p^L - 1\right) > 0$.

The change of the second term of BL_{RB} with respect to $\pi_B^L(1)$ is:

$$\begin{aligned} &\frac{d\left[\left(1-p^R\right)\Pr\left(l_1=1|x_1=1\right)+p^R\Pr\left(l_1=1|x_1=0\right)\right]}{d\pi_B^L(1)}\Pi_B^R\left(l_1=1,s_1^R=0\right) \\ &+\left[\left(1-p^R\right)\Pr\left(l_1=1|x_1=1\right)+p^R\Pr\left(l_1=1|x_1=0\right)\right]\frac{d\Pi_B^R\left(l_1=1,s_1^R=0\right)}{d\pi_B^L(1)} \end{aligned}$$

$\frac{d\left[\left(1-p^R\right)\Pr(l_1=1|x_1=1)+p^R\Pr(l_1=1|x_1=0)\right]}{d\pi_B^L(1)} = -\left(1-p^R\right)p^L\left(1-\lambda_1^L\right)-\left(1-p^L\right)\lambda_1^L$ which is negative.

Furthermore:

$$\frac{d\Pi_B^R\left(l_1=1,s_1^R=0\right)}{d\pi_B^L(1)} = \frac{d\left[\mu_B^R(a_1(1,1)) - \mu_B^R(a_1(1,0))\right]}{d\pi_B^L(1)} > 0$$

as:

$$\frac{da_1(1,1)}{d\pi_B^L(1)} = -\frac{1}{\left(\Pr(1,1|1)+\Pr(1,1|0)\right)^2}AB > 0$$

$$\begin{aligned} A &= \left(1-\lambda_1^L\right)\left[\lambda_1^R\phi_G^R(1|1)+\left(1-\lambda_1^R\right)\phi_B^R(1|1)\right] \\ &\quad \left[\lambda_1^R\phi_G^R(1|0)+\left(1-\lambda_1^R\right)\phi_B^R(1|0)\right] \end{aligned}$$

$$B = 1 - 2p^L < 0.$$

However,

$$\frac{da_1(1,0)}{d\pi_B^L(1)} = -\frac{1}{(\Pr(1,0|1)+\Pr(1,0|0))^2}AB < 0$$

as

$$\begin{aligned} A &= \left(1-\lambda_1^L\right)\left[\lambda_1^R\phi_G^R(0|1)+\left(1-\lambda_1^R\right)\phi_B^R(0|1)\right] \\ &\left[\lambda_1^R\phi_G^R(0|0)+\left(1-\lambda_1^R\right)\phi_B^R(0|0)\right] \end{aligned}$$

$$B = 2p^L - 1 > 0$$

The result

$$\frac{d\Pi_B^R\left(l_1=1, s_1^R=0\right)}{d\pi_B^L(1)} = \frac{d\left[\mu_B^R(a_1(1,1)) - \mu_B^R(a_1(1,0))\right]}{d\pi_B^L(1)} > 0$$

conveys that as the counterpart L is more likely to tell a lie when receiving signal 1 (in case L is biased , seeing 1 from L is more likely to be the truth which reinforce R's report of 1 irrespective whether this is the truth or not. The is the extreme case that a bad L lies always so a 1 report means that L is good for sure.

Putting everything together however we get that the change of the second term of BL_{RB} (i.e. $l_1 = 1$) with respect to $\pi_B^L(1)$ is positive as. The negative effect coming from $\left(1-p^R\right)\Pr(l_1=1|x_1=1)+p^R\Pr(l_1=1|x_1=0)$ when L lies more is not enough to cancel the reinforcement effect of a report of 1 coming from L.

We can conclude thus that the change of BL_{RB} with respect to $\pi_B^L(1)$ is positive.

The change of the bad R reputational cost of lying (denoted as CL_{RB}) with respect to $\pi_B^L(1)$ can be shown to be negative. This is due to two factors:

1. $\frac{dv_R^B(l_1=0,r_1)}{d\pi_B^L(1)} = \underbrace{\frac{dv_R^B(l_1=0,r_1)}{d\lambda_2^R(l_1=0,r_1)}}_{>0}\underbrace{\frac{d\lambda_2^R(l_1=0,r_1)}{d\pi_B^L(1)}}_{<0}+\underbrace{\frac{dv_R^B(l_1=0,r_1)}{d\lambda_2^L(l_1=0,r_1)}}_{>0}\underbrace{\frac{d\lambda_2^L(l_1=0,r_1)}{d\pi_B^L(1)}}_{<0} < 0.$

In determining the sign of $\frac{dv_R^B(l_1=0,r_1)}{d\pi_B^L(1)}$ I used the fact that:
$\frac{dv_R^B(l_1=0,r_1)}{d\pi_B^L(1)} = \frac{dv_R^B(l_1=0,r_1)}{d\lambda_2^L(l_1=0,r_1=0)}\frac{d\lambda_2^L(l_1=0,r_1)}{d\pi_B^L(1)}$ and
$\frac{d\lambda_2^R(l_1=0,r_1)}{d\Pr(x=0|l_1=0)} < 0$ and $\frac{d\Pr(x_1=0|l_1=0)}{d\pi_B^L(1)} = \frac{\left(1-\lambda_1^L\right)\left(2p^L-1\right)}{\left(1+\left(1-\lambda_1^L\right)\pi_B^L(1)\right)^2} > 0.$

2. an expert has always reputational incentives to declare against the her perceived biased. By applying a monotonic transformation on $v_R^B(l_1 = 0, r_1 = 0)$ and using point 1. above we get $\frac{d\Pi R_B^R(l_1, s_1^R=0)}{d\pi_B^L(1)} < 0$.

Putting everything together we get $\frac{dCL_{RB}}{\pi_B^L(1)}$ to be negative and as a result we can conclude that $\frac{dNBL_{RB}}{\pi_B^L(1)} > 0$ for any $\pi_B^R(0)$.

The fact that a bad expert' *net* current benefit from lying is a strictly increasing function in the probability of the counterpart lying for all possible own strategies of the counterpart means that in the first period the decision maker is more likely to ignore an advice from the counterpart as it is tainted by the possibility of bias; this in turn implies a higher weight on expert's report and therefore a higher current payoff. At the same time the future reputation costs are the highest when the counterpart probability of lying is close to 0 as the decision maker is able to compare the report with a likely state.

Given the monotonic properties of the net current benefit of lying with respect to both own strategy and counterpart's strategy we can now find the way the optimal strategies of the biased experts change with the parameters of the model.

5.1 Comparative statics with respect to the time preference parameters

The change of a biased experts optimal strategy with respect to time preference parameters, is determined by

$$\frac{d\pi_B^R(0)}{d\mu_B^R} = -\frac{\det\begin{vmatrix} \frac{dNBL_{RB}}{d\mu_B^R} & \frac{dNBL_{RB}}{d\pi_B^L(1)} \\ \frac{dNBL_{LB}}{d\mu_B^R} & \frac{dNBL_{LB}}{d\pi_B^L(1)} \end{vmatrix}}{\det\begin{vmatrix} \frac{dNBL_{RB}}{d\pi_B^R(0)} & \frac{dNBL_{RB}}{d\pi_B^L(1)} \\ \frac{dNBL_{LB}}{d\pi_B^R(0)} & \frac{dNBL_{LB}}{d\pi_B^L(1)} \end{vmatrix}}$$

$$\frac{d\pi_B^R(0)}{d\mu_B^R} = -\frac{\frac{dNBL_{RB}}{d\mu_B^R}\frac{dNBL_{LB}}{d\pi_B^L(1)}}{\frac{dNBL_{RB}}{d\pi_B^R(0)}\frac{dNBL_{LB}}{d\pi_B^L(1)} - \frac{dNBL_{LB}}{d\pi_B^R(0)}\frac{dNBL_{RB}}{d\pi_B^L(1)}} > 0$$

This is due to the fact that $\frac{dNBL_{RB}}{d\mu_B^R} > 0$ as μ_B^R is just a positive transformation on the expected future payment and $\frac{dNBL_{LB}}{d\pi_B^L(1)} < 0$ (as proved earlier).

Also

$$\frac{d\pi_B^R(0)}{d\mu_B^L} = -\frac{\det\begin{vmatrix} \frac{dNBL_{RB}}{d\pi_B^R(0)} & \frac{dNBL_{RB}}{d\mu_B^L} \\ \frac{dNBL_{LB}}{d\pi_B^R(0)} & \frac{dNBL_{LB}}{d\mu_B^L} \end{vmatrix}}{\det\begin{vmatrix} \frac{dNBL_{RB}}{d\pi_B^R(0)} & \frac{dNBL_{RB}}{d\pi_B^L(1)} \\ \frac{dNBL_{LB}}{d\pi_B^R(0)} & \frac{dNBL_{LB}}{d\pi_B^L(1)} \end{vmatrix}}$$

Therefore

$$\frac{d\pi_B^R(0)}{d\mu_B^L} = -\frac{\frac{dNBL_{RB}}{d\pi_B^R(0)}\frac{dNBL_{LB}}{d\mu_B^L}}{\frac{dNBL_{RB}}{d\pi_B^R(0)}\frac{dNBL_{LB}}{d\pi_B^L(1)} - \frac{dNBL_{LB}}{d\pi_B^R(0)}\frac{dNBL_{RB}}{d\pi_B^L(1)}} > 0$$

As $\frac{dNBL_{LB}}{d\mu_B^L} > 0$ and $\frac{dNBL_{RB}}{d\pi_B^R(0)} < 0$ by the same argument as above

Chapter 3

Corporate Governance: A Double Layered Moral Hazard

1 Introduction

In this paper, I propose a framework where an investor is aware of the existence of different participants in a firm and the possibility of exploiting their existence for increasing the chances of a project's success.

The starting point of this analysis is Tirole (2006) which classifies corporate governance literature into four strands based on how the management may not act in the owner's best interests: "insufficient effort", "extravagant investment" which refers to the problem of empire building, "entrenchment strategies" or actions that hurt the owners but secure top executives in their position and last but not least, "self dealing" when "managers may increase their private benefit from running the firm by engaging in a wide variety of self dealing activities." These are all essentially moral hazard problems. In this paper the focus is on self-dealing, which is traditionally solved by compensating the manager sufficiently for not privately enjoying the funds of the company. I show that this may not be the optimal incentive scheme in some cases. Instead in certain conditions it is optimal to incentivize the manager to use discretionary funds within the firm to motivate the employees to participate in the success of the firm.

The novelty of this model lies in the new feature of self-dealing - usually in contract theory self-dealing is treated exclusively as a problem (see Tirole 2006), while in this model I show that allowing the manager to use some uncontractible amount at his discretion might help the firm by encouraging greater effort.

The manager has access to an uncontractible amount either because the firm operates in different environments with different shareholder protection laws, or the size of the firm is large with many but small shareholders, or the nature of the activity does not allow the investor to contract every possible action e.g. research labs. Thus this amount can be interpreted as being characteristic to the nature of the project or environment or both.

The manager may legally disburse funds under his discretionary control for several different reasons. He may, for instance, pay higher than the typical market compensation and perquisites to himself. Alternatively, he may pay higher wages to his workers, spend on better working conditions for firm's employees, also also on infrastructure , e.g. research facilities, sponsorship of social activities for employees (canteens, gyms, nurseries).

In return for sharing his private benefits, the manager creates a work environment which fosters cooperation, and overall higher efficiency from all the participants in the firm. This type of behavior is supported by Fehr, Gachter, and Kirchsteiger (1997) which empirically show that there is a strong reciprocity between firms and workers, and Fehr, Kirchsteiger, and Riedl (1998) which evidence reciprocal behavior even between anonymously trading partners. Furthermore, Bandiera, Barankay and Rasul (2007, 2009, 2011) demonstrate that working with friends can influence positively or negatively (based on the ability of the friends) an individual's productivity within a firm (in this case a fruit farm in UK) while Giuliano (2005) shows that demographic differences between managers and workers can influence the rates of dismissal and promotion of subordinates. Moreover Mas and Moretti (2009) empirically show that individuals are motivated by social relations and mutual monitoring, suggesting that working in a supportive environment can induce effort, when economic mechanisms are limited. So the extra effort made by the employees should not be interpreted as normal to the regular working contract.

Another motivation for this paper is the fact that managers may fail to follow shareholders' objectives not only in order to pursue their own enrichment, but also to pursue the interests of the other employees of the firm. Bertrand and Mullainathan (2003) show that when the corporate governance mechanisms like take-over threats have a limited effect, the managers act both in their own interest and in the interest of the workers: once a law that limits the threat of take-overs is introduced average blue collar wage increases, total factor productivity declines and the return on capital falls.

The corporate governance literature has analyzed self-dealing through several examples: excessive compensation, managerial perquisites, transfer pricing or self-serving financial transactions such as personal loans to insiders, or even theft of corporate assets. Berle and Means (1932) and Jensen and Meckling (1976) look at managerial consumption of perquisites due to lack of separation of ownership and control, while Baumol (1959) and Jensen (1986) analyze over-investment by management. The ability of the management to divert corporate wealth is discussed in Grossman and Hart (1988), Hart (1995) and Zingales (1994). Many corporate finance studies however, look at self dealing as a consequence of concentrated ownership with negative effects on control. Morck, Wolfenzon, and Yeung (2005) have provided a survey on this topic. This paper however is not motivated by the existence of different forms of ownership and control, but on the ability of the manager to divert company's funds.

As the ultimate goal of the owner is to make the manager use all the available tools for achieving higher profitability, however ignoring the relations between manager and the firms employees may result in wrong incentives for managers. Thus, this paper intends to provide a unifying framework in which managerial incentives are impacted by both manager - shareholder and manager-employees relationships.

My model consists of two levels of moral hazard. The investor cannot contract the amount of discretionary funds the manager will share with the workers. Secondly, the manager cannot observe the actual exerted effort and must condition payment on the basis of an observed outcome. Here, we can see that the manager acts both as an agent and as a principal simultaneously in two different subgames; it is necessary to model this extra layer to recognize the existence of an environment consisting of agents whose actions will have an impact on the success of the firm.

The investor chooses not to run the firm himself and thus not to contract the employees himself because he does not have the ability to do it or because the ownership could be very disperse. Moreover even if the ownership is not disperse, as Tirole again points out "managers have proprietary information that often enables them to get their way. So while shareholders have formal control over a number of decisions, managers often have real control." Even financial and regulatory rules - in US in particular - deter investors (in general institutional ones) to sit in boards due to the possibility of being penalized for inside information when re-selling the shares of the company. Shleifer and Vishny (1997) provide a survey of corporate governance practices.

In terms of determining managerial compensation, I consider this in the standard sense of the well-known principal agent problems that arise from asymmetric information. This paper may be seen as an extension of the standard principal agent model to two layers of moral hazard with the manager being both an agent for the investor and a principal for the employees. This is different from double moral hazard problem where two economic actors are engaged in a joint production.

The double moral hazard problem was identified, defined and further analysed by Lafontaine (1992), Romano (1994), Bhattacharya and Lafontaine (1995), Maruyama, (2003) and represents a situation where the manager is both shareholder and agent and thus reacting to two-sided incentives. This issue was captured in different settings such as franchising relationships by Mathewson and Winter (1985), Lal (1990), Roberts (1996), Lal, Park, and Kim (2000), or more generally in vertical integration. Articles that have surveyed the theories of vertical integration are Holmstrom and Roberts (1998), Whinston (2003), Gibbons (2005), and Lafontaine and Slade (2007).

In this model, the problem I address is not related however to the fact that the manager is both shareholder and agent which would be a direct application of vertical integration to corporate governance. The problem here is related to correct identification of all participants in the firm and designing the right incentives for the manager: a double layered moral hazard.

The next section sets up and solves the model, while Section 3 provides testable predictions and discussions. Section 4 concludes.

2 The Model

2.1 The environment

An investor invests I in a project. The project has a verifiable rate of return r at the end of the period if the project is successful and the investor loses all the investment if the project fails. The investor hires a manager to run the business, but he cannot write a complete management contract due to the complex nature of the project. As Tirole (2006) puts it there are in general four ways in which the management may not act in the owner's best interest: insufficient effort, extravagant investment, entrenchment strategies and self dealing. These are all fundamentally issues of moral hazard problem and there are situations in which firms could face at least two or three of these issues, however for this model tractability reasons will consider only the last one: self dealing.

Self dealing problem in this model takes the form of an amount $B < I$ available to the manager but not contractible for a specific task which the manager may choose to spend on his personal welfare or any other discretionary project. The manager has access to this uncontractible amount B either because the firm operates in different environments with different shareholder protection laws, or the size of the firm is large with many but small shareholders, or the structure of the firm gives high powers to CEOs. It could also be the case that the nature of the activity does not allow the investor to contract every possible action, for example research labs. Thus B can be interpreted as being characteristic to the nature of the project/firm or environment or both. The manager may legally disburse funds under his discretionary control for several different reasons. He may, for instance, pay higher than the typical market compensation and perquisites to himself. Alternatively, he may spend on better work conditions and higher wages for firm's employees (Lenovo's CEO Yang Yuanqing shared $3.25 million of his bonus in 2013 with his employees), but also on infrastructure (both productive and social), e.g. research facilities, sponsorship of social activities both for employees (canteens, gyms, nurseries). Providing exceptional work environment triggers a reciprocal behavior from the employees. This extra effort is not ex-ante observable and hence not contractible.

This first level of moral hazard between the investor and the manager is very similar to that employed in Tirole (2001) and Holmstron and Tirole (1997). In these studies, however the manager was compensated for not using B as private benefit, while in this model the manager will have incentive to use the funds for optimally involving the firm's employees in the success of the project.

The manager, thus has the ability to commit to share a portion of this amount B, labelled B^E, with other the employees. In turn, this payment will compensate for exerting higher effort N at cost $c(N)$. The project is successful with probability $p(N)$ and fails with probability $1-p(N)$. The remaining part of B not shared with the employees will be used by the manager for his own private enjoyment. The right share of managerial private benefit B_p and employees' benefit B^E will be determined optimally. The manager will face a trade-off between higher probability of success (implying a higher expected wage) combined with a lower private benefit (due to higher B^E) and the combination of lower expected wage with higher private benefit.

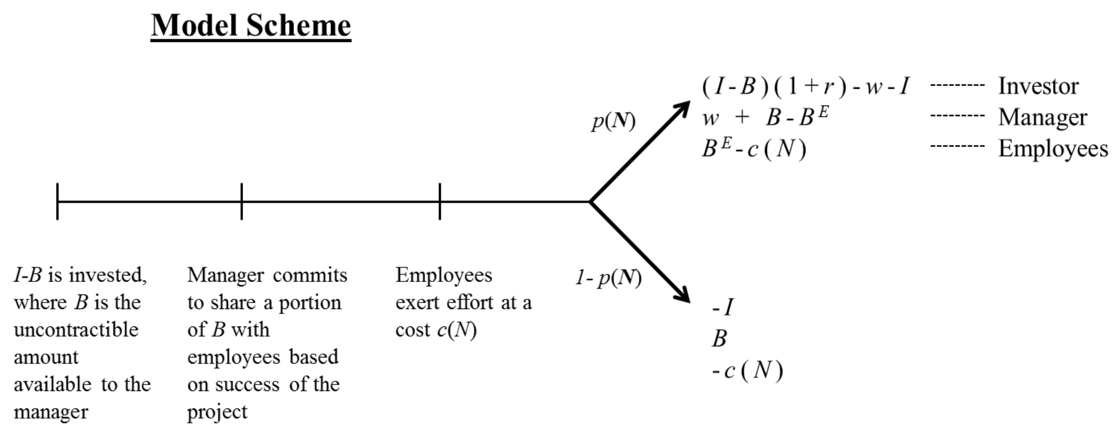
The employees' effort is unobservable, hence uncontractible. If the project is successful the manager will be paid a wage w and the manager will pay the employees B^E while the remaining part of B, B_p will remain with the manager. If the project fails the manager gets zero wage, does not pay anything and keeps the whole B to himself. The minimum wage set to zero corresponds to allowing a limited liability constraint on the manager.

This type of incentive set-up has been lately employed by different corporations in order to attract the employees for the firm's success. One of the latest example is the case of the insurance company Prudential Financial which announced in

February 2014 that 44,000 of its employees who do not normally participate in equity compensation programs will receive a \$1,300 bonus for helping the company to attain its profitability goal ("Prudential Workers Get \$57 Million as CEO Beats Target," Bloomberg News Feb. 6, 2014). Regarding this action John Nadel, an analyst at Sterne Agee and Leach commented that: "rewarding the general, non-executive-management, is a smart move [..] and has to go a long way towards further solidifying morale and loyalty."

My model consists of two levels of moral hazard. First, the investor cannot contract the amount of discretionary funds the manager will share with the workers. Second, the manager cannot observe the actual number of effort exerted by the employees and must condition payment on the basis of an observed outcome. Here, we can see that the manager acts both as an agent and as a principal simultaneously in two different subgames; it is necessary to model this extra layer to recognize the existence an environment consisting of agents whose actions will have an impact on the success of the firm.

The set-up is a three players model with an investor, a manager and the employees of the firm and I restrain myself from issues of moral hazard in teams.



The total effort of the employees determines a particular probability of success $p(N)$. This probability function $p(N)$ could be scaled up by a constant p in order to capture the success of the project when no effort is exerted, $B^E = 0$ and $B_p = B$, but without loss of generality I set this constant to zero. The effort is exerted at a cost $c(N)$.

The investor, in this model, in turn, will maximize his expected return by setting a wage for the manager that increases his incentive to offer more to the employees. The game has a Stackelberg timing.

2.2 Optimal Contracts

2.2.1 First Best

First, I will look at the first best case where both the amount available to the manager B and the effort level are contractible. The optimal effort is straightforward to calculate for a probability function $p{:}R \rightarrow [0,1]$ twice differentiable, concave and increasing in N and a twice differentiable (increasing and convex) cost function $c(N)$:

$$N^* = \arg\max_N p(N)\left[I(1+r)\right] - I - c(N) \tag{1}$$

with first order conditions:

$$\frac{\partial p(N)\left[I(1+r)\right]}{\partial N} = \frac{\partial c(N)}{\partial N} \tag{2}$$

In this case the manager is left with nothing at his discretion and the investors contracts the employees. However the compensation offered to the employees has no allocative rule. The employees receive a constant minimum pay in all states of the world, which makes them exert effort. This case however is hypothetical only as the investor has no executive role in the firm. Next, I look at the allocative role of the contracts (when both B and N are not contractible) offered to the manager by the investor and to the employees by the manager so that both the manager and the investor achieve the highest possible payoffs. In this process, the employees achieve also higher than otherwise possible expected payoffs.

2.2.2 Second Best (B and N non-contractible)

The investor's optimal contracting problem under moral hazard:

$$\max_{w} p(N)\left[(I-B)(1+r)-w\right]-I \tag{3}$$

subject to:

1. individual rationality constraint for the manager:

$$(1-p(N))B+p(N)\left(w+B-B^{E}\right)\geq\bar{u} \tag{4}$$

or

$$B+p(N)\left(w-B^{E}\right)\geq\bar{u}$$

2. incentive compatibility constraint which requires that the manager is optimizing his payoff. This would be the second stage moral hazard problem implied by the manager - employee relationship.

The manager maximizes his expected payoff by offering to the employees a particular payment B^E accounted towards their effort. Thus, the manager's - who is the principal in this case - optimal contracting problem under moral hazard is:

$$\max_{B^W\leq B}(1-p(N))B+p(N)\left(w+B-B^{E}\right) \text{ or} \tag{5}$$

$$\max_{B^W\leq B} B+p(N)\left(w-B^{E}\right) \tag{6}$$

subject to:

1. limited liability constraint for the employees:

$$B^{E}\geq 0$$

2. participation constraints for the employees:

$$p(N)B^{E}-c(N)\geq 0$$

3. incentive compatibility constraint for the employees which states that the effort exerted maximize their private payoff net of their cost of effort:

$$\max_{N} p(N)\, B^E - c(N)$$

The time structure defined above imposes that a effort gets exerted only if there is a positive benefit to be made i.e. $B^E > 0$.

The optimal effort, the managerial wage and the optimal share of the employees' benefit are determined by solving the above problem by backward induction.

For an increasing and concave probability function $p(N)$ and increasing and convex cost function $c(N)$, the optimal number of links $\bar{N}\left(B^E\right)$ is determined by the first order condition:

$$\frac{\partial p(N)}{\partial N} B^E = \frac{\partial c(N)}{\partial N} \tag{7}$$

The optimal compensation scheme can be determined by substituting the optimal effort obtained above into the optimization problem of the manager. Probability of success of the project at the optimal effort becomes a function of employees' compensation and will be denoted by $P\left(B^E\right)$.

$$\bar{B}^E(w) = \arg \max_{B^E \leq B} B + P\left(B^E\right)\left(w - B^E\right) \tag{8}$$

The optimal employees' compensation is determined by the first order condition:

$$\frac{\partial P\left(B^E\right)}{\partial B^E}\left(w - B^E\right) - P\left(B^E\right) = 0 \tag{9}$$

The above condition reflects the relation between the amount that the manager decides to share with the employees and his own salary earned when the project is successful. Basically, if the manager distribute more out the discretionary amount at his disposal he will need to be compensated more if the project is successful. Through this incentive compatibility constraint, the manager effectively says: I will forgo a part of sure B in the favour of my employees if the investor will compensate me more when the project is realized.

Assuming an interior solution, the optimal managerial wage $\bar{w}$ will be determined by substituting the optimal employees' benefit in investor's optimization problem:

$$\max_{w} p\left(\bar{N}\left(\bar{B}^{E}(w)\right)\right)\left[(I-B)(1+r)-w\right]-I \tag{10}$$

such that individual rational constraint is satisfied.

The optimal wage $\bar{w}$ has to furthermore satisfy $B^{E}(\bar{w}) \leq B$, otherwise the maximum available fund available to the manager B is reached. When $B^{E}(\bar{w}) > B$, the bounded employees' payment has to be limited such that the it is not higher than what it is available to the manager i.e. $B^{E} = B$.

It is important to note that even though B^{E} is restricted to the maximum amount that the manager could spend, the relation implied by [9] still needs to be satisfied - wage and employees' benefit have to be in a positive relation, otherwise the manager will not distribute the amount B for the employees to exert effort.

Summarizing the above results:

Proposition 9 *The optimal contract between an investor, manager and social group exists with the following features: the optimal effort of the employee is characterized by* [7], *the employees' benefit is characterized by* [9] *,while the optimal wage contract offered to the manager is determined by [10]; the final return on investment is* $p\left(\bar{N}\right)\left[(I-B)(1+r)-\bar{w}\right]-I.$

In section 3 I provide an example which capture heterogeneous environments with different optimal corporate governance set-ups.

2.2.3 Optimal Amount at Manager's Disposal

If we address a more general question - in the sense that we allow the investor to choose an optimal B which maximizes his profit - then the investor will maximize his profit over both w and over B :

$$\max_{w,B} p\left(N\left(B^E(w)\right)\right)\left[(I-B)(1+r)-w\right]-I$$

such that manager's individual rationality constraint is satisfied:

$$\left(1-p\left(N\left(B^E(w)\right)\right)\right)B+p\left(N\left(B^E(w)\right)\right)\left(w+B-\bar{B}^E\right)\geq \bar{u}$$

.

All the other constraints of the manager and the employee remain the same.

3 Predictions and discussions

Further I look at a tractable case when probability of success has the form: $p(N)=N$ and I assume a quadratic and separable cost of effort for exerting effort:

$$c(N)=m\frac{(N)^2}{2}.$$

The parameter m captures the ease of exerting effort by the employee. This could be specific to the employee in terms of training, abilities or specific to the firm, or the industry. At a more general level the cost of effort could be characteristics to different geographic community and could be due to different education systems, literacy, cultural paths, social norms, ethnicity or even past patterns of migration

The next results describe different optimal contracts $\left(\bar{w},\bar{B}_p\right)$ - between an investor and his manager - and $\left(\bar{B}^E,\bar{N}\right)$ - between the manager and the employee - at different levels of uncontractible amount available to the manager and various level of cost of effort.

Result 1 *If* $0<B\leq\frac{I(1+r)}{4m^2-r-1}$, *the optimal contracts* $\left(\bar{w},\bar{B}_p\right)$ *and* $\left(\bar{B}^E,\bar{N}\right)$ *exist and are characterized by the following:*

a. *the managerial compensation is set at* $\bar{w}=\frac{B}{2m}$ *if* $\bar{u}<\frac{B}{2m}$ *and manager's individual rationality constraint does not bind.*

b. *the managerial compensation is set at* $\bar{w} = \sqrt{m^2 (\bar{u} + B)}$ *if* $\bar{u} \geq \frac{B}{2m}$ *and his individual rationality constraint binds.*

c. *the private managerial benefit* $\bar{B}_p$ *is set at zero.*

d. *the employees' benefit is* $\bar{B}^E = B$.

e. *the optimal level of effort is* $\bar{N} = \frac{B}{m}$.

f. *the investor's payoff:* $\Pi = \frac{B}{m} \left[(I - B)(1 + r) - \frac{B}{2m} \right] - I$.

We are now in a situation when the amount available to the manager is not very high. We see that within this incentive structure, all managerial private benefit is sacrificed for increasing the success of the project which further translates to higher ex-post managerial benefits. This is due to the fact the employees have the ability of exerting higher effort for increasing the firm's success.

In terms of comparative statics, if the uncontractible amount - B - available to the manager is small, by increasing it, more effort is exerted; as a result employees' benefit is higher. The managerial private benefit is set at zero but the managerial optimal compensation increases with B. This incentive scheme basically sets-up an uncertain higher bonus for forgoing a certain private benefit.

Result 2 *If* $I \geq B > \frac{I(1+r)}{4m^2 - r - 1}$, *and* $\frac{(I-B)^2(1+r)^2}{8m^3} - B \geq \bar{u}$, *the optimal contracts* $\left(\bar{w}, \bar{B}_p\right)$ *and* $\left(\bar{B}^E, \bar{N}\right)$ *exist and are characterized by the following:*

a. *the managerial wage is set up at* $\bar{w} = \frac{[(I-B)(1+r)]}{2m}$.

b. *the private benefit is* $\bar{B}^p = B - \frac{[(I-B)(1+r)]}{4m^2}$.

c. *the optimal employees' benefit is* $\bar{B}^E = \frac{[(I-B)(1+r)]}{4m^2}$.

d. *the optimal level of effort is* $\bar{N} = \frac{[(I-B)(1+r)]}{4m^3}$.

e. *the investor's payoff is* $\frac{(I-B)(1+r)(2m-1)}{8m^3} - I$

If the amount available to the manager is above a threshold, he will start distributing a lower share to the employees. In this case both the wage and the employees' benefit decrease with B while the private benefit increases with B.

This result suggests that, when B exceeds a minimum threshold, increasing B leads to a disparity in earnings of the manager vis à vis the benefits shared with employees- i.e. when more of the investment takes an uncontractible form, the greater the inequality will be. The wage is set at lower levels but the manager is compensated with higher personal benefit. In other words, beyond a certain level, the existence of B can no longer be exploited to incentivize the manager. This high uncontractibility of B could be due to the nature of the ownership of the firm (small and dispersed shareholders are less likely to get involved in the activity of the manager), or even due to the nature of firm's activity for example R&D centers where it is not possible to contract due to the uncertainty of the activity. It may also be an outcome of the legal environment, where regulations for accounting disclosure are either less stringent or are poorly enforced or even the structure of the firm where CEOs hold great executive powers.

The cost of effort plays also an important role in this set up as lower cost translates in higher compensation for the employees and lower private benefit for the manager while the investor is overall better-off.

Next I look at a situation when manager's individual rationality constraint binds and the manager needs to be compensated more to stay with the firm.

Result 3 *If* $I \geq B \geq \frac{I(1+r)}{4m^2-r-1}$, *and* $\frac{(I-B)^2(1+r)^2}{8m^3} - B < \bar{u}$ *the optimal contracts* $(\bar{w}, \bar{B}_p)$ *and* $(\bar{B}^E, \bar{N})$ exist and are characterized by the following:

a. *the managerial wage is set at* $\bar{w} = \sqrt{4m^2(\bar{u} - B)}$

b. *the private benefit is* $\bar{B}_p = B - \sqrt{(\bar{u} - B)}$

c. *the total employees' benefit is* $\bar{B}^E = \sqrt{(\bar{u} - B)}$

d. *the optimal level of effort* is $\bar{N} = \frac{\sqrt{(\bar{u}-B)}}{m}$ *and*

e. *the investor's payoff is* $\Pi = \sqrt{(\bar{u} - B)}\left[(I - B)(1 + r) - \sqrt{4m^2(\bar{u} - B)}\right] - I$

As before, after a threshold increasing B does not incetivize the exertion of effort and the manager keeps most of it in terms of private benefit. Increasing B translates here as well, in higher amounts transferred to the manager and lower employees' compensation.

3.1 A possible interpretation of the model: the emergence of different structures of corporate governance

It appears from this simplified example that the level of uncontractible amount available to the manager and employees' cost of exerting effort are important factors in determining the optimal nature of contract for the manager.

If the uncontractible portion of investment B is in a moderate region, or the cost of effort is low, this may lead to shared benefits among employees and the manager. Further we see lower level of inequality between the manager and the employees and higher profits for the investors.

If either the uncontractible amount or the cost of exerting effort is high we see a higher disparity of earning between the manager and the employees. In this case most of the benefits are enjoyed by the manager.

Based on this simple model we could explain the emergence of different forms of corporate governance across the world as different countries have different legal systems which regulate the uncontractible usage of funds but also have different organizational structures based on the employees participation in firms. For example the US and UK systems of corporate governance focus on the owner's best interests while the corporate governance in Japan, Germany and France relates to employees' interests (but also of other stakeholders) rather than profitability exclusively for the owners. In Germany this type of governance is stipulated by the law with firms' employees being represented in the boards of directors. In Japan, however, the strict social norms rather than the law impose stakeholders' interests to be of foremost importance to the firm's objectives. Based on my model one could argue that these systems of governance are optimal from investors' perspectives given the specific cost of exerting effort by the employees. In particular this cost could be characteristics to each region and could be due to different education systems, literacy, cultural paths, social norms, ethnicity or even past patterns of migration .

As an illustration of the results in Section 3 we could analyze governance and organizational structures in countries like Japan where firms are embedded in social environments versus governance systems in more individualistic societies. When describing the structure of Japanese organization Aoki (1990) suggests that firm's governance has a high impact on the efficient allocation of resources. Furthermore, Aoki identifies "the ethnic homogeneity of the Japanese domestic factory" which in my model could be interpreted as a low cost of exerting effort as a crucial factor for the development and effectiveness Japanese corporations.

However, although the social environment could foster lower cost of effort, if the amount uncontractible available to the manager is too high, perhaps due to large size of the firm or weak legal enforcement system, this model predicts high disparity in earning between the management and the employees and stakeholders. Again we see consistency with the Japanese governance systems, but in this case we look at large corporations. In particular, when CEOs are given high powers without proper means of control we observe that they are less accountable to both shareholders and stakeholders even though the social environment fosters efficient allocation of resources.

The latest Japanese financial scandals (i.e. Olympus scandal[8] which concluded in 2013 with three top executives being found guilty and imprisoned for falsifying accounts to cover up losses of $1.7bn, or the latest dismissal of the top executive of Mizuho Financial Group[9], Japan's second-largest lender, for offering more than $2 million in loans to people affiliated with organized crime), show that when the size of the company becomes too large and there are not proper means of

[8]BBC News 3rd July 2013 "Olympus scandal: Former executives sentenced."

[9]Reuters news, 23rd January 2014: "Mizuho replaces core unit CEO after mob loan scandal."

monitoring the boards, this type of governance could lead instead to losses. As pointed out by Yashimori (1995) “large listed corporations in Japan are legally subject to two monitoring mechanisms: statutory auditors and independent certified public accountants. Neither is functioning properly. [...] The root cause of the lack of monitoring by the statutory auditors is that they are selected by the president whom they are supposed to monitor. [...] 90% of the statutory auditors (in large corporations) are indeed chosen by the president for perfunctory approval at the shareholders' meeting.” To tackle this situation in June 2010, the Tokyo Stock Exchange introduced a new rule that all listed companies must appoint at least one independent outside director or a statutory auditor who is independent.

A different situation we might consider is when the costs of effort is high or the uncontractible amount is high then the manager will keep all the private benefit for himself. In this case, the investor might wish to design different incentive schemes such that the manager uses the entire discretionary amount B to complement the existing investment instead of sharing it with the employees or keeping for his own use respectively. This would be just a simple model of moral hazard in which the manager's wage is set at high enough levels such that he has no incentives to deter B from investment as in Tirole (2001). This form of governance is optimal in societies characterized by a more individualistic approach of business such as US or UK.

Designing incentive schemes ignoring the social structure in which the company operates (which the existing literature has done so far) might result however in inefficient allocation of resources than otherwise feasible. This is due not only to the inability of using all the available resources but also to the design of wrong incentives offered to the manager in the form of inappropriate compensation which is too high to incentivize him to look for alternative sources of increasing profitability.

One could also argue that the set-up in this model is related with the concept of stakeholder society where all the participants in the success of the firm are taken into account when decisions are made. This concept originated and developed in the management theory (Freeman 1984, Freeman and Evan 1990, Hillman and Keim, 2001, Godfrey, 2005, Walsh, 2005). The importance of stakeholder influence on incentive in firms has also been recently recognized in economic studies through Tirole 2001, Magill, Quinzii and Rochet 2013, and Allen, Carletti and Marquez 2013. Similarly to Yoshimori (1995) the model captures a positive relation between firm's efficiency and the stakeholder approach of governance as ignoring important participants in the firm's success might translate in the wrong incentives for the management.

4 Conclusion

This study attempts to model corporate governance as a situation where an investor is aware of the existence of different participants and the possibility of exploiting their existence for increasing the chances of a project's success. By corporate governance I understand mechanisms through which managers are made to act in the interest of shareholders.

The common approach to model corporate governance is through principal agent models which could identify optimal compensation contracts for managers. In this model I look at a particular example of the agency problem: self-dealing.

In the contract theory literature, self dealing is described as a circumstance where a manager has access to uncontractible resources and needs therefore to be compensated for not expropriating funds from the company. I extend this approach by considering the existence of an extra organizational layer: the employees. The employees help the firm run smoothly by exerting effort. I argue that an agent's incentives at work are moderated by both his ability of performing his tasks but also by the uncontractible amounts available in the firm. I look thus for the optimal managerial compensation such that the manager is incentivized to use private benefits for making the employees to exert effort. Self-dealing thus loses the negative connotation from the current contracts literature and the manager shares the private benefits with the employees when he is offered the right compensation

I also add to the contract theory literature by extending the standard principal agent model to two layers of moral hazard with one manager being both an agent for the investor and a principal for the employees. This is different than the double moral hazard problem where the shareholder is both principal and agent.

In section 3 I connect this model with the emergence of different corporate governance system across the world and I argue that not only the availability of uncontractible funds but also the cost of exerting effort are determinants in choosing different organizational structures. I argue that this cost of effort is due to fundamental differences between societies - such as history, culture, or geography, including ethnicity or past patterns of migration. This paper tries to go to a more profound level and proposes one plausible set of mechanisms that relate social and legal characteristics (cost of exerting effort and availability of uncontractible resources) to corporate governance and possibly other economic outcomes. This characteristics in question need not be exogenous, but would play the role described in the paper as long as they were sufficiently entrenched. In that sense, this is not a dynamic model, but rather an attempt at reaching one level lower into more fundamental sources of difference between societies.

Although these characteristics are endogenous to varying degrees, I suggest that they will help make cross-sectional comparisons across societies, and possibly intertemporal comparisons for a given society. Indicators for such features may take the form of differences in ethnicity levels, but also legal structures and enforcement quality, openness, democracy, and other such proxies for protection of property rights, the rule of law, and factors that enable individuals to perform their jobs better. There is a great body of literature that examines cross-sectional differences in economic outcomes of countries as a result of differences in social and legal characteristics, however, the precise mechanisms that underlie the processes assumed in this literature that often refers to the term "institutions" are not always clear. For instance, if all agents were to act rationally given knowledge of the effect of institutions on economic well-being, there would be a great deal more rapid convergence in economic outcomes across the world - something we do not see. This paper attempts a connection between the observed macroeconomic evidence about "quality of institutions" and further provides testable predictions about differences in corporate governance mechanisms and wage distributions in society. To a more practical end along this line, it also offers a basic policy framework for transnational corporations in terms of setting the appropriate ratio of managerial wage to managerial "slack" based on the location of activity (cultural environment).

This model can apply to general conditions while the outcome in terms of effort exerted and benefits shared are not forgone conclusions, but depends on the interplay of regional characteristics. In other words heterogeneity in legal and cultures differences would play a central role in determining the governance mechanisms and wage contracts offered by investors.

References

[1] Allen, F., E. Carletti, R. Marquez. 2013. Stakeholder governance, competition and firm value. *Wharton Financial Institution Center Working Paper Series.* 7 13-33.

[2] Aoki, M. 1990. Toward an economic model of the Japanese firm. *Journal of Economic Literature* 28 1-27.

[3] Baumol, J. 1959. Business behaviour, value and growth. *Macmillan, London.*

[4] Bandiera, O., I. Barankay, I., I. Rasul. 2009. Social connections and incentives in the workplace: evidence from personnel data. *Econometrica* 77(4) 1047-1094.

[5] Bandiera, O., I. Barankay, I., I. Rasul. 2007. Incentives for managers and inequality among workers: evidence from a firm level experiment. *Quarterly Journal of Economics* 122(2) 729-773.

[6] Bandiera, O., I. Barankay, I., I. Rasul. 2011. Field experiments with firms. *Journal of Economic Perspectives.* 25 63-82.

[7] Berle, A., G. C. Means.1932. The Modern Corporation and Private Property. *Macmillan, NewYork.*

[8] Bhattacharya, S., F. Lafontaine. 1995. Double sided moral hazard and the nature of share contracts. *Rand Journal of Economics* 26 761-81.

[9] Bertrand, M., S. Mullainathan. 2003. Enjoying the quiet life? Corporate governance and managerial preferences. *Journal of Political Economy.* 111 1043-75.

[10] Fehr, E., G. Kirchsteiger, A. Riedl. 1998. Gift exchange and reciprocity in competitive experimental markets. *European Economic Review.* 42(1) 1-34.

[11] Fehr, E., S. Gachter, G. Kirchsteiger. 1997. Reciprocity as a contract enforcement device: experimental evidence. *Econometrica.* 65. 833-860.

[12] Freeman R.E. 1984. Strategic management: A stakeholder approach. Pitman. Boston.

[13] Freeman R.E, W.M. Evan.1990. Corporate governance: a stakeholder interpretation. *Journal of Behavioral Economics.* 19 337–359.

[14] Gibbons, R. 2005. Four formal(izable) theories of the firm? *Journal of Economic Behavior and Organization*. 58 200–245.

[15] Godfrey, P.C. 2005. The relationship between corporate philanthropy and shareholder wealth: a risk management perspective. *Academy of Management Review.* 30 777–798.

[16] Grossman, S., O. Hart. 1980. Takeover bids, the free-rider problem, and the theory of the corporation. *The Bell Journal of Economics.*11(1). 42–64.

[17] Hart, O. 1995. Firms, contracts, and financial structure. *Clarendon Press, Oxford.*

[18] Hillman A.J., G.D. Keim. 2001. Shareholder value, stakeholder management, and social issues: what's the bottom line? *Strategic Management Journal.* 22(2)125–139.

[19] Holmstrom, B., Roberts, J. 1998 The boundaries of the firm revisited. *Journal of Economic Perspectives*. 12: 73–94.

[20] Holmstrom, B., J. Tirole. 1997. Financial intermediation, loanable funds, and the real sector. *The Quarterly Journal of Economics*. 112(3). 663-91.

[21] Jensen, M. 1986. Agency costs of free cash flow, corporate finance, and takeovers. *American Economic Review.* 76, 323–329.

[22] Jensen, M., W. Meckling. 1976. Theory of the firm: managerial behavior, agency costs and ownership structure. *Journal of Financial Economics.* 3. 305–360

[23] La Porta R., F. Lopez-de-Silanes, A. Shleifer. 1999. Corporate ownership around the world. *Journal of Finance.* 2. 471-517.

[24] Lafontaine, F. 1992. Agency theory and franchising: some empirical results. *RAND Journal of Economics.* 23(2) 263-283.

[25] Lafontaine, F., M. E. Slade. 1997. Retail contracting: theory and practice. *Journal of Industrial Economics.* 45(1). 1-25.

[26] Lafontaine, F., M. E. Slade. 2007. Vertical integration and firm boundaries. *Journal of Economic Literature.* 45(3). 629-685.

[27] Lal, R. 1990. Improving channel coordination through franchising. *Marketing Science.* 9(4). 299-318.

[28] Lal. R., C. S. Park, H. Kim. 2000. Margin or sales. A comparative analysis of two royalty structures in Japanese convenience-store industry. *Japanese Distribution Strategy, Business Press..*

[29] Magill, M., M. Quinzii, J. C. Rochet. 2013. A critique of shareholder value maximization.*Swiss Finance Institute Research Paper.*13-16.

[30] Maruyama, M. 2003. Optimal structure of franchise contracts. *Journal of Economics & Business Administration.* 188(1).11-26.

[31] Mas A., E. Moretti. 2009. Peers at work. *American Economic Review.* 99(1)112-45.

[32] Mathewson, G. F., R. A. Winter. 1985. The economics of franchise contracts. *Journal of Law and Economics.* 28(3). 503-526.

[33] Milgrom, P. R. 1988. Employment contracts, influence activities, and efficient organization design. *Journal of Political Economy.* 96(1) 42-60.

[34] Morck, R., D. Wolfenzon, Yeung, B. 2004. Corporate governance, economic entrenchment and growth. *NBER Working Papers 10692*

[35] Phillips R.A., R.E. Freeman RE, A.C. Wicks AC. 2003. What stakeholder theory is not. *Business Ethics Quarterly.* 13 479–502.

[36] Robert, M. 1996. Incomplete contracts and the choice between vertical integration and franchising. *Journal of Economic Behavior and Organization.* 31.101–115.

[37] Romano R. E. 1994. Double moral hazard and resale price maintenance. *RAND Journal of Economics.* 25 455-466.

[38] Rotemberg, J. J. 1994. Human relations in the workplace. *Journal of Political Economy.*102 684-717.

[39] Shleifer, A., and R. W. Vishny. 1997. A survey of corporate governance. *Journal of Finance.* 737–83

[40] Tirole, J. 2001. Corporate governance. *Econometrica.* 69(1)1-35.

[41] Tirole, J. 2006. The theory of corporate finance. *Princeton University Press.*

[42] Zingales, L.1994. The value of the voting Right: A study of the milan stock exchange experience. *Review of Financial Studies.* 7(1). 125-48.

[43] Walsh, J.P. 2005. Taking stock of stakeholder management. *Academy of Management Review.* 30 426–438.

[44] Whinston, M. D. 2003. On the transaction cost determinants of vertical integration. *Journal of Law, Economics, and Organization.*19. 1–23.

[45] Yoshimori, M.,1995. Whose company is it? The concept of the corporation in Japan and the West. *Long Range Planning.* 28 33-44.

5 Appendix

5.1 Proof of Proposition 1

Proof. It follows directly from the properties of the probability function $p{:}R \to [0,1]$ twice differentiable, concave and increasing in N and cost function $c(N)$ as well twice differentiable, increasing and convex in N. ■

5.2 Proof of the results in section 3

We solve for the optimal contract by backward induction. The success function is $p(N) = N$ and cost function $c(N) = m\frac{(N)^2}{2}$ with $m > \frac{1}{2}$.

For $p(N) = N$ the employees' maximization problem is:

$$\max_N NB^E - m\frac{(N)^2}{2}$$

which implies:

$$N = \frac{B^E}{m} \tag{11}$$

Substituting these optimal links into the probability function we get:

$$p(N) = \frac{B^E}{m}$$

The manager's optimization problem becomes:

$$\max_{B^W \leq B} B + \frac{B^E}{m}\left(w - B^E\right)$$

Thus:

$$B^E = \frac{w}{2m}$$

This shows that there is a positive relation between manager's wage and the employees' compensation.

The condition for an interior solution for the employees compensation is $B^E < B$

or $w < 2mB$.

Now, the investor maximization problem becomes:

$$\max_{w} \frac{w}{2m^2}\left[(I-B)(1+r)-w\right]-I$$

such that

$$B+\frac{w}{2m^2}\left(w-B^E\right)\geq \bar{u} \quad (12)$$

FOC:

$$\frac{[(I-B)(1+r)]}{2m^2}=\frac{w}{m}$$

Thus

$$w=\frac{[(I-B)(1+r)]}{2m}$$

and

$$\begin{aligned} B^W &= \frac{[(I-B)(1+r)]}{4m^2} \\ B^p &= B-\frac{[(I-B)(1+r)]}{4m^2} \end{aligned}$$

There are two constraints that need to be verified:

1. $B^E \leq B$, so in order to have indeed an interior solution for B^E then $\frac{[(I-B)(1+r)]}{4m^2} < B$ or

$$B > \frac{I(1+r)}{4m^2-r-1}$$

Note that $4m^2-1-r>0$ as long as $m > \frac{\sqrt{1+r}}{4}$

2. $B+\frac{w}{2m^2}\left(w-B^W\right) \geq \bar{u}$,so in order to have interior solution for $w : w^2 \geq 4m^2\left(\bar{u}+B\right)$ or $\frac{(I-B)^2(1+r)^2}{8m^3}-B\geq \bar{u}$

If both of this conditions are satisfied investor payoff is:

$$\begin{aligned} \Pi &= \frac{[(I-B)(1+r)]}{4m^2}\left[(I-B)(1+r)-\frac{[(I-B)(1+r)]}{2m}\right]-I \\ &= \frac{(I-B)(1+r)(2m-1)}{8m^3}-I \end{aligned}$$

Π is positive as long as $\frac{I\left[(1+r)(2m-1)-8m^3\right]}{8m^3} > B\,(1+r)\,(2m-1)$.

If now, constrains bind than the above results are the optimal contract and hence

Result 2.

If condition 1 fails then $B^E = B$ and $w = \frac{B}{2m}$ and the investor's payoff is

$$\Pi = \frac{B}{m}\left[(I-B)\,(1+r) - \frac{B}{2m}\right] - I$$

These results represent the optimal contract in **Result 1**.

If condition 2 fails but 1 is satisfied then $w = \sqrt{4m^2\,(\bar{u}-B)}$ and $B^E = \sqrt{(\bar{u}-B)}$. and the investor payoff is:

$$\Pi = \sqrt{(\bar{u}-B)}\left[(I-B)\,(1+r) - \sqrt{4m^2\,(\bar{u}-B)}\right] - I$$

These results represent the optimal contract in **Result 3**.