AVERTISSEMENT

LIENS

Code la Propriété Intellectuelle – Articles L. 122-4 et L. 335-1 à L. 335-10
Loi n°92-597 du 1er juillet 1992, publiée au *Journal Officiel* du 2 juillet 1992
http://www.cfcopies.com/V2/leg/leg-droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 1 Capitole (UT1 Capitole)

Présentée et soutenue par :
Tong SU

le 15 Novembre 2016

Titre :

Three Chapters in Information Economics

École doctorale et discipline ou spécialité :

ED TESC : Économie

Unité de recherche :
Théorie des Jeux

Directeur/trice(s) de Thèse :

Monsieur Christian HELLWIG

Jury :
Monsieur Bruno BIAIS, Professeur, Université Toulouse 1 capitole
Monsieur Christian HELLWIG, Professeur, Université Toulouse 1 capitole
Monsieur Filip MATEJKA, Professeur, CERGE-Prague
Monsieur Alessandro BONATTI, Professeur, MIT-Sloan

Public Communication with Coordination Frictions

By Georgy Lukyanov and Tong Su*

Draft: October 7, 2014

This paper develops a model in which the sender strategically communicates with a group of receivers whose payoffs depend on the sender's information. It is shown that, in the presence of coordination frictions, conflict of interests between the sender and the receivers arises endogenously, in spite of the sender's benevolence. As a result, equilibrium communication is imperfect: extremely good or bad news get disclosed, while relatively "neutral" information is withheld. Consequently, an exogenous bias in the sender's preferences can improve communication and raise welfare.

JEL: C72, D83.

Keywords: Endogenous conflict of interests; public information provision; optimal bias; global games.

In various economic circumstances, an informed agent communicates with the group of people via public messages: credit rating agency issues rating reports to potential investors, Central Bank makes policy announcements in press, a firm publishes annual reports revealing its current performance, a manager reviews team performance in routine meetings with employees, etc. In such situations, revealed information brings private benefits to individuals, as it improves their decision making.

* Lukyanov: Toulouse School of Economics, Manufacture des Tabacs, 21 Allée de Brienne, 31000 Toulouse Cedex, France (email:lukyanov@mit.edu); Su: Toulouse School of Economics, Manufacture des Tabacs, 21 Allée de Brienne, 31000 Toulouse Cedex, France (email:tong.su@tse-fr.eu). We thank Christian Hellwig for extensive feedback and invaluable guidance. We are also grateful to Ryan Chahrour, Jacques Crémer, Harry Di Pei, Alexander Guembel, David Jimenez-Gomez, Shuo Liu, Eric Mengus, Stephen Morris, Alessandro Pavan, Guillaume Plantin, François Salanié, Jakub Steiner, Robert Ulbricht and Xavier Vives for helpful comments and discussions. Financial support from the European Research Council under the European Community's 7th Framework Programme FP7/2007-2013 grant agreement N263790 (InfoMacro) is gratefully acknowledged. All remaining errors are ours.

However, it should be kept in mind that the members of the audience sometimes engage in strategic interaction among themselves: a creditor may not be willing to lend to a firm whenever he learns that it had been denied by others; likewise, peer pressure or synergies induce an employee in an organization to exert effort when he knows that others are also doing that. In a broad sense, we refer to these payoff interdependencies as *coordination frictions*.

In the strategic world, public messages serve a dual role. On the one hand, extra information allows one to take a more proper action; on the other hand, due to its publicity, a message acts as a coordination device facilitating one's prediction of others' behavior[1]. As a result, full information disclosure, even though can be a good policy *ex ante*, is usually suboptimal *ex post*.

We demonstrate that the underlying reason is the conflict of interests between uninformed audience and the informed public agent, which emerges *endogenously* and despite the public agent's benevolence. Let us consider an example: after the 2007 financial crisis, CRAs have been criticized for their misleading "inflated ratings" for securitized debts. However, one defending argument was that the rating agencies deliberately avoided downgrades in afraid of feedback effects which may trigger "multi-notch downgrades" and destabilize markets[2]:

> ...the very fact of a rating downgrade has an effect in the financial market... even if [investors] do not rely on ratings, they pay attention to the downgrade because they consider that other market participants may react negatively to the downgrade... [A] confidence crisis in ratings or massive rating downgrades may totally destabilize the financial markets... therefore, leading CRAs are extremely reluctant to downgrade a company's debt.[3]

[1]For instance, investors usually ask a lower risk premium for AAA bonds not just because good ratings indicate higher solvency, but also because high-rated firms face smaller debt run risks. He and Xiong (2012) develop a dynamic debt run model showing that the pattern of yield spreads is largely driven by creditors' coordination concerns, whereby each of them decides whether or not to roll over. This effect makes yield spreads excessively volatile in comparison to movements in fundamentals.

[2]The feedback effect by credit ratings is studied by Manso (2013).

[3]See Darbellay (2013), pp. 183-185.

Whereas the impact of the conflict of interests on information transmission has been well studied, microfoundations for the conflict itself received little attention in the literature. Our paper develops a theoretical model that derives this conflict from the primitives.

We address the following questions: In what way do coordination frictions affect incentives of the public agent to disclose information? How does it change equilibrium communication? What should be personal characteristics of the sender (in terms of his preferences) that might improve social welfare?

Based on the setup of Crawford and Sobel (1982), we model information transmission within the standard sender/receiver framework. A large population of individuals (receivers), each of whom faces a binary action choice ("risky" versus "safe") is guided by the public agent (sender) who possesses some relevant information. Two assumptions are made throughout the paper. First, it is assumed that sender's information is *hard*, so that it can be either revealed truthfully or not at all. Second, we assume that the sender's message must be *public*: in particular, the sender cannot tailor-make her reports for different subgroups of receivers.

Applying the global games' methodology, we pin down the unique symmetric equilibrium in any continuation game following the sender's message. This equilibrium is of the standard "switching" type, whereby each receiver switches his action from one to another as long as his private signal passes a given threshold, which behaves monotonically in the public message.

For the benchmark case where coordination frictions are absent, we show that a benevolent sender is always willing to reveal her information *ex post*. Consequently, coordination frictions are a primary force driving the conflict of interests. Technically, they create a wedge between receivers' strategy that is desirable from the benevolent sender's perspective, given her information, on the one hand, and the equilibrium strategy played by receivers in the continuation game, given the same information, on the other: *ex-post* truthful revelation is no longer incentive compatible for the sender. As a result, equilibrium communication is coarse: the

sender chooses to withhold a range of medium signal realizations, at the same time disclosing extreme realisations.

Our paper draws novel welfare implications that contrast with the predictions from standard communication models, in which the conflict of interests is typically welfare-detrimental. In our model, an exogenous preference bias in an appropriate direction can partially offset the endogenous conflict of interest, facilitating communication and improving welfare[4].

Related Literature

Strategic information transmission has been extensively studied since the emergence of the cheap-talk literature, pioneered by Crawford and Sobel (1982). As they have shown, conflict of interests between an advisor and a decision maker inhibits efficient information transmission and reduces welfare[5]. From this perspective, our paper can be viewed as providing microfoundations for the endogenous conflict of interests.

Farrell and Gibbons (1989) extend their model to multiple audiences and study the sender's optimal usage of public and private messages. More recent examples along these lines are Eliaz and Forges (2012) and Goltsman and Pavlov (2011). In contrast to these papers, wel restrict the sender to use public messages.

Our paper is also closely related to the work by Che and Kartik (2009), who show that a decision maker can be better off by having an advisor whose belief is different from his own. The reason is that disagreement in the priors induces the sender to acquire information in order to persuade the receiver, which turns out to be valuable for both of them. This conclusion is similar to ours, but its

[4]This normative result sheds light on many real-life situations, such as how people should pick the right advisor or how institutions should be designed for effective communication. For instance, qualities of a good leader are often studied in organizational literature. Bolton et al. (2013) have shown that resoluteness is good for the leader who faces a coordination issue among her followers. Our paper complements their findings by showing that the desirable leader's type is not invariant to the nature of coordination. We defer a more detailed discussion of potential applications until Section V

[5]An early example of the cheap-talk model with multiple audiences can be found in Farrell and Gibbons (1989). For the overview of the work in the field, see Farrell (1995), Farrell and Rabin (1996) and the recent paper by Sobel (2013).

driving force is quite different: the disagreement comes not from differences in beliefs, but from the payoff externalities[6].

One feature that distinguishes our communication protocol from the one adopted in the cheap-talk literature is that the sender's information is non-falsifiable. This places our work within the literature on communication with lying costs. An early contribution was made by Seidmann and Winter (1997), who showed that when misreporting is not costless, a relatively weak monotonicity condition of the sender's preferences in the receiver's actions leads to a fully revealing equilibrium. Austen-Smith and Banks (2000) introduced signalling (along with cheap talk) within the standard setup; more recent examples include Kartik (2007, 2009); Kartik et al. (2007). Through the lens of these models, our work can be placed at the opposite side of the spectrum (relative to cheap talk), where any misrepresentation is prohibitively costly.

A relatively recent literature has studied the social value of public information in an environment, where agents' actions exhibit strategic complementarities. Morris and Shin (2002) and Angeletos and Pavan (2004, 2007) remain classical reference. The main intuition of their papers is that more precise public information can reduce social welfare as the coordination motive makes agents "over-weight" the public signal relative to their private signals. Although we adopt a similar setup, our motivation is different. While their analysis focussed on exogenously provided public information, we derive it from the sender's behavior. As a consequence, we provide a distinct mechanism for efficiency loss, which stems from the sender's commitment problem.

Another building block we use is the global games literature, which studies the role of public and private information in coordination games. Several papers have demonstrate that increasing the transparency of public information can reduce social welfare. In particular, Morris and Shin (2002) show that individuals tend

[6]The idea that a DM can be benefited from the heterogeneity in preferences between him and the advisor has also been pointed out, in various frameworks, by Dewatripont and Tirole (1999), Prendergast (2007) and Landier et al. (2009).

to put too much weight (relative to the Pareto optimum) on the public signal, since it serves a coordination purpose. However, they assume the public agent can commit to the precision of her public message, whereas we do not allow commitment. Yet we obtain a similar result that social welfare is not necessarily increasing in the informativeness of equilibrium reporting strategy.

Some papers in this field have tried to endogenize public information. Angeletos et al. (2006) examine the informational role of policy in the global games' setting: along the lines of Morris and Shin (1998), they analyze a currency attack preceded by the choice of the interest rate made by Central Bank. They show that such policy interventions generate multiple equilibria. The key distinction between their setup and ours is that their "message" is costly: the level of interest rate directly affects the losses CB will suffer in the event of the currency attack. In our model, communication does not entail any direct costs, making it a cheap-talk (rather than a signalling) story.

The rest of the paper is organized as follows. Section I develops the general setup. Section II solves its special, analytically tractable version and discusses comparative statics. Section III provides intuition about endogenous conflict of interests. Section IV addresses welfare implications with exogenously biased public agent. Section V considers several applications of the basic framework. Section VI discusses robustness of our results. Section VII concludes.

I. The Model

A. *Environment*

Consider a uniform continuum of risk-neutral agents (receivers), indexed by $i \in [0,1]$. Each of them has a unit endowment, which can be allocated between the safe and the risky activity. We write $a_i = 0$ when agent i devotes his entire endowment to the safe activity and $a_i = 1$ when only the risky activity is undertaken. In general, we allow $a_i \in [0,1]$.

Payoff for the risky action is given by the function $R(\theta, A)$, which depends on

two variables: the state of economic fundamentals, $\theta \in \Theta$, and aggregate level of the risky action undertaken, $A \equiv \int_0^1 a_i di$. Payoff for the safe action does not depend on θ, but might still depend on A. We denote it by $r(A)$:

$$R : \Theta \times [0,1] \to \mathbb{R}, \quad r : [0,1] \to \mathbb{R}. \tag{1}$$

The fundamental θ is chosen by Nature before agents undertake actions. We assume it has commonly known prior with cdf Ψ[7]. Agent i's realized payoff under state θ and aggregate participation A is given by

$$\widetilde{\pi}_i(a_i; \theta, A) = (1 - a_i) r(A) + a_i R(\theta, A). \tag{2}$$

Before choosing a_i, each agent i observes a noisy private signal $x_i \in \mathcal{X}$, whose conditional distribution is denoted by $F(\cdot|\theta)$. We assume that x_i's are i.i.d. across agents.

There is a public agent (sender) who *might* observe another signal about θ. Specifically, we assume that with probability $p < 1$, she obtains an informative signal $y \in \mathcal{Y}$, drawn from a conditional distribution $H(\cdot|\theta)$. With the complementary probability $1 - p$, her signal remains uninformative. For notational convenience, we refer to this case as $y = \varnothing$. Conditional on θ, the sender's signal is assumed to be independent from $\{x_i\}_{i \in [0,1]}$.

We assume that the sender is benevolent[8]. Specifically, she cares about receivers' welfare, computed using the utilitarian aggregator. For any profile $\{a_i\}_{i \in [0,1]}$, her realized payoff is the sum of receivers' payoffs:

$$\widetilde{\Pi}\left(\{a_i\}_{i \in [0,1]}, \theta\right) = \int_0^1 \widetilde{\pi}_i(a_i; \theta, A) di. \tag{3}$$

Before receivers act, the sender can send a public message m revealing y. We

[7] In what follows, conditional posteriors will be denoted by the same letters, with the variables on which we condition explicitly specified.

[8] In Section IV, we relax this assumption and introduce an exogenous preference bias.

assume that y is hard information and cannot be falsified. However, the sender can voluntarily choose to withhold it[9]. Her reporting strategy can be equivalently described by partitioning the public message space $\mathcal{Y}$ into *disclosure* and *non-disclosure* regions. The latter is denoted by $\mathcal{Y}^{\mathrm{N}} \subseteq \mathcal{Y}$: the sender will voluntarily choose to send $m = \varnothing$ whenever $y \in \mathcal{Y}^{\mathrm{N}}$ is observed. Otherwise, if $y \in \mathcal{Y} \backslash \mathcal{Y}^{\mathrm{N}}$, she reports truthfully: $m = y$.

Correspondingly, we denote the message set by $\mathcal{M} = \mathcal{Y} \cup \{\varnothing\}$. Another assumption is that the sender cannot commit to her information disclosure policy: the choice of $\mathcal{Y}^{\mathrm{N}}$ must satisfy *ex-post* incentive compatibility constraints.

Receiver i chooses his action upon observing private signal x_i and the message m. His strategy is a mapping

$$a_i : \mathcal{X} \times \mathcal{M} \to [0, 1], \tag{4}$$

which for any pair (x_i, m) tells him which action to undertake.

The timing of the game is as follows:

1) The Nature draws θ from Θ according to Ψ.

2) The sender observes $y \in \mathcal{Y}$ (with probability p) or $y = \varnothing$ (with probability $1 - p$), while each receiver i observes $x_i \in \mathcal{X}$.

3) The sender sends a message, $m \in \mathcal{M}$.

4) Each receiver i chooses a_i upon receiving (x_i, m).

5) Payoffs are realized.

We make two sets of assumptions about the primitives:

ASSUMPTION 1: *The function $R(\theta, A)$ possesses the following properties:*

(i) For each $A \in [0, 1]$, $R(\cdot, A)$ is weakly increasing in θ;

[9]The sender who failed to get the signal (which occurs with probability $1 - p$) has no choice but to send $m = \varnothing$. Essentially, her behavior is non-strategic. Therefore, the only possible type of misreporting is to send $m = \varnothing$ whence the signal $y(\neq \varnothing)$ was actually observed.

(ii) For all $\theta \in \Theta$, we can either have $\frac{\partial R(\theta,A)}{\partial A} > 0$ or $\frac{\partial R(\theta,A)}{\partial A} < 0$;

(iii) $r(A)$ is monotone in A;

(iv) There are states, in which one extreme action dominates the other:

$$\inf_{\theta\in\Theta} \sup_{A\in[0,1]} (R(\theta, A) - r(A)) < 0 < \sup_{\theta\in\Theta} \inf_{A\in[0,1]} (R(\theta, A) - r(A)).$$

Condition (i) says that higher θ implies better payoff to the risky action for the receiver. Condition (ii) says that for any $\theta \in \Theta$, the payoff to the risky action either monotonically increases with A (actions exhibit strategic complementarities) or monotonically decreases with A (actions exhibit strategic substitutabilities). Condition (iii) is a regularity assumption implying that the payoff to the safe action is maximized at one of the extremes, $A = 0$ or $A = 1$[10]. Condition (iv) allows us to identify dominance regions for the continuation game played by receivers. It states that for θ low enough, $a_i = 1$ is strictly dominated by $a_i = 0$, regardless of what others do, and vice versa for θ high enough.

To ease exposition, we assume that distributions admit densities, denoted by lowercase letters: ψ for the prior, f for receivers' signals and h for the sender's signal. The following regularity conditions are imposed on F and H:

ASSUMPTION 2: *Conditional distributions $F(\cdot|\theta)$ and $H(\cdot|\theta)$ satisfy:*

i) Monotone Likelihood Ratio Property (MLRP): for all $\theta_1, \theta_2 \in \Theta$ such that $\theta_1 > \theta_2$, the functions

$$\frac{f(x|\theta_1)}{f(x|\theta_2)} \quad \text{and} \quad \frac{h(y|\theta_1)}{h(y|\theta_2)} \tag{5}$$

are increasing in x and y, respectively.

[10] A special but important case is constant $r(A)$.

ii) Precision at the extremes: for all $\theta_1 > \theta_2$,

(6)
$$\inf_{x\in\mathcal{X}} \frac{f(x|\theta_1)}{f(x|\theta_2)} = \inf_{y\in\mathcal{Y}} \frac{h(y|\theta_1)}{h(y|\theta_2)} = 0 \quad \textit{and} \quad \sup_{x\in\mathcal{X}} \frac{f(x|\theta_1)}{f(x|\theta_2)} = \sup_{y\in\mathcal{Y}} \frac{h(y|\theta_1)}{h(y|\theta_2)} = \infty.$$

Condition (i) states that higher signals correspond to "good news" in the sense of Milgrom (1981). Condition (ii) states that for extreme signal realizations, both the sender and the receivers are almost certain that the fundamental is either very high or very low. Those assumptions are satisfied for commonly used signal structures (for instance, Gaussian).

B. Equilibrium Definition

Let us specify the sender's and the receivers' objective functions. Given (x_i, m), receiver i chooses a_i to maximize

$$\mathbb{E}\left[\widetilde{\pi}_i(a_i;\theta,A)|x_i,m\right] = (1-a_i)r(A) + a_i\mathbb{E}[R(\theta,A)|x_i,m]. \tag{7}$$

It should be kept in mind that, from the perspective of receiver i, both the fundamental θ and the aggregate action A are considered random.

We characterize symmetric equilibria, in which all receivers adopt the the same threshold strategy: receiver i chooses $a_i = 1$ if and only if his private signal x_i exceeds a threshold $\hat{x}(m)$; otherwise, he chooses $a_i = 0$. That is[11],

$$a_i(x_i, m) = \mathbb{1}_{x_i \geq \hat{x}(m)}. \tag{8}$$

Therefore, given the state θ and the threshold $\hat{x}(m)$, by the Law of Large Numbers, aggregate risky action A can be written as

$$A(\theta, m) = \int_0^1 a_i(x_i, m)di = \int_{x\geq\hat{x}(m)} dF(x|\theta) = 1 - F(\hat{x}(m)|\theta). \tag{9}$$

[11]Standard arguments of iterated conditional dominance used in the global games' literature establish that any equilibrium strategy must be of this form.

Receiver i's problem 7 boils down to

$$\max_{a_i \in \{0,1\}} \int_{\Theta} \Big((1-a_i) r(A(\theta,m)) + a_i R(\theta, A(\theta,m)) \Big) d\Psi(\theta | x_i, m). \tag{10}$$

Given this threshold strategy, the receiver whose signal realization is exactly equal to the threshold has to be indifferent between $a_i = 0$ or $a_i = 1$:

$$\begin{aligned} 0 =& \mathbb{E}[R(\theta, A) - r(A) | x_i = \hat{x}(m), m] \\ =& \int_{\Theta} \Big(R(\theta, 1 - F(\hat{x}(m)|\theta)) - r(1 - F(\hat{x}(m)|\theta)) \Big) d\Psi(\theta | x_i = \hat{x}(m), m). \end{aligned} \tag{11}$$

Correspondingly, all the receivers who get $x_i > \hat{x}(m)$ strictly prefer $a_i = 1$ while all those who get $x_i < \hat{x}(m)$ prefer $a_i = 0$.

When receivers switch their action around $\hat{x}$, their aggregate payoff in state θ (which is also the benevolent sender's payoff) equals

$$\widetilde{\Pi}(\hat{x}, \theta) = F(\hat{x}|\theta) r(1 - F(\hat{x}|\theta)) + \big(1 - F(\hat{x}|\theta)\big) R(\theta, 1 - F(\hat{x}|\theta)). \tag{12}$$

The sender seeks to maximize expected value of 12, conditional on $y \in \mathcal{Y}$ and anticipating receiver's response, summarised by $\hat{x}(m)$. The sender's strategy is thus to choose m so as to solve

$$\max_{m \in \mathcal{M}} \Pi(\hat{x}(m), y) = \int_{\Theta} \widetilde{\Pi}(\hat{x}(m), \theta) d\Psi(\theta | y), \tag{13}$$

subject to the constraint that for each $y \in \mathcal{Y}$, $m \in \{y, \varnothing\}$ (feasibility) and that her disclosure choice satisfies incentive compatibility.

We now give the formal definition of equilibrium.

DEFINITION 1: *A Perfect Bayesian equilibrium consists of (i) the decision threshold for the receivers, $\hat{x}(m)$, (ii) the information revelation strategy for the sender, $\mathcal{Y}^N \subseteq \mathcal{Y}$, (iii) the conditional posterior $\Psi(\theta|m)$ and, (iv) the measure of aggregate participation $A(\theta, m)$, such that:*

1) Given $\mathcal{Y}^N$, $\forall m \in \mathcal{M}$, the posterior $\Psi(\theta|m)$ is consistent with Bayes' rule:

$$\psi(\theta|m) = \begin{cases} \dfrac{h(y|\theta)\psi(\theta)}{\displaystyle\int_{\Theta} h(y|\widetilde{\theta})\psi(\widetilde{\theta})d\widetilde{\theta}}, & \text{if } m = y, \\ \dfrac{(1-p)\psi(\theta) + p\displaystyle\int_{y\in\mathcal{Y}^N} h(y|\theta)\psi(\theta)dy}{1-p+p\displaystyle\int_{\Theta}\int_{y\in\mathcal{Y}^N} h(y|\widetilde{\theta})\psi(\widetilde{\theta})dyd\widetilde{\theta}}, & \text{if } m = \varnothing. \end{cases} \tag{14}$$

2) Given $A(\theta, m)$ and $\Psi(\theta|m)$, the threshold $\hat{x}(m)$ constitutes equilibrium in the receivers' continuation game.

3) Given (θ, m), aggregate participation $A(\theta, m)$ is determined by 9.

4) Given receivers' action profile $\{a_i(x_i, m)\}_{i=0}^1$, the sender optimally chooses her non-disclosure region so as to maximize 13: $\forall y \in \mathcal{Y}$,

$$\Pi(\hat{x}(\varnothing), y) \geq \Pi(\hat{x}(y), y) \quad \Longleftrightarrow \quad y \in \mathcal{Y}^N.$$

Before proceeding to equilibrium characterization, let us introduce welfare criterion that will be useful in the subsequent analysis. We focus on the *ex-ante* expected aggregate payoff for the receivers. Given the sender's strategy (that is, the non-disclosure region $\mathcal{Y}^{\mathrm{N}}$), ex-ante expected welfare is given by

$$\begin{aligned} V\left(\mathcal{Y}^{\mathrm{N}}\right) = \int_{\Theta} \Bigg\{ & \left(1-p+p\int_{y\in\mathcal{Y}^{\mathrm{N}}} dH(y|\theta)\right)\widetilde{\Pi}(\hat{x}(\varnothing),\theta) \\ & + p\int_{y\notin\mathcal{Y}^{\mathrm{N}}} \widetilde{\Pi}(\hat{x}(y),\theta)dH(y|\theta)\Bigg\} d\Psi(\theta). \end{aligned} \tag{15}$$

II. Public Communication

We start this section by outlining a special case that can be solved analytically. Specifically, we consider a linear payoff structure, together with normally distributed prior and signals. As we argue, our main conclusions can be general-

ized[12].

Suppose that, given θ and A, the payoffs from the risky and the safe actions are, correspondingly:

$$R(\theta, A) = \theta + \rho A \quad \text{and} \quad r(A) = 1 + \rho A. \tag{16}$$

This payoff structure captures the scenario where receivers choosing the risky action impose an externality on the other receivers, regardless of what the others do[13]. Parameter ρ captures the degree of coordination frictions in our model. When $\rho = 0$, which we refer to as the *frictionless* case, each receiver's risky payoff depends only on the fundamental but not on others' actions. In general, we allow ρ to be either positive or negative.

Next, assume that θ, as well as x and y, are normally distributed:

$$\theta \sim \mathcal{N}(\theta_0, \gamma^{-1}), \quad x|\theta \sim \mathcal{N}(\theta, \beta^{-1}), \quad y|\theta \sim \mathcal{N}(\theta, \alpha^{-1}), \tag{17}$$

where γ, β and α represent, respectively, the precision of the prior, the private signal and the public signal.

A. *The Continuation Game*

Suppose that the sender reveals y and the receiver i privately observes x_i. Since he will get ρA regardless of his action, by 11, he will strictly prefer to choose $a_i = 1$ iff

$$\mathbb{E}[\theta|x_i, y] > 1, \tag{18}$$

which boils down to $\frac{\alpha y + \beta x_i + \gamma\theta_0}{\alpha+\beta+\gamma} > 1$.

[12] See Section VI.A where robustness is discussed.

[13] For example, we can think of an effort from one employee in a team, which increases the productivity of other team members, whereby the private benefit from shirking is fixed at 1.

We call

$$\hat{x}(y) \equiv \frac{\alpha+\beta+\gamma}{\beta} - \frac{\alpha y + \gamma\theta_0}{\beta} \tag{19}$$

the *truthful* threshold: the receiver whose private signal $x_i = \hat{x}(y)$ will be indifferent between $a_i = 1$ and $a_i = 0$, given y.

For notational convenience, we denote by $\hat{x}(\varnothing)$ the truthful threshold given the public message $m = \varnothing$[14]. Likewise, we denote by

$$\hat{x}_0 \equiv \frac{1}{\beta}(\beta+\gamma-\gamma\theta_0) \tag{20}$$

the truthful threshold when there is no public signal: $\hat{x}_0$ could be thought of as the cutoff the receivers would have used, if the sender had not had the option to withhold information.

Before we characterize the full game equilibrium, it is useful to introduce the *wishful* threshold, which is the cutoff for the private signal that a benevolent sender would like the receivers to use:

$$x^*(y) \in \arg\max_x \Pi(x,y), \tag{21}$$

where $\Pi(x,y)$ is given by 13 with $\hat{x}(m)$ replaced by x.

When $\rho \neq 0$, risky actions exhibit externalities on the entire group. As the receivers do not internalize them, their equilibrium response (as given by $\hat{x}(\cdot)$) is inefficient: as the next proposition demonstrates, the benevolent sender would like receivers to act more aggressively when these externalities are positive and more conservatively when they are negative.

PROPOSITION 1: *Conditional on y, the wishful threshold $x^*(y)$ is given by*

[14]In general $\hat{x}(\varnothing)$ will depend on the sender's revelation strategy given by $\mathcal{Y}^{\mathrm{N}}$.

$$x^*(y) = (1-\rho)\frac{\alpha+\beta+\gamma}{\beta} - \frac{\alpha y + \gamma\theta_0}{\beta}.$$

PROOF:

In the Appendix.

An immediate corollary of this result concerns the sender's information disclosure strategy in any equilibrium: unless $\rho = 0$, due to discrepancy between $\hat{x}$ and x^*, some sender types $y \in \mathcal{Y}$ will always prefer to withhold information, rendering complete revelation impossible.

COROLLARY 1: *Full information disclosure is an equilibrium strategy for the sender if and only if $\rho = 0$, i.e. if there are no coordination frictions.*

PROOF:

The "if" part follows from Proposition 1: when $\rho = 0$, $\hat{x} \equiv x^*$, meaning that no sender would like to withhold her information *ex post*.

The "only if" part is proven by contradiction. Without loss of generality, take $\rho > 0$. Suppose that the sender were to always reveal y. In equilibrium, $m = \varnothing$ would then indicate that the sender indeed got no signal, implying $\hat{x}(\varnothing) = \hat{x}_0$. However, since for all $y \in \mathcal{Y}$, $\hat{x}(y) > x^*(y)$ and the range of $\hat{x}$ and x^* is $\mathbb{R}$, for any $\hat{x}_0$, there exists such $y_0 \in \mathcal{Y}$ that $\hat{x}(y_0) > \hat{x}_0 > x^*(y)$. But the sender who gets y_0 will be better off withholding her information, contradicting that y is always revealed.

The case for $\rho < 0$ can be treated analogously.

The intuition for the above corollary is that coordination frictions (as captured by ρ) create an endogenous conflict of interests between the sender and the receivers, due to which public information provision in equilibrium turns out to be coarse. This endogenous conflict of interests will be discussed in greater detail in Section III

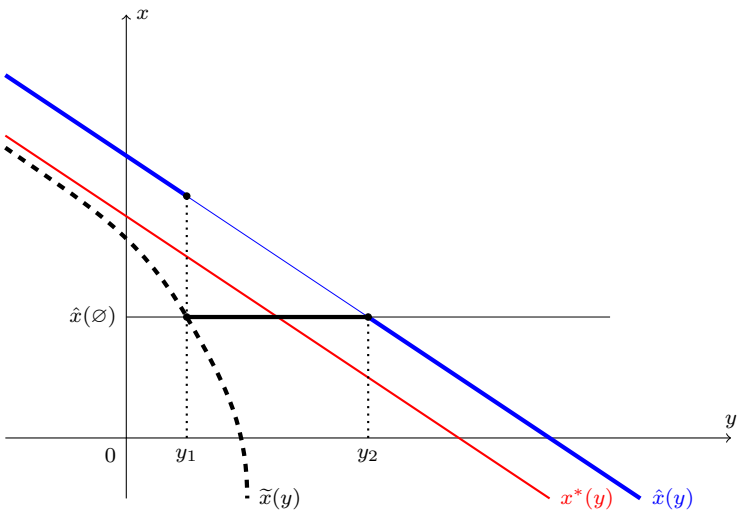


Figure 1. : Equilibrium thresholds and $\mathcal{Y}^{\mathrm{N}}$ (case with $\rho > 0$).

B. Equilibrium and the non-disclosure region

In this section we characterize the full game equilibrium where the sender's information revelation strategy is pinned down by her incentive compatibility conditions.

THEOREM 1: *When $\rho \neq 0$, an equilibrium of the full game is solved by the triple $\{\hat{x}(\varnothing), y_1, y_2\}$ satisfying $y_1 \neq y_2$ and the following 3 conditions:*

$$1 = \mathbb{E}\left[\theta | x_i = \hat{x}(\varnothing), y \in \{\varnothing\} \cup \mathcal{Y}^N\right], \tag{22}$$

$$\hat{x}(\varnothing) = \hat{x}(y_2), \tag{23}$$

$$\Pi(\hat{x}(\varnothing), y_1) = \Pi(\hat{x}(y_1), y_1). \tag{24}$$

where $\mathcal{Y}^N = [y_1, y_2]$ if $\rho > 0$ and $\mathcal{Y}^N = [y_2, y_1]$ if $\rho < 0$.

PROOF:

In the Appendix.

Equilibrium characterization provided in Theorem 1 can be most easily understood with the help of Figure 1[15]. The blue line represents $\hat{x}(y)$, the threshold that the receivers choose when y is disclosed. The red line $x^*(y)$ illustrates the sender's wishful threshold. As can be seen, in case of positive externalities ($\rho > 0$), the red line lies everywhere below the blue line, showing that the sender would wish to over-report y, if this were possible.

The dashed line $\widetilde{x}(y)$ represents an *auxiliary* threshold implicitly defined by $\Pi(\widetilde{x}(y), y) = \Pi(\hat{x}(y), y)$. In words, the sender with signal y is indifferent between picking $\widetilde{x}(y)$ and $\hat{x}(y)$, also strictly preferring any $x \in (\widetilde{x}(y), \hat{x}(y))$. Since the three threshold functions are monotonic, an arbitrary $\hat{x}(\varnothing)$ will cross each of them exactly once, yielding a corresponding pair $y_1(\hat{x}(\varnothing))$ and $y_2(\hat{x}(\varnothing))$. Type y sender prefers $x^*(y)$ to anything else, while being indifferent between $\hat{x}(y)$ and $\widetilde{x}(y)$. As a result, given $\hat{x}(\varnothing)$, the sender will withhold her information for all $y \in (y_1(\hat{x}(\varnothing)), y_2(\hat{x}(\varnothing)))$ and disclose otherwise.

Condition 22 is needed to ensure that, in equilibrium, the threshold corresponding to the empty message $\hat{x}(\varnothing)$ is consistent with receivers' Bayesian updating, given the non-disclosure region $\mathcal{Y}^{\mathrm{N}}$. Corollary 1 explains why full disclosure cannot be an equilibrium strategy for the sender as long as $\rho \neq 0$. When the receivers observe an empty public message $m = \varnothing$, they know that either the sender indeed observed no signal, or that $y \in \mathcal{Y}^{\mathrm{N}}$. Each receiver uses this knowledge, along with his private signal, to update his posterior on θ.

Finally, we have to specify out-of-equilibrium beliefs. Since $p < 1$, there is always a strictly positive probability that the sender does not get any signal, and as she cannot make faulty reports, $m = \varnothing$ will be observed with strictly positive probability in *any* equilibrium. Hence, the only out-of-equilibrium play one can conceive is when the sender reports $y \in \mathcal{Y}^{\mathrm{N}}$, i.e. some news she was supposed to remain silent about. We assume that if the sender were revealing such signal $m = y \in \mathcal{Y}^{\mathrm{N}}$, as a probability 0 event in equilibrium, the receivers will update

[15] A similar construction was used by Che and Kartik (2009).

their belief as if they see $m = \varnothing$. Clearly, there will be no deviation for any type of sender given this particular out-of-equilibrium belief.

Although our assumptions guarantee uniqueness of equilibrium in every continuation game, the full game can nevertheless have multiple equilibria. In fact, our model has a self-fulfilling nature of the message $m = \varnothing$. To see this more clearly, suppose that the sender were to expect the receivers to play relatively high $\hat{x}(\varnothing)$ upon receiving no news. As can be inferred from Figure 1, this implies a relatively low non-disclosure region (low y_1 and y_2), in turn yielding pessimistic posteriors on θ upon $m = \varnothing$ and, as a consequence, justifying high $\hat{x}(\varnothing)$ in the first place. However, the propositions etsablished in Section IV do not depend on the particular equilibrium selection.

III. Endogenous Conflict of Interests

In this section, we give a more detailed account of the conflict of interests that is endogenously created by coordination frictions. Specifically, we consider various economic scenarios with different payoff structures.

As we have demonstrated in the previous section, when the receivers' risky actions exhibit positive (negative) externalities, the sender has an incentive to induce higher (lower) participation in the risky action. Let us define by $\Delta(y)$ the wedge between the wishful threshold (the one that the sender would want the receivers to use, conditional on her information y) and the receivers' equilibrium threshold:

$$\Delta(y) \equiv x^*(y) - \hat{x}(y). \tag{25}$$

Under the payoff structure used in Section II, we had

$$\Delta(y) = -\rho\left(1 + \frac{\alpha + \gamma}{\beta}\right). \tag{26}$$

The direction of the endogenous conflict of interests depends on the sign of ρ.

However, the scale is a constant which does not depend on the sender's information y. Intuitively, this is due to the fact that each receiver's payoff is affected in the same way regardless of which a_i he undertakes.

As a result, the wedge between the marginal social benefit from a risky action and the marginal private benefit is a constant, as well as the wedge between x^* and $\hat{x}$. In general, under other payoff structures the endogenous conflict of interests can be *state-dependent*. It is also worth noting that $\Delta(y)$ is increasing in the precision of the public signal relative to the private signal. The reason is that each receiver's action becomes more sensitive to the public information, requiring a larger change of the private threshold in order to correct the wedge.

A. *Alternative payoff structures*

Previously we assumed that coordination frictions symmetrically affected the risky payoff and the safe payoff: there was a common term ρA appearing in both $R(\theta, A)$ and $r(A)$ functions. We now present several alternative payoff structures applicable in various economic scenarios.

Strategic complementarities

In many situations, aggregate risky action exhibits spillover effects only to those receivers who also chose it[16]. In such situations, receivers' actions exhibit strategic complementarities. Compared to 16, the payoff structure would be

$$R(\theta, A) = \theta + \rho A \quad \text{and} \quad r(A) = 1. \tag{27}$$

Parallel to 18, receiver i is indifferent between $a_i = 0$ and $a_i = 1$ when

$$\mathbb{E}[\theta + \rho(1 - F(\hat{x}(y)|\theta))|x_i = \hat{x}(y), y] = 1. \tag{28}$$

[16]For instance, when a given creditor decides to rollover his debt to the firm, the firm's liquidity constraint becomes less tight, yielding a lower probability of default. As a result, this creditor's action imposes positive externalities to other creditors who have also invested in this firm; however, it does not affect those who have invested in safe assets instead.

The next proposition shows that under this payoff structure, the endogenous conflict of interests depends on the sender's information:

PROPOSITION 2: *When the payoff structure is given by 27, and assuming*

$$\rho \in \left[0, \sqrt{\frac{\pi\beta(\alpha+2\beta+\gamma)}{2(\alpha+\beta+\gamma)(\alpha+\gamma)^2}}\right),$$

both $\hat{x}(y)$ and $x^*(y)$ are unique and monotonically decreasing for all $y \in \mathcal{Y}$. Moreover, $\Delta(y)$ possesses the following properties:

(i) When receivers' actions are strategic complements (substitutes), $\Delta(y)$ is negative (positive), for all $y \in \mathcal{Y}$.

(ii) Furthermore,

$$\lim_{y\to-\infty} \Delta(y) = 0 \quad \textit{and} \quad \lim_{y\to+\infty} \Delta(y) = -\rho \cdot \frac{\alpha+\beta+\gamma}{\beta}.$$

PROOF:

In the Appendix.

Proposition 2 shows that the conflict of interests is state-dependent: the wedge between the wishful threshold $x^*(y)$ and truthful threshold $\hat{x}(y)$ is small for low realizations of y and converges to 0 as $y \to -\infty$; on the other end, as $y \to +\infty$, the wedge $\Delta(y)$ converges to a constant.

To gain some intuition, notice that the marginal social benefit from the aggregate risky action (rA) is proportional to the measure receivers who have already chosen $a_i = 1$. When the sender observes low y, she expects a small A. As a result, she has little incentive to locally manipulate participation: if y is extremely low, most receivers will not choose the risky action anyway, and so we have $\Delta(y) \to 0$. On the other hand, when the sender observes high y, she anticipates that many receivers will choose $a_i = 1$, and hence the marginal social benefit will also be high. The wedge converges to a constant because $\mathbb{E}[A|y] \to 1$ as $y \to +\infty$.

Complementarities with negative externalities: Runs

Some applications (the most prominent is the model of a bank run) demonstrate that complementarites and externalities may operate in the opposite directions. Specifically, in the context of a bank run, withdrawals by others raise each individual's incentive to withdraw, but massive withdrawals also hurt him.

In that case, we can write the receiver's payoff function as

$$\widetilde{\pi}(a_i;\theta,A) = a_i + (1-a_i)R(\theta,A), \tag{29}$$

where a_i stands for the amount that depositor i chooses to withdraw early. Greater aggregate early withdrawal decreases the return to the late withdrawal, implying $R_A < 0$. Therefore, depositors' actions exhibit strategic complementarities along with negative externalities to those who do not withdraw.

However, notice that for this case, $a_i = 1$ should be appropriately referred to as the "safe" action.

Discontinuity in the Payoff: Regime Change

Our basic framework can be adapted to the following version of the speculative currency attack model.

Consider the interaction between the Central Bank (sender) and the group of speculators (receivers). Each of them can choose either not to attack the current currency peg ($a_i = 0$), receiving a normalized return of 1; or to attack ($a_i = 1$) and receive payoff R that depends on whether the peg has been abandoned or not. Assume that it is abandoned if and only if the aggregate attack is sufficiently high compared to how stable this regime is, namely $A \geq 1-\theta$. To summarize:

$$r(A) \equiv 1 \quad \text{and} \quad R(\theta,A) = \begin{cases} \delta \leq 1, & \text{if } A < 1-\theta, \\ \bar{R} \geq 1, & \text{if } A \geq 1-\theta. \end{cases} \tag{30}$$

Let us assume that the sender always receives Gaussian signal y, whereas re-

ceivers get i.i.d. Gaussian signals $\{x_i\}$. Receivers' strategies in the continuation game under this payoff structure will be characterized by a pair of thresholds $\hat{x}(y)$ and $\hat{\theta}(y)$. In equilibrium, receiver i will choose to attack iff his private signal $x_i > \hat{x}(y)$, and the regime is abandoned iff $\theta > \hat{\theta}(y)$.

Naturally, both thresholds will depend negatively on y. Furthermore,

$$\lim_{y \to +\infty} \hat{\theta} \to 0 \quad \text{and} \quad \lim_{y \to -\infty} \hat{\theta} \to 1.$$

In this regime change game, there is also a (negative) endogenous conflict of interests between a benevolent sender and receivers. The reason is that, since the payoff from abandoning the peg ($\bar{R}$) always exceeds the payoff from maintaining it (δ) for receivers who choose to attack, *ex post* the sender would be willing to abandon the peg more often, i.e. for a larger set of realised θ.

As a result, the sender would wish the receivers to attack more aggressively, given the sender's information is publicly revealed. This is summarized in the following proposition:

PROPOSITION 3: *When the payoff structure is given by 30, the conflict of interests is negative:* $\Delta(y) < 0$. *Furthermore,*

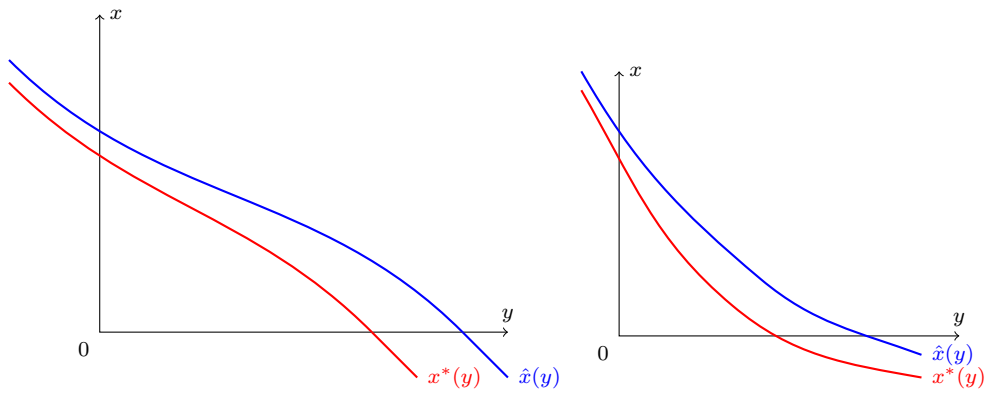
$$\Delta(y) \to 0 \quad \textit{as} \quad y \to -\infty \quad \textit{or} \quad y \to +\infty$$

PROOF:

In the Appendix.

The endogenous conflict of interests in the regime change game is also state-dependent. It is becomes vanishingly small at both extremes, for very high and very low y[17]. However, since the sender evaluates this event conditional on her signal, the probability that the realized θ is close to $\hat{\theta}$ converges to 0 for extreme y's. In other words, coordination concerns only matter when the sender has

[17]Recall that the sender's incentive to manipulate comes from the fact that receivers are always better off if $\hat{\theta}$ is marginally lower, namely when the regime is abandoned more often.



(a) Strategic complementarity (b) Regime change

Figure 2. : Endogenous conflict of interests in alternative cases.

an intermediate posterior about θ, being less certain about the outcome of the currency attack.

IV. Exogenous Preference Bias

In this section, we generalize our model by introducing an exogenous bias into the sender's preferences. We analyze its welfare implications in communication games with coordination frictions. Our main result establishes that a small exogenous bias in an appropriate direction (offsetting an endogenous one) can mitigate the conflict of interests, achieving better information transmission and raising welfare. For technical simplicity, we return to the framework with externalities outlined in Section II However, in Section VI.A we argue that our conclusions also apply to other payoff structures.

Assume that the sender's bias comes from the private benefit (or cost) that she gets whenever receiver i plays $a_i = 1$[18]. Denote this private benefit by b.

[18]For example, a credit rating agency may collect extra fees from the issuer if its debt is favorable by the market. Likewise, a firm's CEO may have her own preference over different projects.

Now the sender's expected payoff is given by

$$\Pi(x,y,b)=\int_{-\infty}^{+\infty}\Big\{F(x|\theta)r(1-F(x|\theta)) + \Big[1-F(x|\theta)\Big]\Big(R\big(\theta,1-F(x|\theta)\big)+b\Big)\Big\}d\Psi(\theta|y). \tag{31}$$

As before, let $x^*(y,b)=\arg\max_x \Pi(x,y,b)$ denote the sender's wishful threshold, given her exogenous bias b. It is straightforward to see that $\frac{\partial}{\partial b}x^*(y,b)<0$. When the sender's benefit is positive, she would like to boost participation (by implementing an even smaller threshold) compared to the case of a benevolent sender. Conversely, when the risky action imposes some extra costs on her ($b<0$), she would prefer to abate participation by raising x^*.

Replacing 24 with 31 throughout, equilibria of the communication game with an exogenously biased sender can be solved in a similar fashion following Theorem 1.

PROPOSITION 4: *When the sender has exogenous bias b and externality parameter is ρ, the equilibrium is solved by a triplet $\{y_1(b,\rho),y_2(b,\rho),\hat{x}(\varnothing,b,\rho)\}$, which coincides with $\{y_1(0,\rho+b),y_2(0,\rho+b),\hat{x}(\varnothing,0,\rho+b)\}$.*

PROOF:

In the Appendix.

Proposition 4 shows that the characterization of an equilibrium with the sender's bias b is the same as the characterization of an equilibrium with benevolent sender as in Theorem 1, but with a different degree of coordination frictions: $\rho+b$ instead of ρ. An immediate implication is that full information disclosure becomes possible whenever $b=-\rho$.

A. Welfare-Improving Bias

In this section, we show that sender's bias can not only improve information transmission, but also raise aggregate welfare. First, we present a lemma showing

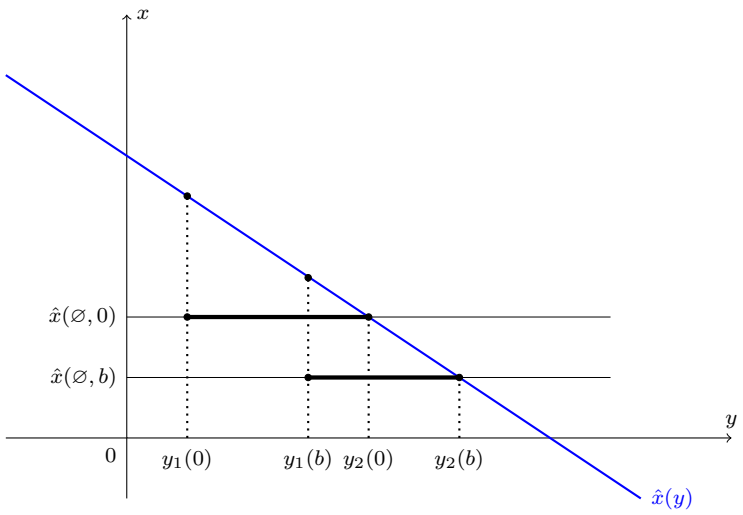


Figure 3. : Equilibrium thresholds with the biased sender ($\rho > 0, b < 0$)

how equilibrium threshold played by receivers when they get no public message, $\hat{x}(\varnothing, b)$, changes with b.

LEMMA 1: *The threshold $\hat{x}(\varnothing, b)$ is strictly increasing in b:*

$$\frac{\partial \hat{x}(\varnothing, b)}{\partial b} > 0.$$

PROOF:

In the Appendix.

To get a sense of the mechanism at work, recall Figure 1. A decrease in b leaves the equilibrium threshold $\hat{x}(y)$ intact, at the same time shifting up the desired threshold $x^*(y, b)$. Suppose receivers continued to play the same $\hat{x}(\varnothing)$. Then the sender who observes y_1 would strictly prefer to reveal it. This implies that y_1 should have increased. Since the receivers' posterior of θ (conditional on $m = \varnothing$) first-order stochastically increases in y_1, it in turn calls for a decline in $\hat{x}(\varnothing)$ in order to keep the marginal receiver's posterior mean equal to 1.

Figure 3 provides an intuition for why the sender's preference bias may correct the inefficiency and improve welfare. The blue line represents the truthful threshold $\hat{x}(y)$. The black lines represent equilibrium thresholds upon an empty message for two values of b ($b = 0$ and $b < 0$). The two thick segments correspond to the non-disclosure regions[19].

When $\rho > 0$, a benevolent sender ($b = 0$) has an incentive to implement a lower threshold: $x^*(y, 0) > \hat{x}(y, 0)$. However, when the sender incurs a private cost from the receivers' risky actions (i.e. when $b < 0$), the threshold $x^*(y, b)$ shifts up accordingly. This partially cancels out the wedge caused by the endogenous conflict of interests. As a result, equilibrium communication can be more efficient compared to the case of a fully benevolent sender[20].

The next proposition establishes normative implications of introducing a preference bias.

PROPOSITION 5: *Compared to the case of benevolent sender ($b = 0$), a small negative preference bias ($b < 0$) achieves higher* ex-ante *welfare V, whenever risky actions exhibit positive externalities ($\rho > 0$). Likewise, a small positive preference bias ($b > 0$) achieves higher welfare, whenever risky actions exhibit negative externalities ($\rho < 0$).*

PROOF:

In the Appendix.

The intuition is as follows. Since $\hat{x}(y)$ does not depend on b, receivers' response for $m = y$ outside $\mathcal{Y}^{\mathrm{N}}$ does not change. Hence, the entire impact on V comes from the change in the bounds y_1 and y_2 and from shift in $\hat{x}(\varnothing)$.

Let us think of the fictitious sender type who knows *only* that $y \in \mathcal{Y}^{\mathrm{N}}$ (that is, observes an interval of possible y's) and chooses $m = \varnothing$. Such a sender would suffer a second-order loss from a marginal change in b from $b = 0$. However, the

[19] Note that $y_1(0)$ ($y_1(b)$) is pinned down by the intersection between $\hat{x}(\varnothing, b)$ and $\widetilde{x}(y, b)$ ($\hat{x}(\varnothing, b)$ and $\widetilde{x}(y, b)$), with $\widetilde{x}(y, 0)$ and $\widetilde{x}(y, b)$ being omitted from the figure.

[20] In fact, we can see from Proposition 4 that, as long as the sender's bias $b = -\rho$, the sender is able to fully disclose her information in equilibrium.

corresponding gain for the sender who is genuinely uninformed (i.e. the one who sees $y = \varnothing$) from the shift in the receivers' response $\hat{x}(\varnothing)$ is of the first order. Therefore, overall aggregate welfare goes up.

Proposition 5 implies that the receivers can be better off with an exogenously biased public agent compared to the case of a benevolent public agent. The reason is that a small exogenous bias in an appropriate direction offsets the endogenous bias, helping the sender overcome inefficient equilibrium information transmission. Note that whether a bias is welfare-improving or not depends on the nature of coordination: when receivers' risky actions exhibit positive externalities, they are better off dealing with a "conservative" sender, while an "aggressive" sender would do a better job whenever risky actions exhibit negative externalities.

The bottom-line message is that benevolence is not necessarily welfare-improving *per se*, which is relevant for many economic situations. For example, a credit rating agency that internalizes the cost of coordination failure in the financial market will be reluctant to provide bad news to investors. Even though this concern can be appreciated *ex post*, it also impedes communication *ex ante*. Our theory suggests that investors would be better off if the credit rating agency had a negative preference bias – that is, using a more conservative rating policy, giving good ratings less often.

V. Applications

This section provides several economic applications that fit into the structure of our general model outlined in Section I We discuss the positive and normative content of our key results in the context of disclosure policies by the credit rating agencies, announcements made by the Central Bank and the role of leadership in organizations.

A. Credit Rating Agencies and Investors

The credit rating industry as an important player in the financial market, especially the debt market, has drawn much criticism after the 2007-2008 financial crisis. Investors blamed rating agencies for their failure in providing accurate ratings on structured securities[21]. In addition, many have pointed out that rating inflation was a result of the conflict of interests between CRAs and investors, since the former are typically paid by issuers[22].

Currently, there are many policy-oriented studies on CRAs reform, focusing on the question how to improve the accuracy of the credit ratings[23]. One approach focusses on correcting CRAs' biases. For example, Diomande et al. (2009) and Lynch (2010) propose public-funded credit rating agencies, which would ideally represent the best interests of investors. However, we argue that this approach will not be effective enough, since even an unbiased rating agency will have an incentive to misreport.

Indeed, there are other strategic motives for CRAs to perform not only as passive information providers.

First, CRAs can have real (feedback) effects on the cost of capital for issuing firms, whose financial contracts are often contingent on their credit ratings (see Kliger and Sarig (2000), Kisgen and Strahan (2010) or Chen et al. (2012)). As a result, a downgrade of a particular firm's bonds not only reveals more information for investors, but also puts that firm at a risk of a credit-cliff.

Second, as institutional investors are constrained by regulation to invest only into bonds rated above investment-grades, rating inflation can actually favor those

[21]There are many empirical studies, including Cornaggia and Cornaggia (2013), Jiang et al. (2012) and Benmelech and Dlugosz (2010), suggesting that credit ratings tend to be inflated.

[22]Though being regulated, CRA's are still profit-maximizing firms whose main gains come from the fees paid by security issuers. As demonstrated by Bolton et al. (2012), the trusting nature of investors, due to either naiveness or regulatory constraints, makes rating inflation sustainable in equilibrium. Skreta and Veldkamp (2009) show that competition among rating agencies exacerbates rating inflatiobn, even if the ratings are unbiased and independent, since the issuers can selectively announce the most favorable rating, the phenomenon that has been known as "rating shopping".

[23]See Medvedev and Fennell (2011) for a survey on related policy proposals and debates. Also see White (2010) for an excellent review on the institutional background of credit rating agencies.

investors who can thereby earn higher yields (see Calomiris (2009), Opp et al. (2013) and Cornaggia and Cornaggia (2012)).

Third, credit ratings can act as a coordination device, and hence be self-fulfilling. In Boot et al. (2006), investors ask a high premium for a low-graded firm, in turn inducing this firm to engage in excessive risk.

For these reasons, even in absence of conflict of interests, a rating agency (like a benevolent sender in our model) has an incentive for *ex-post* information manipulation. In this case, what our theory suggests is that exogenous negative biases can be welfare-improving. That is, in order to overcome inefficient communication issue, CRAs should be exogenously conservative, i.e. having additional private incentives to reveal bad news.

In practice, such conservatism can be implemented through an asymmetric regulatory procedure: for example, more strict assessments when a CRA makes up-gradings, compared to when it makes down-gradings.

B. Monetary policy and forward guidance

The idea that public statements made by the Central Bank concerning its intentions can guide agents' expectations about future policy conduct has been reflected in the literature on forward guidance[24].

Relating to this issue, one may regard the sender from our model as the Central Bank engaging in inflation or interest rate targeting, while the group of receivers can be thought of as firms making some strategic choice (such as the price, investment scale, entry/exit decision etc).

The literature distinguishes two forms of forward guidance: one is called *Odyssean*, whereby the Central Bank commits to a particular action in the future; the other is called *Delphic*, which is confined to forecasts and likely future policy actions, leaving the policymaker's hands untied. The virtue of the former is credibility

[24]Recent contributions include Campbell et al. (2012); Goodhart (2013); Praet (2013); Williams (2013) and most notably Woodford (2013). Empirical studies on managing market expectations may be found in Neuenkirch (2012, 2013).

and relative simplicity (cf. the Romer proposal[25]), however it also leaves little room for monetary discretion. The latter's main advantage is its flexibility, which comes at the expense of higher uncertainty.

Although our model does not allow the sender to make her statements deliberately vague (e.g. by adding noise to her report), save for the extreme form of vagueness when nothing is reported ($m = \varnothing$), our results suggest that the policymaker would wish to make more precise announcements during the periods of inflationary booms or severe downturns.

In that respect, our paper contributes to the discussion of the variation in monetary policy in normal times and during crises (see Fahr et al. (2013) for the recent survey).

C. Organizations

There is a vast economic literature on leadership in organizations[26]. One important challenge for a leader is how she should strategically convey her information to team members in order to achieve a more efficient team (see Dewan and Myatt (2008), Majumdar and Mukand (2007) and Ferreira and Rezende (2007)). Our model captures this feature, while the coordination issue is also a classic team problem (as in Hermalin (1998)).

When team members' efforts are complementary in production, one member will exert high effort only if the others are exerting high efforts as well. Typically, there will exist multiple equilibria, including a productive one, in which all team members exert high effort, and an unproductive one, in which all team members exert little or no effort.

Our model shows that when a team leader is benevolent, she will be reluctant to reveal "bad" news, which may cause her team members to coordinate into an unproductive equilibrium[27]. However, this concern for the leader makes her

[25]For details, see Romer (2011).
[26]See Bolton and Dewatripont (2011) for a survey.
[27]In reality, there are also other reasons that a benevolent leader does not want to reveal bad news, for

use her information inefficiently, while the team suffers from an informational efficiency loss. Our model suggests that this efficiency loss can be mitigated if the leader had a biased interest.

In fact, many studies have demonstrated that personal characteristics of a leader play a very important role in organizations. Bolton et al. (2012) present a model having a flavor similar to ours. In their model, the leader continuously receives information about the unknown economic state and takes a corresponding action. The coordination among team members is insufficient due to a time inconsistency issue, because the leader can adapt her initial action in response to the new information arriving in the future. Hence, the team can benefit from the resolute leader, the one who would prefer to stick to her initial plan.

Landier et al. (2009) also reach a conclusion similar to ours by showing that it can be beneficial for an organization to have a leader and team members with heterogeneous preferences, though for a different reason. In their model, the preference heterogeneity (the dissent) is valuable since it helps disciplining the leader, inducing her to make her decisions based on the objective information rather than on her own preferences.

VI. Discussion

This section provides an informal discussion of the several aspects of our model. We start by addressing robustness issues concerning receivers' payoff specification and distributional assumptions. Then we briefly discuss out-of-equilibrium beliefs, potential sources of multiplicity and the nature of information. Finally, we mention one possible extension of the current framework: to compare with the persuasion literature, we describe an alternative setup where the sender can commit to the reporting strategy before observing the signal.

example, if bad news destroy the confidence of her team members and hence their performance (Compte and Postlewaite (2004)).

A. Robustness

For the main body of our paper we have assumed linearity of the payoff $R(\theta, A)$ and $r(A)$, coupled with normality of θ, x_i's and y. This provided us with an extra structure and facilitated equilibrium characterization. However, the main conclusions of our paper go through with the more general specification, such as those alternative payoff structures we have presented in Section III.A.

Regarding to the welfare implication, it should be noted that the essential elements used in the proof is that the sender's indifference condition on her cutoff types as well as the condition that receivers do correct Bayesian updating in equilibrium. These, however, are satisfied according to the definition of equilibrium and are not subjective to any specific payoff structure or information structure.

It should also be noted that, even though the qualitative results are robust in more general settings, the quantitative parts do not necessarily survive. For example, the intuition that coordination frictions create endogenous conflict of interests, which impedes a benevolent public agent to reveal her information *ex post*, is still valid beyond the linear-Gaussian setup. However, the exact shape of $\Delta(y)$ may various across different assumptions.

B. Hard versus Soft Information

In our model, the messages sent by the sender are assumed to be verifiable. That is, the sender's information is hard. The games of hard information disclosure is first introduced in Milgrom (1981). Yet, the information is often modeled to be soft as well. In cheap-talk games, a message has no literal meaning and a receiver can never tell the sender is lying or not by reading the message itself.

In this paper, we do not take no stance on which modeling approach is more plausible than the other. We choose the first approach merely because it simplifies our analysis. It is should not be a surprise that our story also goes through in a cheap-talk setting. Suppose that the sender can only use soft messages to reveal her information to receivers.

Due to endogenous conflict of interests, the equilibrium communication would exhibit a "multiple-step" partition of sender's signal space as a feature of cheap-talk games. Usually we will have multiple equilibria, but it is possible that there is a unique babbling equilibrium if the coordination friction is large enough. When sender also has exogenous bias, which is offsetting the endogenous one, the equilibria set expands by including more informative equilibrium with more "steps".

C. Communication with Commitment

In our setting, the sender decides her message after the signal $y \in \mathcal{Y}$ has already been observed. One may ask what if the sender were able to commit to a revelation policy before her signals are realized, as commonly assumed in persuasion games.

In this case, it is no longer a game played by sender and receivers, but a planning problem faced by the sender alone. As suggested by Kamenica and Gentzkow (2011), in that case the sender's maximization program could be equivalently formulated as looking for the optimal choice of a distribution of posteriors, which is Bayesian-consistent with the prior. They have shown that full revelation is optimal if and only if the sender's objective is convex in the posteriors. With quadratic loss utilities, as in Crawford and Sobel (1982), one can easily show that it is optimal for the sender to reveal all information if she can commit.

However, the $\Pi(x, y)$ function in our model does not have much regularities hence the convexity analysis is very difficult to perform even in the simple Linear-Gaussian case. Whereas Proposition 5 suggests that there is scope for improvement by marginally more infroamtion disclosure, yet, whether full information revelation is the optimal policy for the sender remains unclear to us.

VII. Concluding Remarks

We constructed a model describing communication of the sender with the large audience. As we have shown, as long as there exists some interdependence in the receivers' payoffs (coordination frictions), sender's objectives would typically

diverge from those of any given receiver *having the same information*. As a consequence, the sender would sometimes wish not to share her knowledge, despite the fact that her goal is to maximize collective well-being.

One distinctive feature of the communication equilibrium we construct is the structure of the non-disclosure region: as we show, the sender would prefer to disclose relatively extreme news and withhold some information "in the middle". This construction resembles the one obtained by Che and Kartik (2009).

Although mechanical reasoning is similar, economic intuition differs: in their setup, the sender's bias was constant and exogenously given. The sender would wish the receiver to be as close to his most preferred action as possible, which can sometimes be best achieved by withholding information.

In our setup, the sender's bias was linked to the primitives: the relative precision of her information versus that of the receivers' and the degree of coordination frictions. Public message both improves individual decision making and facilitates coordination: when the sender's news are extremely good or bad, the former concern dominates (the sender prefers *almost* everyone to do the same thing) and the news get disclosed; when the news are relatively neutral, the latter concern dominates, and the sender prefers to withhold her information.

A natural implication of the endogenous conflict of interests is that there would generally exist an offsetting *exogenous* bias correcting this distortion. Informally, depending of its direction, such bias could be referred to as the overall "conservativeness" (or else the "liberal attitude") of the public agent. For instance, if the CRA's were to expect that in case when the credit ratings overstate true economic situation, their reports would fuel investment that is excessive from the optimal standpoint, then the adoption of conservative rating policy would be welfare-enhancing.

At a broader level, our model applies to any environment where knowledge is dispersed, information aggregation is limited and coordination has some value. It is well-known that traditionally, the price system accomplishes both tasks: it

aggregates information and coordinates behavior.

In many instances, for technical or institutional reasons, prices fail to do their job: either not all relevant information gets incorporated into prices, or if it does, the prices do poor coordinating job; or worse still, these two functions could be incompatible with each other. Whatever the underlying reasons might be, every now and then some "talking" is needed to supplement the prices.

Toulouse School of Economics

REFERENCES

Angeletos, GeorgeMarios and Alessandro Pavan, "Transparency of Information and Coordination in Economies with Investment Complementarities," *American Economic Review, Papers and Proceedings*, May 2004, *94* (2), 91–98.

_ and _ , "Efficient Use of Information and Social Value of Information," *Econometrica*, July 2007, *75* (4), 1103–1142.

_ , Christian Hellwig, and Alessandro Pavan, "Signaling in a Global Game: Coordination and Policy Traps," *Journal of Political Economy*, June 2006, *114* (3), 452–484.

Austen-Smith, David and Jeffrey S. Banks, "Cheap Talk and Burned Money," *Journal of Economic Theory*, March 2000, *91* (1), 1–16.

Benmelech, Efraim and Jennifer Dlugosz, "The credit rating crisis," in "NBER Macroeconomics Annual 2009, Volume 24," University of Chicago Press, 2010, pp. 161–207.

Bolton, Patrick and Mathias Dewatripont, "Authority in organizations: a survey," *Handbook of Organizational Economics*, 2011.

_ , **Markus K. Brunnermeier, and Laura Veldkamp**, "Leadership, Coordination, and Corporate Culture," *Review of Economic Studies*, 2013, *80* (1), 512–537.

_ , **Xavier Freixas, and Joel Shapiro**, "The Credit Ratings Game," *Journal of Finance*, February 2012, *67* (1), 85–112.

Boot, Arnoud W. A., Todd T. Milbourn, and Anjolein Schmeits, "Credit Ratings as Coordination Mechanisms," *Review of Financial Studies*, Spring 2006, *19* (1), 81–118.

Calomiris, Charles, "A recipe for ratings reform," *The Economists' Voice*, 2009, *6* (11).

Campbell, Jeffrey R, Charles L Evans, Jonas DM Fisher, Alejandro Justiniano, Charles W Calomiris, and Michael Woodford, "Macroeconomic Effects of Federal Reserve Forward Guidance [with Comments and Discussion]," *Brookings Papers on Economic Activity*, 2012, pp. 1–80.

Che, Yeon-Koo and Navin Kartik, "Opinions as Incentives," *Journal of Political Economy*, October 2009, *117* (5), 815–860.

Chen, Zhihua, Aziz Lookman, N Schürhoff, and Duane Seppi, "Ratings Matter: Evidence from the Lehman Brothers' Index Rating Redefinition," 2012. working paper.

Compte, Olivier and Andrew Postlewaite, "Confidence-enhanced performance," *The American Economic Review*, 2004, *94* (5), 1536–1557.

Cornaggia, Jess and Kimberly J Cornaggia, "Estimating the costs of issuer-paid credit ratings," *Review of Financial Studies*, 2013, *26* (9), 2229–2269.

Cornaggia, Kimberly and Jess Cornaggia, "Does the bond market want informative credit ratings?," 2012. working paper.

Crawford, Vincent P. and Joel Sobel, "Strategic Information Transmission," *Econometrica*, November 1982, *50* (6), 1431–1451.

Darbellay, Aline, *Regulating Credit Rating Agencies* Elgar Financial Law, Edward Elgar Pub, 2013.

Dewan, Torun and David P Myatt, "The qualities of leadership: Direction, communication, and obfuscation," *American Political Science Review*, 2008, *102* (03), 351–368.

Dewatripont, Mathias and Jean Tirole, "Advocates," *Journal of Political Economy*, February 1999, *107* (1), 1–39.

Diomande, M Ahmed, James S Heintz, and Robert N Pollin, "Why US Financial Markets Need a Public Credit Rating Agency," *The Economists' Voice*, 2009, *6* (6).

Eliaz, Kfir and Françoise Forges, "Disclosing Information to Interacting Agents," December 2012. working paper.

Fahr, Stephan, Roberto Motto, Massimo Rostagno, Frank Smets, and Oreste Tristani, "A monetary policy strategy in good and bad times: Lessons from the recent past," *Economic Policy*, 2013, *28* (74), 243–288.

Farrell, Joseph, "Talk is Cheap," *American Economic Review, Papers and Proceedings*, May 1995, *85* (2), 186–190.

_ **and Matthew Rabin**, "Cheap Talk," *Journal of Economic Perspectives*, Summer 1996, *10* (3), 103–118.

_ **and Robert Gibbons**, "Cheap Talk with Two Audiences," *American Economic Review*, December 1989, *79* (5), 1214–1223.

Ferreira, Daniel and Marcelo Rezende, "Corporate strategy and information disclosure," *The RAND journal of economics*, 2007, *38* (1), 164–184.

Goltsman, Maria and Gregory Pavlov, "How to Talk to Multiple Audiences," *Games and Economic Behavior*, May 2011, *72* (1), 100–122.

Goodhart, Charles, "Debating the Merits of Forward Guidance," *Forward Guidance*, 2013, p. 151.

He, Zhiguo and Wei Xiong, "Dynamic Debt Runs," *Review of Financial Studies*, November 2012, *25* (6), 1799–1843.

Hermalin, Benjamin E, "Toward an economic theory of leadership: Leading by example," *American Economic Review*, 1998, pp. 1188–1206.

Jiang, John Xuefeng, Mary Harris Stanford, and Yuan Xie, "Does it matter who pays for bond ratings? Historical evidence," *Journal of Financial Economics*, 2012, *105* (3), 607–621.

Kamenica, Emir and Matthew Gentzkow, "Bayesian Persuasion," *American Economic Review*, October 2011, *101*, 2590–2615.

Kartik, Navin, "A Note on Cheap Talk and Burned Money," *Journal of Economic Theory*, September 2007, *136* (1), 749–758.

_ , "Strategic Communication with Lying Costs," *Review of Economic Studies*, October 2009, *76* (4), 1359–1395.

_ **, Marco Ottaviani, and Francesco Squintani**, "Credulity, Lies, and Costly Talk," *Journal of Economic Theory*, May 2007, *134* (1), 93–116.

Kisgen, Darren J and Philip E Strahan, "Do regulations based on credit ratings affect a firm's cost of capital?," *Review of Financial Studies*, 2010, *23* (12), 4324–4347.

Kliger, Doron and Oded Sarig, "The information value of bond ratings," *The Journal of Finance*, 2000, *55* (6), 2879–2902.

Landier, Augustin, David Sraer, and David Thesmar, "Optimal Dissent in Organizations," *Review of Economic Studies*, April 2009, *76* (2), 761–794.

Lynch, Timothy E, "Deeply and Persistently Conflicted: Credit Rating Agencies in the Current Regulatory Environment," *Indiana Legal Studies Research Paper*, 2010, (133).

Majumdar, Sumon and Sharun W Mukand, "The Leader as Catalyst," 2007. working paper.

Manso, Gustavo, "Feedback Effects of Credit Ratings," *Journal of Financial Economics*, August 2013, *109* (2), 535–548.

Medvedev, Andrei and Damien Fennell, "An economic analysis of credit rating agency business models and ratings accuracy," *Financial Services Authority Occasional Paper*, 2011, *41*.

Milgrom, Paul R., "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, Autumn 1981, *12* (2), 380–391.

Morris, Stephen and Hyun Song Shin, "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks," *American Economic Review*, June 1998, *88* (3), 587–597.

_ **and** _ , "Social Value of Public Information," *American Economic Review*, December 2002, *92* (5), 1521–1534.

Neuenkirch, Matthias, "Managing financial market expectations: the role of central bank transparency and central bank communication," *European Journal of Political Economy*, 2012, *28* (1), 1–13.

_ , "Central bank transparency and financial market expectations: The case of emerging markets," *Economic Systems*, 2013, *37* (4), 598–609.

Opp, Christian C, Marcus M Opp, and Milton Harris, "Rating agencies in the face of regulation," *Journal of Financial Economics*, 2013, *108* (1), 46–61.

Praet, Peter, "Forward guidance and the ECB," *Forward guidance: Perspectives from central bankers, scholars and market participants, a voxEU. org eBook, CEPR*, 2013.

Prendergast, Canice, "The Motivation and Bias of Bureaucrats," *American Economic Review*, March 2007, *97* (1), 180–196.

Romer, Christina D, "Dear Ben: It's Time for Your Volcker Moment," *The New York Times*, 2011, *29*.

Seidmann, Daniel J. and Eyal Winter, "Strategic Information Transmission with Verifiable Messages," *Econometrica*, January 1997, *65* (1), 163–169.

Skreta, Vasiliki and Laura Veldkamp, "Ratings Shopping and Asset Complexity: A Theory of Ratings Inflation," *Journal of Monetary Economics*, July 2009, *56* (5), 678–695.

Sobel, Joel, *Advances in Economics and Econometrics: Tenth World Congress*, Vol. 1, Cambridge University Press, May

White, Lawrence J., "Markets: The Credit Rating Agencies," *Journal of Economic Perspectives*, Spring 2010, *24* (2), 211–226.

Williams, John C, "Forward policy guidance at the Federal Reserve," *Forward Guidance*, 2013, p. 43.

Woodford, Michael, "Forward Guidance by Inflation-Targeting Central Banks," 2013.

Appendix

For Online Publication

Proof of Proposition 1

Given sender's information is y and for an arbitrary threshold x, by 13 the receivers' welfare equals:

$$\begin{aligned}\Pi(x,y) &= \int_\Theta \Big\{ F(x|\theta)\big[1+\rho(1-F(x|\theta))\big] \\ &\quad + (1-F(x|\theta))\big[\theta+\rho(1-F(x|\theta))\big]\Big\}\Psi(\theta|y) \\ &= \rho + \frac{\alpha y+\gamma\theta_0}{\alpha+\gamma} + \widetilde{F}(x|y)\left(1-\rho-\frac{\alpha y+\gamma\theta_0}{\alpha+\gamma}\right) + \frac{1}{\alpha+\gamma}\widetilde{f}(x|y),\end{aligned} \tag{1}$$

where $\widetilde{f}(\cdot|y)$ and $\widetilde{F}(\cdot|y)$ are, correspondingly, the pdf and the cdf of the normal distribution with mean $\frac{\alpha y+\gamma\theta_0}{\alpha+\gamma}$ and variance $(\alpha+\gamma)^{-1}+\beta^{-1}$.

Taking derivative w.r.t. x,

$$\Pi_x = \widetilde{f}(x|y)\left(1-\rho-\frac{\alpha y+\beta x+\gamma\theta_0}{\alpha+\beta+\gamma}\right). \tag{2}$$

Define $x^*(y)$ such that $\frac{\alpha y+\beta x+\gamma\theta_0}{\alpha+\beta+\gamma} = 1-\rho$. It is clear that $\Pi_x > 0$ for $x < x^*$ and $\Pi_x < 0$ for $x > x^*$. Hence, $\Pi(x,y)$ is single-peaked in x, with the maximum achieved at $x = x^*$.

This completes the proof.

Proof of Theorem 1

We prove the theorem for $\rho > 0$ (the $\rho < 0$ case can be treated analogously). First, let us introduce the lemma characterizing existence of the auxiliary threshold $\widetilde{x}(y)$ and its properties.

LEMMA 2: *There exists $\widetilde{x}(y) < x^*(y)$, such that $\Pi(\widetilde{x}(y),y) = \Pi(\hat{x}(y),y)$, if and*

only if

$$y < \overline{y} \equiv \frac{1}{\alpha}\left[(\alpha+\gamma)\left(1-\frac{\hat{z}(\tau_B\rho)}{\tau_B}\right)-\gamma\theta_0\right], \tag{3}$$

with $\tau_B \equiv \sqrt{\frac{(\alpha+\beta+\gamma)(\alpha+\gamma)}{\beta}}$ *and* $\hat{z}$ *solving:*

$$\frac{\phi(z)}{\Phi(z)} = -z + \tau_B\rho, \tag{4}$$

where Φ *and* ϕ *stand for cdf and pdf of the standard normal distribution.*

Moreover, $\tilde{x}(y)$ *is monotonically decreasing in* y *and*

$$\lim_{y\to-\infty}\tilde{x}(y) \to \hat{x}(y), \quad \lim_{y\to\overline{y}}\tilde{x}(y) \to -\infty. \tag{5}$$

PROOF:

Fix an arbitrary $y \in \mathcal{Y}$. Recall that $\Pi(x,y)$ is single-peaked in x and when $\rho > 0$, $\hat{x}(y) > x^*(y)$. Therefore, $\widetilde{x}(y)$ exists if and only if

$$\lim_{x\to-\infty}\Pi(x,y) < \Pi(\hat{x}(y),y). \tag{6}$$

By 1, this implies

$$\widetilde{F}(\hat{x}(y)|y)\left(1-\rho-\frac{\alpha y+\gamma\theta_0}{\alpha+\gamma}\right)+\frac{1}{\alpha+\gamma}\widetilde{f}(\hat{x}(y)|y) > 0. \tag{7}$$

Substituting for $\widetilde{F}$ and $\widetilde{f}$, we get:

$$\begin{aligned}&\Phi\left(\frac{\hat{x}(y)-\frac{\alpha y+\gamma\theta_0}{\alpha+\gamma}}{\sqrt{(\alpha+\gamma)^{-1}+\beta^{-1}}}\right)\left(1-\rho-\frac{\alpha y+\gamma\theta_0}{\alpha+\gamma}\right)\\ &+\frac{1}{\alpha+\gamma}\frac{1}{\sqrt{(\alpha+\gamma)^{-1}+\beta^{-1}}}\phi\left(\frac{\hat{x}(y)-\frac{\alpha y+\gamma\theta_0}{\alpha+\gamma}}{\sqrt{(\alpha+\gamma)^{-1}+\beta^{-1}}}\right) > 0.\end{aligned} \tag{8}$$

Plugging $\hat{x}(y) = \frac{1}{\beta}(\alpha + \beta + \gamma - \alpha y - \gamma\theta_0)$ and defining

$$\tau_B \equiv \sqrt{\frac{(\alpha+\beta+\gamma)(\alpha+\gamma)}{\beta}} \quad \text{and} \quad z \equiv \tau_B \left(1 - \frac{\alpha y + \gamma\theta_0}{\alpha+\gamma}\right),$$

the above condition boils down to

$$\Phi(z)(z - \tau_B\rho) + \phi(z) > 0, \tag{9}$$

or

$$\frac{\phi(z)}{\Phi(z)} > -z + \tau_B\rho. \tag{10}$$

Notice that the left-hand side of the inequality, $\frac{\phi(z)}{\Phi(z)}$, is the mirror-image (with z replaced by $-z$) of the hazard rate of the standard normal distribution. We know that it asymptotically converges to $-z$ as $z \to -\infty$ and converges to 0 as $z \to +\infty$.

Hence, for any $\rho > 0$, there is a unique cutoff $\hat{z}$, such that 10 holds if and only if $z > \hat{z}$. This means that $\widetilde{x}(y)$ exists if and only if y is smaller than the cutoff $\overline{y}$ given by 3.

For any $y \in (-\infty, \overline{y})$ we have $\widetilde{x}(y) < \hat{x}(y)$ implicitly defined by $\Pi(\widetilde{x}(y), y) = \Pi(\hat{x}(y), y)$. Call $\Delta x \equiv \hat{x}(y) - \tilde{x}(y) > 0$. By 1, this implies

$$\begin{aligned} &\widetilde{F}(\widetilde{x}(y)|y)\left(1 - \rho - \frac{\alpha y + \gamma\theta_0}{\alpha+\gamma}\right) + \frac{1}{\alpha+\gamma}\widetilde{f}(\widetilde{x}(y)|y) \\ =&\widetilde{F}(\hat{x}(y)|y)\left(1 - \rho - \frac{\alpha y + \gamma\theta_0}{\alpha+\gamma}\right) + \frac{1}{\alpha+\gamma}\widetilde{f}(\hat{x}(y)|y), \end{aligned} \tag{11}$$

or

(12)
$$\tau_B\left(\frac{\alpha y+\gamma\theta_0}{\alpha+\gamma}+\rho-1\right)=\frac{\phi\left(\tau_B\left(1-\frac{\alpha y+\gamma\theta_0}{\alpha+\gamma}\right)\right)-\phi\left(\tau_B\left(1-\frac{\alpha y+\gamma\theta_0}{\alpha+\gamma}\right)-\frac{\Delta x}{\sigma}\right)}{\Phi\left(\tau_B\left(1-\frac{\alpha y+\gamma\theta_0}{\alpha+\gamma}\right)\right)-\Phi\left(\tau_B\left(1-\frac{\alpha y+\gamma\theta_0}{\alpha+\gamma}\right)-\frac{\Delta x}{\sigma}\right)},$$

which boils down to:

$$\tau_B\rho-z=\frac{\phi(z)-\phi(z-\Delta z)}{\Phi(z)-\Phi(z-\Delta z)}, \tag{13}$$

or

$$\tau_B\rho-z-\frac{\phi(z)}{\Phi(z)}=\frac{\Phi(z-\Delta z)}{\Phi(z)-\Phi(z-\Delta z)}\left(\frac{\phi(z)}{\Phi(z)}-\frac{\phi(z-\Delta z)}{\Phi(z-\Delta z)}\right) \tag{14}$$

where $z\equiv\tau_B\left(1-\frac{\alpha y+\gamma\theta_0}{\alpha+\gamma}\right)$, $\Delta z\equiv\frac{\Delta x}{\sigma}$ and $\sigma\equiv\sqrt{(\alpha+\gamma)^{-1}+\beta^{-1}}$.

The left-hand side of 14 is monotonically decreasing in z and equal to 0 at $z=\hat{z}$. When $z\to+\infty$, the left-hand side of 14 converges to $-\infty$. Therefore, on the right-hand side the denominator must vanish:

$$\lim_{z\to+\infty}\Delta z\to 0.$$

This means that

$$\lim_{y\to-\infty}\tilde{x}(y)\to\hat{x}(y).$$

Similarly, when $z\to\hat{z}$, the right-hand side converges to 0: $\lim_{z\to\hat{z}}\Delta z\to+\infty$, which in turn implies $\lim_{y\to\overline{y}}\tilde{x}(y)\to-\infty$.

Notice that the left-hand side of 13 is decreasing in z. Since $\frac{\phi(\cdot)}{\Phi(\cdot)}$ is a decreasing function, the right-hand side is decreasing in Δz.

Next, notice that $\frac{\phi(\cdot)}{\Phi(\cdot)}$ is convex: for a fixed Δz, $\frac{\phi(z)}{\Phi(z)}$ and $\frac{\phi(z-\Delta z)}{\Phi(z-\Delta z)}$ become closer as z increases. Therefore, the right-hand side is increasing in z. As a result, $\frac{\partial\Delta z}{\partial z}<0$, and so $\frac{\partial\Delta x(y)}{\partial y}>0$. Hence, $\frac{\partial\tilde{x}(y)}{\partial y}=\frac{\partial\hat{x}(y)}{\partial y}-\frac{\partial\Delta x(y)}{\partial y}<0$.

By Lemma 2, fix an arbitrary $\hat{x}(\varnothing)$ and the non-disclosure region is given by an interval since $\tilde{x}(y)$ and $\hat{x}(y)$ are monotonically decreasing. So the triplet $\{\hat{x}(\varnothing), y_1, y_2\}$ solving 22-24 gives an equilibrium. By construction, it satisfies the four equilibrium conditions outlined in Definition 1. The rest of proof shows existence.

Given an arbitrary x, it pins down a unique pair $(y_1(x), y_2(x))$ by 23 and 24. Since 22 has always a unique solution for any pair of (y_1, y_2), there exists a function $\Lambda(x) = x'$ satisfying,

$$1 = \mathbb{E}\left[R(\theta, A)|x', y \in \{\varnothing\} \cup [y_1(x), y_2(x)]\right] \tag{15}$$

An equilibrium exists if and only if there exists a fixed point in function $\Lambda(\cdot)$.

As $x \to +\infty$, by Lemma 2 we have $y_1(x) \to y_2(x)$. So

$$\mathbb{E}\left[R(\theta, A)|x', y \in \{\varnothing\} \cup [y_1(x), y_2(x)]\right] \simeq \mathbb{E}\left[R(\theta, A)|x', y = \varnothing\right]$$

This means that $\Lambda(+\infty) = \hat{x}_0 < +\infty$.

As $x \to -\infty$, by Lemma 2,

$$\mathbb{E}\left[R(\theta, A)|x', y \in \{\varnothing\} \cup [y_1(x), y_2(x)]\right] \simeq \mathbb{E}\left[R(\theta, A)|x', y \in \{\varnothing\} \cup [\hat{y}, +\infty)\right]$$

Clearly we have $\Lambda(-\infty) > -\infty$.

Combining above two conditions, by continuity Λ crosses 45-degree line at least once. This completes the proof.

Proof of Proposition 2

Consider the function

$$\xi(x, y; r) \triangleq \rho \frac{\alpha y + \beta x + \gamma \theta_0}{\alpha + \beta + \gamma} + \rho \int_{-\infty}^{+\infty} [1 - F(x|\theta)]\, d\Psi(\theta|x, y), \tag{16}$$

which denotes, for a given y, the receiver's expected payoff from playing the threshold strategy around x, given all other receivers also play around x.

The threshold $\hat{x}(y)$ is implicitly defined by

$$\xi\left(\hat{x}(y), y; \rho\right) = 1. \tag{17}$$

We can write $\xi(\cdot; \rho)$ as

$$\xi\left(x, y; \rho\right) = \frac{\alpha y + \beta x + \gamma\theta_0}{\alpha + \beta + \gamma} + \rho\left[1 - \widetilde{F}(x|x, y)\right], \tag{18}$$

where $\widetilde{F}(\cdot|x, y)$ is the cumulative density of a normal distribution with mean $\frac{\alpha y+\beta x+\gamma\theta_0}{\alpha+\beta+\gamma}$ and variance $(\alpha + \beta + \gamma)^{-1} + \beta^{-1}$.

Differentiating with respect to x, we get

$$\begin{aligned} \frac{\partial}{\partial x}\xi\left(x, y; \rho\right) &= \frac{\beta}{\alpha + \beta + \gamma} - \rho\widetilde{f}(x|x, y)\left[1 - \frac{\beta}{\alpha + \beta + \gamma}\right] \\ &\geq \frac{\beta}{\alpha + \beta + \gamma} - \rho\frac{\alpha + \gamma}{\alpha + \beta + \gamma}\sqrt{\frac{\beta(\alpha + \beta + \gamma)}{2\pi(\alpha + 2\beta + \gamma)}} > 0. \end{aligned} \tag{19}$$

Next, since the second term in $\xi(\cdot; r)$, $[1 - \widetilde{F}(x|x, y)]$, is bounded, while the first term is linear in x with the positive slope,

$$\lim_{x\to-\infty}\xi\left(x, y; r\right) \to -\infty \quad \text{and} \quad \lim_{x\to+\infty}\xi\left(x, y; r\right) \to +\infty. \tag{20}$$

Therefore, $\xi(\cdot; r)$ single-crosses 1, and hence for any $y \in \mathcal{Y}$, there exists a unique solution $\hat{x}(y)$. For monotonicity of $\hat{x}(y)$, observe that from the Implicit Function Theorem,

$$\frac{d\hat{x}(y)}{dy} = -\frac{\xi_y}{\xi_x} < 0, \tag{21}$$

since $\xi_x > 0$, whereas

$$\xi_y \equiv \frac{\partial}{\partial y}\xi\left(x, y; \rho\right) = \frac{\alpha}{\alpha + \beta + \gamma}\left[\rho\widetilde{f}(x|x, y)\right] > 0.$$

Hence, $\frac{\partial \hat{x}(y)}{\partial y} < 0$.

Next, we show that the wishful threshold $x^*(y)$ is also unique and decreasing in y. Plug the payoff function from 27 into $\Pi(x, y)$ and differentiate with respect to x:

(22)
$$\begin{aligned}\frac{\partial}{\partial x}\Pi(x, y) &= \int_{-\infty}^{+\infty}\left\{f(x|\theta)\Big[1 - \theta - 2\rho(1 - F(x|\theta))\Big]\right\}d\Psi(\theta|y)\\ &= \frac{\int_{-\infty}^{+\infty} f(x|\theta)h(y|\theta)d\Psi(\theta)}{\int_{-\infty}^{+\infty} h(y|\theta)d\Psi(\theta)}\int_{-\infty}^{+\infty}\Big[1 - \theta - 2\rho(1 - F(\theta|x))\Big]d\Psi(\theta|x, y)\\ &\propto 1 - \xi(x, y; 2\rho).\end{aligned}$$

So, there exists a unique $x^*(y)$ such that

$$\xi(x^*(y), y; 2\rho) = 1. \tag{23}$$

Furthermore, note that $\frac{\partial}{\partial x}\Pi(x, y) > 0$ for all $x < x^*(y)$ and $\frac{\partial}{\partial x}\Pi(x, y) < 0$ for all $x > x^*(y)$. Therefore, for any $y \in \mathcal{Y}$, the function $\Pi(\cdot, y)$ is single-peaked and maximized at $x^*(y)$.

For the properties of the conflict of interests, $\Delta(y)$, parametrize the equilibrium threshold $\hat{x}(y)$ by ρ and take the derivative of $\hat{x}(y; \rho)$ with respect to ρ using the fact that $\hat{x}(\cdot)$ is implicitly defined by 17 via the $\xi(\cdot; \rho)$ function:

$$\frac{\partial}{\partial \rho}\hat{x}(y; \rho) = -\frac{\xi_\rho}{\xi_x} = -\frac{1 - \overbrace{\int_{-\infty}^{+\infty} F(x|\theta)d\Psi(\theta|x, y)}^{\widetilde{F}(x|x,y)}}{\xi_x} < 0, \tag{24}$$

since $\xi_x > 0$, whereas $\widetilde{F}(x|x,y) < 1$. So, when $\rho > 0$,

$$x^*(y) \equiv \hat{x}(y; 2\rho) < \hat{x}(y;\rho) \equiv \hat{x}(y).$$

Symmetrically, $x^*(y) > \hat{x}(y)$ whenever $\rho < 0$.

In order to establish the limiting behavior of $\Delta(\cdot)$ as $y \to -\infty$ or $y \to +\infty$, recall that for all $y \in \mathcal{Y}$, $\hat{x}(y)$ and $x^*(y)$ are characterized by:

$$\frac{\alpha y + \beta \hat{x} + \gamma\theta_0}{\alpha+\beta+\gamma} + \rho\Big[1 - \widetilde{F}(\hat{x}|\hat{x}, y)\Big] = 1, \tag{25}$$

$$\frac{\alpha y + \beta x^* + \gamma\theta_0}{\alpha+\beta+\gamma} + 2\rho\Big[1 - \widetilde{F}(x^*|x^*, y)\Big] = 1. \tag{26}$$

Subtracting 25 from 26:

$$\Delta(y) = \rho \cdot \frac{\alpha+\beta+\gamma}{\beta} \cdot \Big[2\widetilde{F}(x^*(y)|x^*(y), y) - \big(1 + \widetilde{F}(\hat{x}(y)|\hat{x}(y), y)\big)\Big]. \tag{27}$$

Consider the limiting behavior of $\widetilde{F}(x|x,y)$. The change in y has a triple effect on $\widetilde{F}$. First, since

$$x|x,y \sim \mathcal{N}\left(\frac{\alpha y + \beta x + \gamma\theta_0}{\alpha+\beta+\gamma}, \frac{1}{\alpha+\beta+\gamma} + \frac{1}{\beta}\right),$$

for fixed x we necessarily have

$$\lim_{y\to-\infty} \widetilde{F}(x|x,y) = 1 \quad \text{and} \quad \lim_{y\to+\infty} \widetilde{F}(x|x,y) = 0. \tag{28}$$

Next, consider the total derivative of $\widetilde{F}(\cdot)$ with respect to x:

$$\frac{d}{dx}\widetilde{F}(x|x,y) = \widetilde{f}(x|x,y)\left[1 - \frac{\beta}{\alpha+\beta+\gamma}\right] > 0.$$

Finally, coupled with the fact that $\hat{x}(y)$ and $x^*(y)$ are decreasing in y:

$$\lim_{y\to-\infty}\widetilde{F}(\hat{x}(y)|\hat{x}(y),y)=\lim_{y\to-\infty}\widetilde{F}(x^*(y)|x^*(y),y)=1$$

and

$$\lim_{y\to+\infty}\widetilde{F}(\hat{x}(y)|\hat{x}(y),y)=\lim_{y\to+\infty}\widetilde{F}(x^*(y)|x^*(y),y)=0.$$

Plugging this into 27, we get

$$\lim_{y\to-\infty}\Delta(y)=0\quad\text{and}\quad\lim_{y\to+\infty}\Delta(y)=-\rho\cdot\frac{\alpha+\beta+\gamma}{\beta}.$$

This completes the proof.

Proof of Proposition 3

For all $y\in\mathcal{Y}$, the regime change game is solved by a pair $(\hat{x}(y),\hat{\theta}(y))$ satisfying:

$$\begin{aligned}1&=\delta F(\hat{\theta}|\hat{x},y)+\bar{R}(1-F(\hat{\theta}|\hat{x},y)),\\1-\hat{\theta}&=1-F(\hat{x}|\hat{\theta}).\end{aligned}\tag{29}$$

The receivers' aggregate payoff conditional on y (which is also the sender's objective) for an arbitrary threshold x is

$$\begin{aligned}\Pi(x,y)=&\int_\Theta 1\cdot F(x|\theta)d\Psi(\theta|y)+\int_{\theta<\widetilde{\theta}(x)}\delta(1-F(x|\theta))d\Psi(\theta|y)\\&+\int_{\theta>\widetilde{\theta}(x)}\bar{R}(1-F(x|\theta))d\Psi(\theta|y),\end{aligned}\tag{30}$$

where $\widetilde{\theta}$ indicates the threshold above which regime will be abandoned, given the attack threshold x. $\widetilde{\theta}(x)$ is uniquely solved by $\theta=F(x|\theta)$, and it is straightforward to check that $\partial\widetilde{\theta}/\partial x>0$: more aggressive attack leads to more frequent regime abandonment.

Taking derivative of 30 with respect to x,

$$\begin{aligned}\frac{\partial \Pi(x,y)}{\partial x} &= \int_{\Theta} 1 \cdot f(x|\theta) d\Psi(\theta|y) - \int_{\theta<\widetilde{\theta}(x)} \delta f(x|\theta) d\Psi(\theta|y) \\ &- \int_{\theta>\widetilde{\theta}(x)} \bar{R} f(x|\theta) d\Psi(\theta|y) + \frac{\partial \widetilde{\theta}(x)}{\partial x}(\delta - \bar{R})(1 - F(x|\widetilde{\theta}))\psi(\widetilde{\theta}|y) \\ &= \widetilde{f}(x|y)\left[1 - \delta \Pr\{\theta < \widetilde{\theta}|x,y\} - \bar{R}\Pr\{\theta > \widetilde{\theta}|x,y\}\right] \\ &+ \widetilde{\theta}_x(\delta - \bar{R})(1-\widetilde{\theta})\psi(\widetilde{\theta}|y)\end{aligned} \tag{31}$$

Plugging $x = \hat{x}$ and $\widetilde{\theta} = \hat{\theta}$ from 29, we obtain the first part of 30 being equal to 0, while the second part being negative. As a result, $x^*(y) > \hat{x}(y)$, and hence $\Delta(y) < 0$ for all $y \in \mathcal{Y}$. Also note that as $y \to +\infty$ or $y \to \infty$, $\psi(\hat{\theta}|y) \to 0$, since $\hat{\theta}$ must lie in $(0,1)$.

Therefore, $\Pi_x(\hat{x}, y) \to 0$ and $\Delta(y) \to 0$.

Proof of Proposition 4

With $r(A) = 1 + \rho A$ and $R(\theta, A) = \theta + \rho A$, rearranging 31:

$$\begin{aligned}\Pi(x,y,b,\rho) &= \int_{-\infty}^{+\infty} \Big\{ F(x|\theta)[1 + \rho(1 - F(x|\theta))] \\ &+ (1 - F(x|\theta))[\theta + \rho(1 - F(x|\theta)) + b] \Big\} d\Psi(\theta|y) \\ &= \int_{-\infty}^{+\infty} \Big\{ F(x|\theta)[1 + (\rho + b)(1 - F(x|\theta))] \\ &+ (1 - F(x|\theta))[\theta + (\rho + b)(1 - F(x|\theta))] \Big\} d\Psi(\theta|y) \\ &= \Pi(x,y,0,\rho + b)\end{aligned}$$

This means that the objective of the sender with bias b, given externality parameter ρ, is identical to that of a benevolent sender with the externality parameter $\rho + b$. Hence, the two equilibria should coincide.

Proof of Lemma 1

Let us rewrite the three conditions characterizing equilibrium in Theorem 1, explicitly taking into account the dependence of $\{\hat{x}(\varnothing), y_1, y_2\}$ on b:

$$1 = \mathbb{E}\left[R(\theta, A) | x = \hat{x}(\varnothing, b), y \in \{\varnothing\} \cup [y_1(b), y_2(b)]\right], \tag{32}$$

$$\hat{x}(\varnothing, b) = \hat{x}(y_2(b)), \tag{33}$$

$$\Pi(\hat{x}(\varnothing, b), y_1(b)) = \Pi(\hat{x}(y_1(b), b), y_1(b)). \tag{34}$$

Recall that the auxiliary threshold $\widetilde{x}(y, b)$ is implicitly defined by $\Pi(\widetilde{x}, y, b) = \Pi(\hat{x}(y), y, b)$, with $\widetilde{x} \neq \hat{x}(y)$. Then 34 can be rewritten as $\hat{x}(\varnothing, b) = \widetilde{x}(y, b)$.

Consider a marginal decrease from b to $b - \varepsilon$ in the $\rho > 0$ case. We have $\widetilde{x}(y, b-\varepsilon) > \widetilde{x}(y, b)$ for any $y \in \mathcal{Y}$. Hence, $\hat{x}(\varnothing, b) = \widetilde{x}(y_1', b-\varepsilon)$ makes $y_1' > y_1(b)$. This means that the posterior $\Psi(\theta | y \in \{\varnothing\} \cup [y_1', y_2(b)])$ first-order stochastically dominates $\Psi(\theta | y \in \{\varnothing\} \cup [y_1(b), y_2(b)])$.

As a result, for 32 to hold, we need a marginal decrease from $\hat{x}(\varnothing, b)$ to x'. Note that $\widetilde{x}(y, b)$ and $\hat{x}(y)$ are decreasing in y. In the new iteration, $x' = \hat{x}(y_2'') = \widetilde{x}(y_1'', b - \varepsilon)$ makes $y_1'' > y_1'$ and $y_2'' > y_2$, resulting in a further decrease to x''. As the iteration continues, the $\{x'\}$ sequence converges to $\hat{x}(\varnothing, b - \varepsilon) < \hat{x}(\varnothing, b)$ so $\frac{\partial \hat{x}(\varnothing, b)}{\partial b} > 0$.

A similar exercise can be done for the case of $\rho < 0$.

Proof of Proposition 5

From 15, receivers' equilibrium welfare when the sender's bias equals b is

$$\begin{aligned} V(\mathcal{Y}^{\mathrm{N}}(b)) = \int_{\Theta} \Bigg\{ & \left(1 - p + p \int_{y \in \mathcal{Y}^{\mathrm{N}}(b)} dH(y|\theta)\right) \widetilde{\Pi}(\hat{x}(\varnothing, b), \theta) \\ & + p \int_{y \notin \mathcal{Y}^{\mathrm{N}}(b)} \widetilde{\Pi}(\hat{x}(y), \theta) dH(y|\theta) \Bigg\} d\Psi(\theta). \end{aligned} \tag{35}$$

Differentiating with respect to b,

$$\begin{aligned}\frac{\partial V}{\partial b} = \int_{\Theta} \Bigg\{ & \left(1 - p + p \int_{y \in \mathcal{Y}^{\mathrm{N}}(b)} dH(y|\theta)\right) \widetilde{\Pi}_x(\hat{x}(\varnothing, b), \theta) \frac{\partial \hat{x}(\varnothing, b)}{\partial b} \\ & + p\widetilde{\Pi}(\hat{x}(\varnothing, b), \theta) \left(h(y_2(b)|\theta) \frac{\partial y_2(b)}{\partial b} - h(y_1(b)|\theta) \frac{\partial y_1(b)}{\partial b}\right) \\ & - p\widetilde{\Pi}(\hat{x}(y_2(b), \theta) h(y_2(b)|\theta) \frac{\partial y_2(b)}{\partial b} + p\widetilde{\Pi}(\hat{x}(y_1(b), \theta) h(y_1(b)|\theta) \frac{\partial y_1(b)}{\partial b} \Bigg\} d\Psi(\theta).\end{aligned}$$

According to Theorem 1, at $b = 0$:

$$\begin{aligned}\left.\frac{\partial V}{\partial b}\right|_{b=0} &= \hat{x}_b(\varnothing, 0) \int_{\Theta} \left\{ \left(1 - p + p \int_{y \in \mathcal{Y}^{\mathrm{N}}} dH(y|\theta)\right) f(x|\theta)(1 - \rho - \theta) \right\} d\Psi(\theta) \\ &\propto \hat{x}_b(\varnothing, 0) \left(1 - \rho - \mathbb{E}[\theta | \hat{x}(\varnothing, 0), y \in \{\varnothing\} \cup [y_1, y_2]]\right).\end{aligned}$$

By the definition of $\hat{x}(\varnothing, 0)$, $\mathbb{E}[\theta|\hat{x}(\varnothing, 0), y \in \{\varnothing\} \cup [y_1, y_2]] = 1$ and by Lemma 1, $\hat{x}_b(\varnothing, b) > 0$ implies $\frac{\partial V}{\partial b} < 0$ for $\rho > 0$ and $\frac{\partial V}{\partial b} > 0$ for $\rho < 0$.

This means that the receivers' welfare is locally improved by the sender with the small negative bias if their actions exhibit positive externalities, and by the sender with small positive biases if actions exhibit negative externalities.

Mechanism Design and Information Disclosure*

Tong Su† Takuro Yamashita‡

November 17, 2014

Abstract

We consider the problem of optimal information disclosure in mechanism design where the principal can commit to his disclosure policy (possibly because he has hard information) as well as to his mechanism. We first provide a characterization result for the optimality of the full disclosure policy. Applying this result, in a generalized auction setting we show that the principal (seller) always prefers to disclose all the relevant information to the agents. In a bilateral trade setting where his objective is surplus, under a mild condition on the environment, he does not find optimal to reveal all the information. In a voting application where voters choose between either the status quo or a reform, we show that the principal should reveal all information regarding to the aggregate benefit from the reform but reveal no information about individual benefit for each agent.

1 Introduction

When a party has information relevant to the others, whether this party has an incentive to disclose such information is an important but complicated problem. For example, consider a company selling a experience good to consumers, where the company has superior information than the consumers about some aspects of the good such as its quality. Disclosing such information would be beneficial for the consumers, but of course, the company does not necessarily have such an incentive if the disclosure may negatively affect his profit. As another

*Preliminary draft. Please do not circulate without permission.

†Toulouse School of Economics, tong.su@tse-fr.eu

‡Toulouse School of Economics, takuro.yamashita@tse-fr.eu

example, consider a government and investors in financial markets. The government may have exclusive information about future regulations, better estimates of future growth rates, and so on, which are relevant to future values of certain financial products. Whether the benevolent government should disclose all the relevant information may depend on the nature of the environment.

This question about transparency has been examined for specific games, and the answer naturally varies with which games are considered. For example, Morris and Shin (2002) and Angeletos and Pavan (2007) consider coordination games, and show that even the benevolent outsider may not have an incentive to disclose all the relevant information. In first-price auction, several papers such as Landsberger et al. (2001) and Kaplan and Zamir (2002) show that, when the auctioneer has partial information about bidders' valuations, for example their rankings, she can increase her expected revenue by strategically reveal her information across bidders.

These studies suggest that the incentive for transparency would be affected by the frictions intrinsic to specificity of games. A natural direction then may be to investigate whether the informed party has an incentive to disclose all the relevant information *if the party can design the rule of the game by himself*. For this purpose, we take a mechanism design approach, assuming that this informed party is at the same time the designer of a mechanism.

In this sense, this paper is close to the informed-principal literature. However, as opposed to the standard approach in this literature, we assume that the principal can commit to his disclosure strategy (as well as to a mechanism) at the *ex ante* stage. As we discuss later, this commitment assumption would make sense in situations where the principal can show some "hard evidence" to the agents. Moreover, the commitment assumption makes it possible to distinguish the principal's *ex ante* incentive of transparency from her *interim* or *ex post* incentive of "mimicking other types", while both of them exist in the standard approach in the literature. In this sense, we study the informed party's incentive for information disclosure that is essentially different from his signaling incentive.

To be more specific, we consider the following model. First, the principal ex ante commits to his disclosure policy as well as his mechanism. Then, his information is realized, and he sends a public message to the agents according this disclosure policy. Each agent observes this public message, the mechanism, and his own private information, and he sends a message to the mechanism. An allocation is realized and the game ends.

Under certain conditions on the environment, we first provide a simple result that char-

acterizes the optimality of the full disclosure policy. The characterization result is stated in terms of the convexity of the principal's value function (Theorem 1). In the context of Bayesian persuasion, Kamenica and Gentzkow (2011) show that the convexity or concavity of the informed party's preference is a key property in understanding the optimal disclosure policy. In our setting, because the principal does not only disclose information but also designs a mechanism, their result does not directly apply. Nevertheless, under appropriate conditions, a modified version of their result applies.

Next, motivated by examples discussed in the first paragraph, we study three applications. The first application is an auction problem à la Myerson (1981) where a seller (the principal) has relevant information about the bidders' values, and aims to maximize a weighted sum of revenue and surplus. Applying Theorem 1, we show that full transparency is optimal (Theorem 2). We also show that the optimal mechanism is a standard Myerson's auction (i.e., a second-price auction with a reserve price) after fully disclosing the principal's information.[1]

The second application is a bilateral trade problem à la Myerson and Satterthwaite (1983) where the principal (mediator) has relevant information about the traders' values, and aims to maximize surplus. Applying Theorem 1, under a mild condition on the distribution, we show that full transparency is never optimal (Theorem 3). In particular, the principal is always better off by hiding her information when the realization of her information is in a region where a trade between the agents is efficient with a sufficiently high probability. We also show that the optimal mechanism is a Myerson-Satterthwaite trade mechanism (i.e., a

[1]Some papers found a qualitatively similar result, but in a different environment. The closest to ours would be Skreta (2011), who shows that a second-price auction with a reserve price is optimal after full disclosure. However, her information structure is quite different from ours: the principal knows a noisy signal of each agent's valuation, while each agent knows his own value. Thus, the principal's disclosure to agent i can only be about the noisy signals of the values of $-i$. Given that an optimal mechanism is a dominant-strategy incentive compatible, any disclosure policy would not matter for the agents' incentives, including the full disclosure. In our case, the principal knows some information that any agent does not know, and therefore, a logic to obtain the optimality of full disclosure is quite different.

Another paper that is relevant is about the design of bidders' information precision as in Bergemann and Pesendorfer (2007). They consider situations where the principal can control the precision of the agents' private information, while the principal cannot observe at all the realization of them. In our setting, we consider a different situation where the seller observes his own information. The literature on sequential screening as in Courty and Li (2000) and Eső and Szentes (2007) is also relevant though different, because they assume that the additional information provided by the principal can be priced, which is not allowed in our paper.

mechanism that allows for a trade if and only if the value of trade is sufficiently large) after partially disclosing the principal's information.

The last application is a voting problem where a group of agents vote for two alternatives while the principal wants to maximize agents' total welfare. Azrieli and Kim (2013) has shown that in such an environment, the optimal voting mechanism is a weighted majority rule where the one agent's voting weight is determined by his payoff conditional as if his choice is selected. Applying to this result, what we show is that when the principal has information on agents' payoffs, she should fully disclose the aggregate payoffs among agents but should not disclose any individual payoffs.[2]

The paper is structured as follows. In Section 2, we introduce a model where the principal commits to a mechanism and a disclosure policy. In Section 3, we provide a simple condition that characterizes the optimality of full transparency of the principal, in terms of the convexity of the principal's value function (Theorem 1). In Section 4 and 5, we apply Theorem 1 to auction (Section 4), bilateral trade (Section 5) and voting (Section 6). Section 7 discusses some modeling assumptions of the paper and conclude.

2 Mechanism Design with An Informed Principal

There is a set $I = \{1, \ldots, N\}$ of N agents. A principal assigns an allocation $x \in X$ for the agents. For example, in auction, the principal may be an auctioneer, and an allocation consists of a probability of giving a good to each agent as well as his payment. In bilateral trade, the principal may be a mediator, and an allocation consists of a trade probability between a buyer and a seller as well as the trading price.

Each agent $i \in I$ has a private information, or *type*, $v_i \in V_i$. Let $V = \Pi_{i=1}^N V_i$ and a type profile is denoted by $v = (v_i)_{i=1}^N \in V$. The principal also has private information relevant to this economy, denoted by $\theta \in \Theta$. We assume that both V and Θ are compact and convex subsets of a Euclidean space. At the ex ante stage, the principal and agents share the same

[2]Alonso and Câmara (2014) also study the optimal information revelation problem in a voting application, however they only consider fixed voting rule and assume the principal always prefers one alternative then the other. A more related paper is by Li et al. (2014) who show that either full disclosure or no disclosure will be optimal if the principal, who cares both voters' welfare and her desirable alternative, can optimally design voting rule. Our result on optimal information disclosure is different from theirs since we allow the principal's information to be multi-dimensional.

prior distribution on the variables $(v, \theta) \in V \times \Theta$. We assume that $(v_1 \ldots, v_N, \theta)$ are mutually independently distributed,[3] the prior for v admits a density f_V, and the prior for θ admits a density f_Θ.

The principal's utility function (e.g., revenue or surplus) is denoted by $u_0(x, v, \theta)$, and the utility for each agent i is denoted by $u_i(x, v, \theta)$.

The principal has two tools to achieve her objective. First, the principal can send a public message $m \in \mathcal{M}$ to all agents, which we call her *information disclosure policy*. Formally, we define her information disclosure policy as a joint probability measure $\phi \in \Delta(\Theta \times \mathcal{M})$ (i.e., as a distributional strategy of Milgrom and Weber (1982)), whose marginal over Θ is F_Θ (i.e., for each measurable $A \subseteq \Theta$, $\phi(A, \mathcal{M}) = F_\Theta(A)$). In order for the agents' conditional probability measure over Θ to be well-defined for each $m \in \mathcal{M}$, we additionally require that ϕ admits a pair (μ, ψ) that satisfies the following conditions. Intuitively, μ is the marginal of ϕ over $\mathcal{M}$ and ψ is the conditional measure over Θ given each $m \in \mathcal{M}$, induced by ϕ.

Assumption 1. ϕ is *feasible* if there exists (μ, ψ) such that

(i) $\mu \in \Delta(\mathcal{M})$,

(ii) For each m, $\psi(\cdot|m) \in \Delta(\Theta)$,

(iii) For each measurable $A \subseteq \Theta$, $\psi(A|\cdot)$ is a ($\mathcal{M}$-) measurable mapping, and

(iv) For each measurable $A \subseteq \Theta$ and $B \subseteq \mathcal{M}$, $\int_B \psi(A|m) d\mu(m) = \phi(A \times B)$.

The third condition assures that $\int_{m \in B} \psi(A|m) d\mu$ is well-defined, which is used in the last condition. The last condition corresponds to the Bayesian plausibility of Kamenica and Gentzkow (2011). This can also be interpreted as a martingale property of beliefs or as the law of iterated expectation. Given that the principal's message is publicly observable among the agents, we omit the subscript i for the conditional ψ. We assume that the principal can commit to this disclosure strategy before her observing θ. It may be a reasonable assumption, for example, if each $\psi \in \Delta(\Theta)$ is associated with non-falsifiable "evidence" that the principal can show to the agents. Hence, observing such evidence, the agents would update their belief as ψ. Thus, our model can be interpreted as a situation where any ψ can be justified by some evidence.

[3]Note that, when θ is multidimensional, we allow for dependence within θ.

In addition to her information disclosure policy, the principal also designs a mechanism. In general, a (direct) mechanism is denoted by $g : \mathcal{M} \times \Theta \times V \to X$, where $g_m(\theta, v) \in X$ denotes an allocation if the principal announces m as a public message, reports θ, and the agents report v. Until Section 7, however, we restrict our analysis only to a class of mechanisms that are not contingent on the principal's report θ (hence the outcome is simply $g_m(v)$).

The assumption may be reasonable in some situations. For example, it may be sometimes limited for the principal to directly intervene in the agents' play of a mechanism (e.g., by law or regulation). Another example may be such that agents do not know the principal's preference. In this case, if a mechanism is made contingent on the principal's message, the agents may choose messages that induce outcomes undesirable for the principal based on some beliefs of them about the principal's preference. If the principal does not know the agents' beliefs about the principal's preference, he may not want to make the mechanism contingent on his own message.

Finally, although it is certainly a restrictive assumption, we show in Section 6 that our conclusion about the optimality of full disclosure in applications would not change even if we allow for a general class of mechanisms.

In summary, the timing of the game is the following. (i) The principal commits to a disclosure policy ϕ (or (μ, ψ)) and a mechanism g; (ii) θ is realized and the principal publicly announces a message m according to ϕ; (iii) each agent i observes v_i and m, and chooses whether to participate in the mechanism (if not, he is always given utility 0) and his message in the mechanism; finally, (iv) an allocation is determined by the mechanism and the agents' messages.

The problem for the principal is given as follows.

$$\begin{aligned}
\max_{\phi, g} \quad & \int_{\Theta \times \mathcal{M}} \left(\int_V u_0(g_m(v), v, \theta) dF_V \right) d\phi(\theta, m) \\
\text{sub. to} \quad & \int_{V_{-i} \times \Theta} u_i(g_m(v), v, \theta) dF_i(v_{-i}) d\psi(\theta|m) \geq \int_{V_{-i} \times \Theta} u_i(g_m(v_i', v_{-i}), v, \theta) dF_i(v_{-i}) d\psi(\theta|m) \\
& \int_{V_{-i} \times \Theta} u_i(g_m(v), v, \theta) dF_i(v_{-i}) d\psi(\theta|m) \geq 0.
\end{aligned}$$

As is standard, the first constraint in the problem corresponds to each agent's Bayesian incentive compatibility condition. Because each agent observes a public message m by the principal, his expected utility is computed using his posterior $\psi(\cdot|m)$ over θ. The second constraint is each agent's interim individual rationality or participation condition.

3 Optimality of full disclosure

The principal's problem can be interpreted as the following two-step problem. First, she chooses a disclosure policy ϕ. Then, after announcing m, she chooses a mechanism $g_m : V \to X$, assuming that the agents' belief over Θ is given by $\psi(\cdot|m)$.

This second step is a standard mechanism design problem. Given $\psi \in \Delta(\Theta)$, let

$$\begin{aligned}\gamma^*(\psi) = \arg\max_{\gamma: V\to X} \quad & \int_\Theta \int_V u_0(\gamma(v), v, \theta) dF_V(v) d\psi(\theta) \\ \text{sub.to} \quad & \int_{V_{-i}\times\Theta} u_i(\gamma(v), v, \theta) dF_i(v_{-i}) d\psi(\theta) \geq \int_{V_{-i}\times\Theta} u_i(\gamma(v_i', v_{-i}), v, \theta) dF_i(v_{-i}) d\psi(\theta) \\ & \int_{V_{-i}\times\Theta} u_i(\gamma(v), v, \theta) dF_i(v_{-i}) d\psi(\theta) \geq 0,\end{aligned}$$

be the optimal mechanism given ψ as the agents' belief over Θ. We assume that such $\gamma^*(\psi)$ exists for all ψ. Let $S^*(\psi)$ denote the maximized value of this second-step problem. In an extreme case where ψ is a degenerated distribution that assigns probability one for a specific value θ, we also denote them by $\gamma^*(\theta)$ and $S^*(\theta)$.

The following lemma is straightforward but useful.

Lemma 1. If (ψ, g) is an optimal strategy for the principal, then $g_m = \gamma^*(\psi(\cdot|m))$ for μ-almost every m.

The lemma means that we can disentangle the mechanism design problem from the disclosure problem.

Proof. Let $(\psi, \tilde{g})$ be an optimal strategy of the principal. Then, for each m, $\tilde{g}$ satisfies the incentive compatibility and individual rationality conditions. Thus,

$$\begin{aligned} & \int_{\Theta\times\mathcal{M}} \left(\int_V u_0(\tilde{g}_m(v), v, \theta) dF_V \right) d\phi(\theta, m) \\ = & \int_{\mathcal{M}} \int_\Theta \left(\int_V u_0(\tilde{g}_m(v), v, \theta) dF_V \right) d\psi(\theta|m) d\mu(m) \\ \leq & \int_{\mathcal{M}} S^*(\psi(\theta|m)) d\mu(m) \\ = & \int_{\mathcal{M}} \int_\Theta \left(\int_V u_0(g_m(v), v, \theta) dF_V \right) d\psi(\theta|m) d\mu(m). \end{aligned}$$

The LHS is the expected utility of the principal given $(\psi, \tilde{g})$, while the RHS is that given (ψ, g). By construction, $g_m = \gamma^*(\psi(\cdot|m))$ satisfies the incentive compatibility and individual rationality conditions for each m. Therefore, if $(\psi, \tilde{g})$ is a solution, then so is (ψ, g). □

Hence, the ex-ante maximization problem for the principal boils down to an information disclosure problem. We now present a simple condition that basically characterizes the optimality for full disclosure, for a class of environments including the auction and bilateral trade applications we discuss below.

We say that the model is *linear in* θ if u_0 and u_i are linear functions of θ. Although it is restrictive, it includes some standard applications as we see in the subsequent sections. In a linear model, suppose that the principal follows a strategy (ϕ, g), and given an announcement m, the agents' belief over Θ is given by $\psi(\cdot|m)$. Then, the agents' incentive compatibility and individual rationality depend only on the expected value of θ with respect to $\psi(\cdot|m)$, instead of the entire distribution of $\psi(\cdot|m)$. To see this formally, for each $\psi \in \Delta(\Theta)$, let $\mathbb{E}_\psi[\theta] = \int_\Theta \theta d\psi(\theta)$. Then, agent i's expected utility when his type is v_i and he reports v_i' is

$$\int_{V_{-i}\times\Theta} u_i(\gamma(v_i', v_{-i}), v, \theta) dF_i(v_{-i}) d\psi(\theta) = \int_{V_{-i}} u_i(\gamma(v_i', v_{-i}), v, \mathbb{E}_\psi[\theta]) dF_i(v_{-i}),$$

and therefore, his incentive compatibility and individual rationality depend only on $\mathbb{E}_\psi[\theta]$.

Moreover, the principal's expected utility also depends only on $\mathbb{E}_\psi[\theta]$. Therefore, for any given $\psi \in \Delta(\Theta)$, we have $\gamma^*(\psi) = \gamma^*(\mathbb{E}_\psi[\theta])$, and $S^*(\psi) = S^*(\mathbb{E}_\psi[\theta])$. This property is useful in characterizing optimality of full disclosure based on convexity of the value function for degenerated distributions, $S^*(\theta)$, $\theta \in \Theta$.[4]

Theorem 1. In the linear model, full disclosure is optimal if $S^*(\theta)$ is convex on Θ, and full disclosure is not optimal if there is a convex subset $A \subseteq \Theta$ such that (i) $F_\Theta(A) > 0$ and (ii) $S^*(\theta)$ is strictly concave on A.

Recall that $S^*(\theta)$ is the value function of the principal after revealing her information θ to the agents (i.e., ψ is degenerated). To examine the optimality of full disclosure, the theorem says that we only need to calculate the optimal mechanism *assuming* that everyone knows θ. This is a useful property in the following sense. Often in applications, we assume that v follows a distribution that satisfies nice "regularity" properties such as the monotone hazard rate condition. Given that F_V is exogenously given, such an assumption may be acceptable as a simplifying assumption. However, assuming such nice properties for ψ can be problematic, because it restricts feasible disclosure strategies of the principal. An optimal disclosure policy may involve a complicated ψ that does not satisfy such nice properties. The theorem says

[4]In the sense that the following theorem does not provide an "if and only if" condition, rigorously it is not a characterization result. However, we believe it is close to a tight condition.

that we do not worry about such an issue, because it is sufficient to only investigate cases where θ is fully revealed to the agents, as long as optimality of full disclosure is concerned.

4 Auction

In this section, we apply Theorem 1 to an auction problem. There is a seller and N bidders. We assume that $\theta = (\theta_1, \ldots, \theta_N)$ is an N-dimensional vector. An allocation is denoted by $a = (x_i, p_i)_{i=1}^N$ where x_i is the amount of good that agent i obtains and $p_i \in \mathbb{R}$ is agent i's payment to the seller.

Each bidder i has valuation

$$\tilde{v}_i = v_i + \theta_i$$

for the good, and his utility takes a linear form $\tilde{v}_i x_i - p_i$. We can interpret θ_i as related to a quality measure known by the seller, and v_i as a private or idiosyncratic component of i's value. The cumulative density function and probability density function for $v_i \in V_i$ are given by F_i and f_i respectively. We assume that each agent i's "virtual valuation", i.e. $v_i - \frac{1-F_i(v_i)}{f_i(v_i)}$, is monotonically increasing in v_i. In addition, we assume $(v_1, \ldots, v_N)$ are mutually independent, but θ_i, θ_j can potentially be correlated.

We consider a general auction here. Namely we allow for an arbitrary feasible set of allocations, that is $x \in \mathcal{X}$. The seller incurs cost $c(x)$ when providing a vector x of goods to agents. The seller's objective is a weighted sum of revenue and surplus, $u_0(a, v, \theta) = \omega(\sum_i p_i) + (1-\omega)(\sum_i \tilde{v}_i x_i) - c(x)$, for some weight $\omega \in [0, 1]$.

Note that our general auction model includes, in particular, single-good auction with revenue-maximizing seller ($\mathcal{X} = \{x | \sum x_i \leq 1\}$, $\omega = 1$ and $c(x) = c_0 \sum x_i$) and public good auction with benevolent seller ($\mathcal{X} = \{x | x_1 = \cdots = x_N\}$, $\omega = 0$ and $c(x)$ is a convex cost function)

Clearly, the auction model exhibits linearity in θ. Therefore, in order to apply Theorem 1 we only need to check if the seller's value (her utility obtained under an optimal auction mechanism) is convex. If θ is common knowledge, the seller's problem is given as follows,

$$\begin{aligned} \max_{x_i(\cdot), p_i(cdot)} \quad & \mathbf{E}_v[\omega(\sum_i p_i) + (1-\omega)(\sum_i (v_i + \theta_i) x_i) - c(x)] \\ \text{s.t.} \quad & x_i, p_i \text{ satisfy (IC), (IR) for each agent } i \\ & x \in \mathcal{X} \end{aligned} \tag{4.1}$$

The following lemma simplifies seller's problem after replacing p_i from agents' incentive compatibility conditions.

Lemma 2. The seller's problem in 4.1 is equivalently given by

$$\max_{x\in\mathcal{X}} \mathbf{E}_v[\sum_i (v_i - \omega\frac{1-F_i(v_i)}{f_i(v_i)} + \theta_i)x_i - c(x)] \tag{4.2}$$

Proof. This lemma is a standard generalization of Myerson (1981). The derivation is omitted here and we only prove that x_i^* obtained from problem 4.2 is indeed increasing in v_i. Denote $\tau_i^\omega(v_i,\theta_i) \equiv v_i - \omega\frac{1-F_i(f_i)}{f_i(v_i)} + \theta_i$. By our assumption, τ_i^1 is increasing in v_i. So τ_i^ω is increasing in v_i for all $\omega \in (0,1)$.

Consider two value vector for agents $v = (v_1,\ldots,v_n,\ldots,v_N)$ and $v = (v_1,\ldots,v_n',\ldots,v_N)$ with $v_n' > v_n$. Since $x^*(\cdot)$ is obtained from pointwise maximization, we have,

$$\sum_i \tau_i^\omega(v_i,\theta_i)x_i^*(v) - c(x^*(v)) \geq \sum_i \tau_i^\omega(v_i,\theta_i)x_i^*(v') - c(x^*(v')) \tag{4.3}$$

Similarly we have,

$$\sum_i \tau_i^{\omega\prime}(v_i,\theta_i)x_i^*(v') - c(x^*(v')) \geq \sum_i \tau_i^{\omega\prime}(v_i,\theta_i)x_i^*(v) - c(x^*(v)) \tag{4.4}$$

Recall that $\tau_i^\omega = \tau_i^{\omega\prime}$ for all $i \neq n$. Combining above two inequalities we obtain

$$(\tau_n^{\omega\prime} - \tau_n^\omega)(x_n^*(v') - x_n^*(v)) \geq 0 \tag{4.5}$$

□

Denote seller's value as $S(\theta) = \mathbf{E}_v[u_0(a^*(\theta),v,\theta)]$, where $a^*(\theta) = (x_i^*(v_i,\theta), p_i^*(v_i,\theta))_{i=1}^N$ is the optimal allocation as the solution of seller's problem in 4.1. We now show that $S(\theta)$ is convex in θ_i for each i.

Theorem 2. Full disclosure is optimal for the seller.

Proof. Consider θ,θ' and $\alpha \in (0,1)$. Denote $\bar\theta \equiv \alpha\theta + (1-\alpha)\theta'$ We need to show that

$$\alpha S(\theta) + (1-\alpha)S(\theta') \geq S(\bar\theta) \tag{4.6}$$

Or,

$$\alpha\mathbf{E}_v[u_0(a^*(\theta),v,\theta)] + (1-\alpha)\mathbf{E}_v[u_0(a^*(\theta'),v,\theta')] \geq \mathbf{E}_v[u_0(a^*(\bar\theta),v,\bar\theta)] \tag{4.7}$$

Recall that u_0 is linear in θ, we have the right-hand-side,

$$\mathbf{E}_v[u_0(a^*(\bar{\theta}), v, \bar{\theta}] = \alpha \mathbf{E}_v[u_0(a^*(\bar{\theta}), v, \theta)] + (1-\alpha)\mathbf{E}_v[u_0(a^*(\bar{\theta}), v, \theta')] \tag{4.8}$$

Since θ_i is orthogonal to v_i, agent i's incentive compatibility condition is independent of θ_i. In particular, this means that $a^*(\bar{\theta})$ is also a feasible allocation for the seller with information θ or θ'. By optimality we have,

$$\begin{aligned} \mathbf{E}_v[u_0(a^*(\theta), v, \theta)] &\geq \mathbf{E}_v[u_0(a^*(\bar{\theta}), v, \theta)] \\ \mathbf{E}_v[u_0(a^*(\theta'), v, \theta')] &\geq \mathbf{E}_v[u_0(a^*(\bar{\theta}), v, \theta')] \end{aligned} \tag{4.9}$$

Combining 4.8 and 4.9 gives the convexity of seller's value function. By Theorem 1, full information disclosure is optimal for the seller.

□

5 Bilateral Trade

In this section, we consider a bilateral trade setting. The principal is a mediator or social planner who designs a mechanism to achieve trade surplus. There are two agents, a buyer with value $\tilde{v} = v + \theta$ and a seller with value c, where $v \in [0,1]$ is the buyer's private information, $c \in [0,1]$ is the seller's private information, and $\theta \in [0,1]$ is the principal's private information. We assume that they are mutually independent, and the density for v is $f(v)$, the density for c is $g(c)$. As opposed to the auction case, a feasible trade allocation must satisfy budget balance condition. Hence, we denote a feasible allocation by $(p,t) \in [0,1] \times \mathbb{R}$, where p is the probability of trade, and t is the payment from the buyer to the seller.

5.1 The first-best scenario

We first show that, in the first-best scenario where v, c are observable to the principal, the principal is always better off by fully disclosing θ. This would be our benchmark to compare with the second-best scenario where v, c are the agents' private information.

Proposition 1. Full disclosure is optimal for the mediator when v, c are observable to her.

Proof. It suffices to show that the principal's value function assuming θ being fully disclosed is convex. Under the first-best scenario, the agents trade if and only if $v + \theta > c$, so we have

$$S^{FB}(\theta) = \int_0^\theta \int_0^1 (v+\theta-c) f(v) g(c) dv dc + \int_\theta^1 \int_{c-\theta}^1 (v+\theta-c) f(v) g(c) dv dc.$$

Thus,

$$\begin{aligned}
S_{\theta}^{FB} &= \int_0^{\theta}\int_0^1 f(v)g(c)dvdc + \int_{\theta}^1\int_{c-\theta}^1 f(v)g(c)dvdc \\
&\quad + \int_0^1 (v+\theta-\theta)f(v)g(\theta)dv - \int_{\theta-\theta}^1 (v+\theta-\theta)f(v)g(\theta)dv \\
&\quad + \int_{\theta}^1 (c-\theta+\theta-c)f(c-\theta)g(c)dc \\
S_{\theta\theta}^{FB} &= \int_0^1 f(v)g(\theta)dv + \int_{\theta}^1 f(c-\theta)g(c)dc \\
&\quad > 0
\end{aligned}$$

Therefore, $S^{FB}(\theta)$ is convex. □

The intuition here is similar to the auction example. Higher θ increases surplus per trade as well as increases trading probability. However, since the marginal trade gives nearly zero surplus so the second channel has no first order effect. Hence, the second order effect is determined by the increases in trading probability and thus positive.

5.2 The second-best scenario

We now show that, when v, c are the agents' private information, then the principal's (second-best) value function is not convex. Moreover, there exists a region of θ on which the value function exhibits strict concavity, and thus, by the second statement of Theorem 1, full disclosure is not optimal. Define the buyer's virtual value as $v - \gamma(v)$ where $\gamma(v) = \frac{1-F(v)}{f(v)}$, and the seller's virtual value as $c + \phi(c)$ where $\phi(c) = \frac{G(c)}{g(c)}$. As regularity conditions, we assume that $v - \gamma(v)$ is increasing in v, and $c + \phi(c)$ is increasing in c. Moreover, $f(v), g(c)$ are bounded away from zero for all $c, v \in [0, 1]$, and are differentiable.

Theorem 3. Full disclosure is not optimal for the mediator in the second-best scenario.

Proof. First, given that θ being disclosed to the agents, the second-best mechanism is a simple modification of the optimal mechanism of Myerson and Satterthwaite (1983). That is, letting $\lambda(\theta)$ be the Lagrange multiplier for the expected budget balance condition, the

optimal allocation and the value function $S^{SB}(\theta)$ satisfies

$$S^{SB}(\theta) = \int_{c=0}^{1} \int_{v=v^*(c,\theta)}^{1} [(v+\theta-c)(1+\lambda(\theta)) - \lambda(\theta)(\phi(c)+\gamma(v))] dF dG \tag{5.1}$$

$$\int_0^1 \int_{v^*}^1 (v+\theta-c-\phi(c)-\gamma(v)) dF dG = 0 \tag{5.2}$$

$$(v^*(\theta,c)+\theta-c)(1+\lambda(\theta)) - \lambda(\theta)(\phi(c)+\gamma(v^*)) = 0 \tag{5.3}$$

The first equation is the definition of $S^{SB}(\theta)$, the second equation stands for the expected budget constraint condition, and the third condition defines the optimal trade rule $v^*(\theta, c)$, meaning that the agents trade if and only if the buyer reports $v > v^*(\theta, c)$ given θ and the seller's report c.

Note that, by (5.2), v^* is continuous and continuously differentiable both in θ, c. (5.3) then implies that λ is continuous and continuously differentiable in θ. Our goal is to show that the second derivative of S^{SB}, $S^{SB}_{\theta\theta}$, is negative for θ close to 1:

$$S^{SB}_{\theta} = \int_0^1 \int_{v^*(c,\theta)}^1 (1+\lambda(\theta)) f g dv dc \tag{5.4}$$

$$S^{SB}_{\theta\theta} = \lambda' \int_0^1 \int_{v^*}^1 f g dv dc - (1+\lambda) \int_0^1 g f(v^*) v^*_\theta dc \tag{5.5}$$

We first show the following lemma, which states that, as $\theta \to 1$, the agents trade for most of the realizations (v, c).

Lemma 3. There is $\overline{\theta} \in (0,1)$ such that for $\theta > \overline{\theta}$, there exists $\hat{c}(\theta) > 0$ such that $v^*(\theta, c) = 0$ for all $c < \hat{c}$. Moreover, $\hat{c}(\theta) \to 1$ is differentiable, and as $\theta \to 1$, $\hat{c}(\theta) \to 1$.

Proof. (of the lemma)

As $\theta \to 1$, $\lambda \to 0$, and hence $\frac{\lambda}{1+\lambda} \to 0$. From (5.3) we have,

$$(\hat{v} - \frac{\lambda}{1+\lambda}\gamma(\hat{v})) + \theta - (c + \frac{\lambda}{1+\lambda}\phi(c)) = 0$$

For $\forall c < 1$ and $v = 0$, we have

$$\begin{aligned} &(0 - \frac{\lambda}{1+\lambda}\gamma(0)) + \theta - (c + \frac{\lambda}{1+\lambda}\phi(c)) \\ =&\theta - c - \frac{\lambda}{1+\lambda}\left(\frac{1}{f(0)} + \frac{G(c)}{g(c)}\right) \end{aligned} \tag{5.6}$$

Since $c < 1$ and $\frac{1}{f(0)} + \frac{G(c)}{g(c)} < +\infty$, we can find large enough θ(θ close to 1) such that above expression is positive. This means that the point $(0, c)$ lies strictly below the trading

line $v^*(\theta, c)$, or equivalently, we have $v^*(\theta, c) = 0$. Differentiability of $\hat{c}$ is straightforward from the argument above. □

By Lemma 3, for each $c < 1$, for large enough θ (i.e., close to 1), $v^*(\theta, c)$ becomes constant (zero) in θ (but note that it is not necessarily the case for $c = 1$). Thus, for as $\theta \to 1$, the second term in (5.5) vanishes, and therefore, in order to show the second derivative of S^{SB} is negative, it suffices to show that $\lambda'(\theta)$ becomes strictly negative as $\theta \to 1$.

Taking derivative of (5.3) with respect to c we have,

$$(v_c - 1)(1 + \lambda) - \lambda(\phi' + \gamma' v_c) = 0,$$

where v_c means the derivative of $v^*(\theta, c)$ with respect to c, or equivalently,

$$v_c = \frac{1 + \lambda + \lambda\phi'}{1 + \lambda - \lambda\gamma'}.$$

So as $\theta \to 1$, we have $v_c \to 1$, meaning that the trading line converges to a straight line.

Similarly, taking derivative of (5.3) respect to θ, we have the following: letting v_θ denote the derivative of $v^*(\theta, c)$ with respect to θ,

$$(v_\theta + 1)(1 + \lambda) + (v + \theta - c)\lambda_\theta - \lambda_\theta(\phi + \gamma) - \lambda\gamma' v_\theta = 0$$

or

$$\lambda_\theta = \frac{(v_\theta + 1)(1 + \lambda) - \lambda\gamma' v_\theta}{\phi + \gamma - (v^* + \theta - c)}$$

Since λ does not depend on c, we can evaluate the above expression at $c = 1$. Let $\theta \to 1$. Then we have $v^*(\theta, 1) \to 0$ and $\lambda \to 0$. Thus,

$$\lambda_\theta \xrightarrow{\theta \to 1} \frac{v_\theta + 1}{\phi + \gamma}$$

So λ_θ is bounded below 0 in the limit if and only if $\lim_{\theta \to 1} v_\theta(\theta, 1) < -1$.

In fact, from the budget balance equation (5.2), we have,

$$\begin{aligned} 0 &= \int_0^1 \int_{v^*}^1 (v + \theta - c - \phi(c) - \gamma(v)) dF dG \\ &= \int_0^{\hat{c}(\theta)} \int_0^1 (v + \theta - c - \phi(c) - \gamma(v)) dF dG + \int_{\hat{c}(\theta)}^1 \int_{v^*}^1 (v + \theta - c - \phi(c) - \gamma(v)) dF dG \end{aligned}$$

Taking derivative respect to θ we have

$$\begin{aligned} 0 =& \hat{c}_\theta \int_0^1 (v + \theta - \hat{c} - \phi(\hat{c}) - \gamma(v)) g(\hat{c}) dF + \int_0^{\hat{c}(\theta)} \int_0^1 1 dF dG \\ &- \hat{c}_\theta \int_0^1 (v + \theta - \hat{c} - \phi(\hat{c}) - \gamma(v)) g(\hat{c}) dF + \int_{\hat{c}(\theta)}^1 \int_{v^*}^1 1 dF dG \\ &- \int_{\hat{c}(\theta)}^1 v_\theta(\theta, c)(v^* + \theta - c - \phi(c) - \gamma(v^*)) f(v^*) dG, \end{aligned}$$

or

$$\int_{\hat{c}(\theta)}^{1} v_\theta(\theta, c)(v^* + \theta - c - \phi(c) - \gamma(v^*))f(v^*)dG = \int_0^{\hat{c}(\theta)} \int_0^1 1dFdG + \int_{\hat{c}(\theta)}^1 \int_{v^*}^1 1dFdG.$$

As $\theta \to 1$, the RHS converges to 1. For the LHS, we have $\hat{c}(\theta) \to 1$. Moreover, we can find $\theta^* \in (0,1)$ such that for any $\theta > \theta^*$, and for any $c \in [\hat{c}, 1]$, we have $v_\theta(\theta, c) < 1$. Because $v^* + \theta - c$ converges to 0 and $\phi(c) + \gamma(v^*)$ converges to a positive constant as $\theta \to 1$ (and $c \to 1$ accordingly), $v^* + \theta - c - \phi(c) - \gamma(v^*)$ must be negative (though bounded) for any $\theta > \theta^*$ and $c \in [\hat{c}, 1]$. To equate both sides of the equation, $v_\theta(\theta, c)$ must becomes negative without bound. Therefore, $S^{SB}(\theta)$ is strictly concave on a convex subset of Θ, $[\theta^*, 1]$. □

The intuition for above result is as follows. By Equation 5.5, we can see that $S(\theta)$ is increasing in θ by two effects. First, when θ increases slightly, total trade surplus becomes higher since the trade surplus per existing trade is increased by the same amount. Second, as $\lambda(\theta) > 0$ for $\theta < 1$, with higher θ the planner can implement more trade which also increase the total surplus. As we discussed previously, the second effect does not exist in the auction setting or in the first-best scenario for bilateral trade. The curvature of $S(\theta)$ is determined by the total change of the two effects in Equation 5.5. Obviously, the first effect becomes stronger for higher θ since more trades are implemented but increases in a slower speed as $\theta \to 1$. The change of the second effect is captured by λ_θ which is negative. We have shown that $\lambda_\theta \overset{\theta \to 1}{\to} -\infty$. This means that for high enough θ the change in the second effect dominates the change in the first effect, and the total change is negative, hence concave of $S(\theta)$ for a region where θ is close to 1.

5.3 Analytical Example

We illustrate Theorem 3 by the following example with uniformly distributed values. Let $v, c \sim U[0,1]$. By Myerson and Satterthwaite (1983), the second-best rule is given by,

$$p^*(v,c) = \begin{cases} 1 & \text{if} \quad v \geq c + \alpha(\theta) \\ 0 & \text{if} \quad o.w. \end{cases} \tag{5.7}$$

where α is determined by the budget balance condition $\int_\theta^{\theta+1} \int_0^1 (2v - 1 - \theta - 2c)\, p^*(v,c) dc dv = 0$.

Obviously, we have $\alpha(0) = \frac{1}{4}$ and $\alpha(1) = 0$. A visual illustration is given below for both $\alpha(\theta) > \theta$ and $\alpha(\theta) < \theta$ cases.

[figures here]

Note that,

$$\begin{aligned}
&\int_{\theta}^{\theta+1}\int_{0}^{1}(2v-1-\theta-2c)\,p^*(v,c)dcdv\\
=&2\int_{\theta}^{\theta+1}\int_{0}^{1}(v-c)\,p^*(v,c)dcdv-(\theta+1)\int_{\theta}^{\theta+1}\int_{0}^{1}p^*(v,c)dcdv\\
=&2\cdot\text{Total surplus}-(\theta+1)\cdot\text{Total probability of trade}
\end{aligned}$$

For case 2(a) where $\alpha>\theta$, we have,

$$\begin{aligned}
0=&\int_{\theta}^{\theta+1}\int_{0}^{1}(2v-1-\theta-2c)\,p^*(v,c)dcdv\\
=&2\int_{\alpha}^{\theta+1}u(\theta+1-u)du-(\theta+1)\frac{1}{2}(\theta+1-\alpha)^2\\
=&2\left[(\theta+1)\frac{(\theta+1)^2-\alpha^2}{2}-\frac{(\theta+1)^3-\alpha^3}{3}\right]-(\theta+1)\frac{1}{2}(\theta+1-\alpha)^2\\
=&\frac{\theta+1-\alpha}{6}\left[3(\theta+1)(\theta+3\alpha+1)-4\left((\theta+1)^2+(\theta+1)\alpha+\alpha^2\right)\right]\\
=&\frac{1}{6}(\theta+1-\alpha)^2(4\alpha-\theta-1)
\end{aligned}$$

So we obtain $\alpha(\theta)=\frac{\theta+1}{4}$ for $\theta\in[0,\frac{1}{3}]$. The second-best trading surplus is given by

$$\begin{aligned}
S^{SB}(\theta)=&\frac{1}{2}(\theta+1)\frac{1}{2}\cdot(\theta+1-\alpha(\theta))^2\\
=&\frac{9}{64}(\theta+1)^3
\end{aligned}$$

For case 2(b) where $\alpha<\theta$ and $\theta\in[\frac{1}{3},1]$, we have,

$$\begin{aligned}
0=&2\left(\int_{\theta}^{\theta+1}u(\theta+1-u)du+\int_{\alpha}^{\theta}u(1+u-\theta)du\right)-(1+\theta)\left(1-\frac{1}{2}(1-(\theta-\alpha))^2\right)\\
=&2\left(\frac{\theta+\theta+1}{2}-\frac{1}{2}+\int_{\theta-1}^{\alpha}u(1+u-\theta)du\right)-(1+\theta)\left(1-\frac{1}{2}(1-(\theta-\alpha))^2\right)\\
=&2\left[\theta-\frac{1-\theta}{2}\left(\alpha^2-(\theta-1)^2\right)-\frac{1}{3}\left(\alpha^3-(\theta-1)^3\right)\right]-(\theta+1)+\frac{1}{2}(1+\theta)(1+\alpha-\theta)^2\\
=&\theta-1+\frac{\alpha+1-\theta}{6}\left[3(1+\theta)(\alpha+1-\theta)-6(1-\theta)(\alpha-(1-\theta))-4\left(\alpha^2-\alpha(1-\theta)+(1-\theta)^2\right)\right]\\
=&\theta-1-\frac{1}{6}(\alpha+1-\theta)(4\alpha-\theta-5)
\end{aligned}\tag{5.8}$$

So in this case we cannot analytically solve $\alpha(\theta)$ as well as $S^{SB}(\theta)$. However, we can still examine the convexity of S^{SB}. Note that,

$$S^{SB}(\theta)=\frac{1+\theta}{2}\left(1-\frac{(\alpha(\theta)(\theta)+1-\theta)^2}{2}\right)$$

And by (5.8) and implicit function theorem we have,

$$\frac{\partial\alpha(\theta)}{\partial\theta} = -\frac{6+2(\alpha+1-\theta)(4\alpha-\theta-5)+(\alpha+1-\theta)^2}{-2(\alpha+1-\theta)(4\alpha-\theta-5)-4(\alpha+1-\theta)^2}$$

or

$$\begin{aligned}\frac{\partial\alpha(\theta)}{\partial\theta} - 1 =& \frac{6-3(\alpha+1-\theta)^2}{2(\alpha+1-\theta)(4\alpha-\theta-5)+4(\alpha+1-\theta)^2} \\ =& \frac{1-\frac{(\alpha+1-\theta)^2}{2}}{(\alpha+1-\theta)(2\alpha-\theta-1)}\end{aligned}$$

So,

$$\begin{aligned}\frac{\partial S^{SB}(\theta)}{\partial\theta} =& \frac{1}{2}\left(1-\frac{(\alpha+1-\theta)^2}{2}\right)+\frac{\theta+1}{2}\left(-(\alpha+1-\theta)(\frac{\partial\alpha}{\partial\theta}-1)\right) \\ =& \frac{1}{2}\left(1-\frac{(\alpha+1-\theta)^2}{2}\right)\left(1-\frac{\theta+1}{2\alpha-\theta-1}\right) \\ =& \left(1-\frac{(\alpha+1-\theta)^2}{2}\right)\frac{1+\theta-\alpha}{1+\theta-2\alpha}\end{aligned}$$

One can see that $S_\theta^{SB}(\frac{1}{3}^+) = \frac{3}{4} = S_\theta^{SB}(\frac{1}{3}^-)$ and $S_\theta^{SB}(1^-) = 1 = S_\theta^{SB}(1^+)$. So $S_\theta^{SB}(\theta)$ is continuous for all $\theta \in [0,1]$.

For the second derivative, we have

$$\begin{aligned}\frac{\partial^2 S^{SB}(\theta)}{\partial\theta^2} =& \left(-(\alpha+1-\theta)(\frac{\alpha}{\theta}-1)\right)\frac{1+\theta-\alpha}{1+\theta-2\alpha}+\left(1-\frac{(\alpha+1-\theta)^2}{2}\right)\frac{(1-\alpha_\theta)(1+\theta-2\alpha)-(1+\theta-\alpha)}{(1+\theta-2\alpha)^2} \\ =& -\left(1-\frac{(\alpha+1-\theta)^2}{2}\right)\frac{1}{2\alpha-1-\theta}\cdot\frac{1+\theta-\alpha}{1+\theta-2\alpha}+\left(1-\frac{(\alpha+1-\theta)^2}{2}\right)\frac{-\alpha+\alpha_\theta(1+\theta)}{(1+\theta-2\alpha)^2} \\ =& \frac{1-\frac{(\alpha+1-\theta)}{2}}{(1+\theta-2\alpha)^2}\left(1+\theta-\alpha-\alpha+\alpha_\theta(1+\theta)\right) \\ \propto& (\theta+1)(\alpha_\theta+1)-2\alpha\end{aligned}$$

We have $S_{\theta\theta}^{SB}(\frac{1}{3}^+) = \frac{9}{8} = S_{\theta\theta}^{SB}(\frac{1}{3}^-)$. However $S_{\theta\theta}^{SB}(1^-) = -\infty$. Therefore, for θ close to one, S^{SB} becomes strictly concave, and thus, full disclosure is not optimal. The shape of second-best surplus is given by the figure below. The first-best surplus in this uniform-distribution example is given by $S^{FB}(\theta) = \theta + \frac{(1-\theta)^3}{6}$, which is globally convex.

[figures here]

As suggested by above figure, in the second-best scenario, the optimal disclosure strategy for the principal is to fully reveal θ for $\theta \in [0,\theta_0)$ (where $\theta_0 = 0.8683$), and pool all θ for $\theta \in (\theta_0, 1]$. Compared to fully disclosing all θ, this optimal partial disclosure increases the

expected surplus by 0.009%. Hence, in this uniform example, although full disclosure is not optimal, it is "not so far" from the optimum. It may be an interesting future question to investigate if such an observation that full disclosure is "not so far" from the optimum is true for other distributions.

6 Voting

In this section, we consider a simple voting example where Theorem 1 is applied to determine the optimal information revelation strategy. Consider two voters, $i = 1, 2$, collectively choosing whether or not to take a *reform* for a certain policy. If the reform is not chosen, there is a *status quo* policy. Without loss of generality, we can normalize each voter's utility from the status quo to 0 while voter i obtains utility $\tilde{u}_i = u_i + \theta_i$ if the reform is chosen. We assume u_i is a random variable with distribution $U[a, b]$, $a < 0 < b$[5], and its realization is only observed by voter i himself. The additional utility term θ_i is not observed by voter i but observed by the mechanism designer. We assume $u_1, u_2, \theta_1, \theta_2$ are all independent and $\theta_i \sim U[-b, -a]$.[6] The mechanism designer's objective is to maximize ex ante social welfare, by deciding the information revelation about $\theta = (\theta_1, \theta_2)$ as well as by designing the voting mechanism.

We first look at the mechanism designing problem for the designer when θ is common knowledge. According to Azrieli and Kim (2013), the optimal voting mechanism is a weighted majority rule such that the social choice function,

$$f(\tilde{u}_1, \tilde{u}_2) = \begin{cases} \text{reform} & \textit{if} \quad \sum_{i:\tilde{u}_i>0} \omega_i > \sum_{i:\tilde{u}_i<0} \omega_i \\ \text{status quo} & \textit{otherwise} \end{cases}$$

where $\tilde{u}_i$ is the reported utility by voter i and ω_i is the weight given for voter i's voting. The optimal voting weight is,

$$\omega_i = \begin{cases} \mathbf{E}[\tilde{u}_i | \tilde{u}_i > 0] & \textit{if} \quad \tilde{u}_i > 0 \\ -\mathbf{E}[\tilde{u}_i | \tilde{u}_i < 0] & \textit{if} \quad \tilde{u}_i < 0 \end{cases}$$

[5]In general a_i and b_i can be different across voters. However, the results from the asymmetric case are qualitatively the same so here we only focus on the symmetric case.

[6]Note that under this assumption, $\tilde{u}_i$ can always be either negative or positive for any realization of θ_i. Later we will discuss how much we can generalize the assumptions about distributions after the main result of this section is presented.

To simplify notations, we introduce the conditional mean of voter i's utility when the realization is positive. When θ is public knowledge, we have,

$$\begin{aligned}\overline{u}_i(\theta_i) &\equiv \mathbf{E}[\tilde{u}_i|\tilde{u}_i > 0] \\ &= \theta_i + \mathbf{E}[u_i|u_i > -\theta_i] \\ &= \frac{\theta_i + b}{2}\end{aligned} \tag{6.1}$$

Similarly, the conditional mean when realized utility is negative is given by

$$\underline{u}_i(\theta_i) = \frac{\theta_i + a}{2} \tag{6.2}$$

The probability that the realized $\tilde{u}_i$ being positive is given by,

$$\begin{aligned}p_i(\theta_i) &\equiv \text{Prob}\{\tilde{u}_i > 0\} \\ &= \text{Prob}\{u_i > -\theta_i\} \\ &= \frac{b + \theta_i}{b - a}\end{aligned} \tag{6.3}$$

Since $\theta_i \in [-b, -a]$, we always have $\overline{u}_i \geq 0$ and $\underline{u}_i \leq 0$. Following Azrieli and Kim (2013), the optimal voting rule only considers the sign of the reported utility by one voter, while putting different weight across voters. Figure 1 summaries the optimal voting rule (whether or not to choose reform) for different realization of θ.

$\{\Theta_A, \Theta_B\}$ is a partition of the full support $\Theta = [-b, -a] \times [-b, -a]$. Under the optimal voting rule, reform will be chosen: i) if and only if both voters report positive utility when $\theta \in \Theta_A$ and; ii) if and only if at least one voter reports positive utility when $\theta \in \Theta_B$. For a given θ, the ex ante social welfare is defined as the expected total utility for voters.

$$S(\theta) = \int_a^b \int_a^b (u_1 + \theta_1 + u_2 + \theta_2) \mathbb{1}_{\{f(u_1+\theta_1, u_2+\theta_2)=\text{reform}\}} d\Phi(u_2) d\Phi(u_1)$$

It is easy to verify that for θ in each subset of Θ, the social welfare function is given by,

$$S(\theta) = \begin{cases} p_1(\theta_1)p_2(\theta_2)(\overline{u}_1(\theta_1) + \overline{u}_2(\theta_2)) & \text{if} \quad \theta \in \Theta_A \\ \frac{a_1+b_1}{2} + \theta_1 + \frac{a_2+b_2}{2} + \theta_2 - (1 - p_1(\theta_1))(1 - p_2(\theta_2))(\underline{u}_1(\theta_1) + \underline{u}_2(\theta_2)) & \text{if} \quad \theta \in \Theta_B \end{cases} \tag{6.4}$$

Now suppose that θ is only observed by the mechanism designer who can decide how to reveal it to voters. Note that since $\overline{u}_i(\theta_i)$, $\underline{u}_i(\theta_i)$ and $p_i(\theta_i)$ are all linear in θ_i, Theorem 1 implies that the optimal information revelation depends on the concavity or convexity of $S(\theta)$ function.

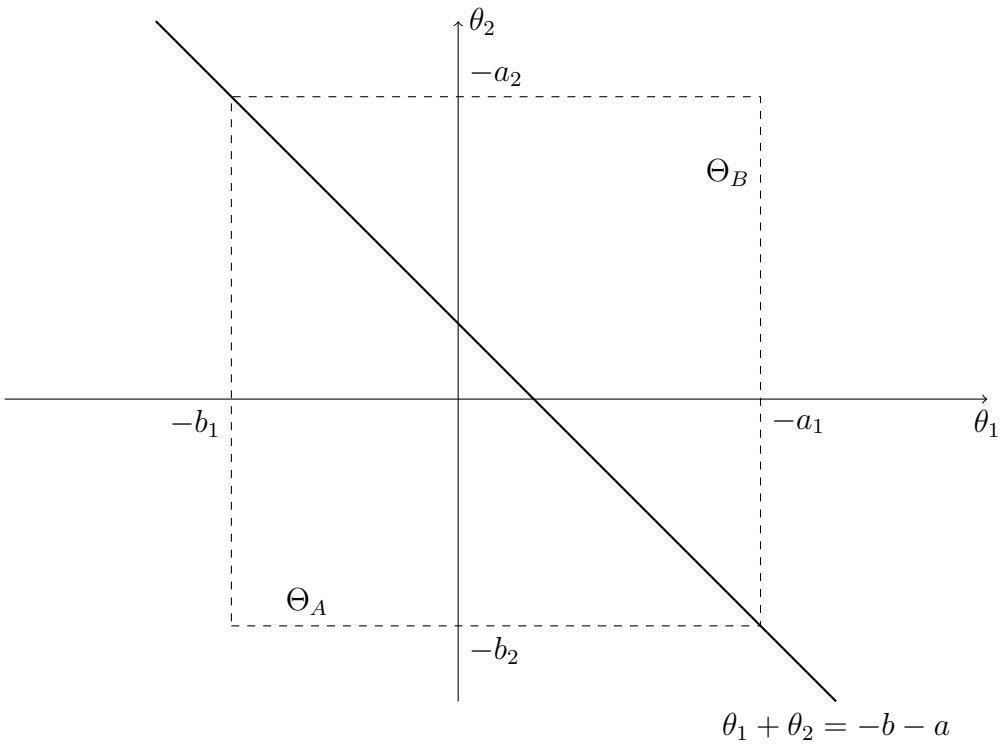


Figure 1: Optimal Voting Rule for Different Realizations of θ

Theorem 4. It is not optimal for social welfare if the designer reveals full information.

Proof. We prove this result by showing that there is a deviation from full information disclosure which strictly improves social welfare. In particular, under the subset $\{\theta|\theta \in \Theta_A, \theta_1+\theta_2 = \hat{\theta}\}$, the ex ante utility is,

$$\begin{aligned} S(\theta) &= p_1(\theta_1)p_2(\theta_2)(\overline{u}_1(\theta_1) + \overline{u}_2(\theta_2)) \\ &= \frac{b+\theta_1}{b-a}\frac{b+\theta_2}{b-a}\left(\frac{b+\theta_1}{2} + \frac{b+\theta_2}{2}\right) \\ &= \frac{(b+\theta_1)(b+\hat{\theta}-\theta_1)(b+\hat{\theta})}{(b-a)^2} \end{aligned} \tag{6.5}$$

Clearly, $S(\theta_1, \hat{\theta}-\theta_1)$ is concave in θ_1. As a result, the designer should not reveal additional information once the aggregate utility $\hat{\theta}$ is revealed to voters. The argument when $\theta \in \Theta_B$ is similar so the maths is omitted here. Hence, full information revelation is not optimal because an alternative information policy, which only reveals $\theta_1 + \theta_2$, performs strictly better.

□

In above proof we have shown that the designer should provide no additional information to voters once the aggregate utility $\hat{\theta} = \theta_1 + \theta_2$ is revealed. This means that the social welfare could be improved if the designer manipulates information to reduce heterogeneity in voters' preferences. The intuition is as follows. As a consequence of optimal voting rule, the reform, whenever it is chosen, always gives higher utilities in expectation compared to status quo. Hence, for a given aggregate benefit $\hat{\theta}$, it is optimal to maximize the aggregate probability that the reform is chosen. So asymmetry across voters is bad since it makes approval harder in average.

To see this, consider an example where $u_i \sim U[-1, 1]$ and $\theta = (\theta_1, \theta_2)$ takes two values, either $(-1, 1/2)$ or $(1/2, -1)$. Under full information revelation, the reform will be chosen with probability 0 because optimal voting requires both voters' approval but the voter with $\theta_i = -1$ always rejects. However, the reform can indeed be a better outcome in average. So the voters are better off if the reform will be chosen with a positive probability, namely, when the designer does not disclose the realized value of θ.

Since we have shown that the designer should not reveal individual utility once the aggregate utility $\hat{\theta} = \theta_1 + \theta_2$ is revealed, the reader may now wonder should the designer reveals every $\hat{\theta}$? The answer is yes. To see this, we are going to show that the social welfare is convex in $\hat{\theta}$.

First, note that with the assumption that θ_1 and θ_2 are independent and are uniformly distributed, the expected θ given the sum of θ_1 and θ_2 is,

$$\mathbf{E}[\theta|\theta_1 + \theta_2 = \hat{\theta}] = (\frac{\hat{\theta}}{2}, \frac{\hat{\theta}}{2}) \tag{6.6}$$

In fact, it is easy to check that these expected values are also local maximizers:[7]

$$\mathbf{E}[\theta|\theta_1 + \theta_2 = \hat{\theta}] = \underset{\{\theta|\theta_1+\theta_2=\hat{\theta}\}}{\arg\max} \; S(\theta) \tag{6.7}$$

Let us define the welfare function $\widetilde{S}(\hat{\theta})$ as,

$$\widetilde{S}(\hat{\theta}) = S(\mathbf{E}[\theta|\theta_1 + \theta_2 = \hat{\theta}])$$

So

$$\widetilde{S}(\hat{\theta}) = \begin{cases} \frac{(2b+\hat{\theta})^3}{8(b-a)^2} & if \quad -2b \leq \hat{\theta} \leq -b-a \\ a+b+\hat{\theta} - \frac{(2a+\hat{\theta})^3}{8(b-a)^2} & if \quad -b-a < \hat{\theta} \leq -2a \end{cases} \tag{6.8}$$

[7]This is due to the assumption that θ_i is uniformly distributed. Otherwise, conditional on $\theta_1 + \theta_2$, full pooling on θ is still optimal but $\mathbf{E}[\theta|\theta_1 + \theta_2]$ is typically not a maximizer.

It is easy to check the function $\widetilde{S}(\cdot)$ is piece-wisely convex. Also, at thresholds $\hat{\theta} = -b - a$ we have

$$\widetilde{S}'(-b - a^-) = \frac{3}{8} < 1 - \frac{3}{8} = \widetilde{S}'(-b - a^+) \tag{6.9}$$

So $\widetilde{S}(\cdot)$ is globally convex. This means that the designer should reveals full information regarding to $\hat{\theta}$. Without a formal proof, our above analysis suggests a candidate of optimal information revelation in the voting example: the designer fully reveals the aggregate utility $\theta_1 + \theta_2$ but reveals no additional information on individual utilities. This particular revelation strategy is a reasonable conjecture given the property of $S(\theta)$ function in this voting example, which is strictly concave on one dimension while strictly convex on the other dimension.

6.1 Remarks on Assumptions

1. u_i follows uniform distribution is crucial in our setting. Otherwise $\overline{u}_i(\theta_i)$, $\underline{u}_i(\theta_i)$ and $p_i(\theta_i)$ are not necessary to be linear in θ. And Theorem 1 is no longer applicable.

2. From the proof we can see that no revelation given $\theta_1 + \theta_2 = \hat{\theta}$ is quite robust. However, full revelation on $\hat{\theta}$ depends on the condition

$$\mathbf{E}[\theta|\theta_1 + \theta_2 = \hat{\theta}] = \underset{\{\theta|\theta_1+\theta_2=\hat{\theta}\}}{\arg\max} \; S(\theta)$$

   If this condition is violated we could not use Equation **??** and many things could happen.

   (a) For example, when θ_1 and θ_2 are very negatively correlated. Revealing the sum of θ_1 and θ_2 can itself generate asymmetry across voters. So no revelation on $\hat{\theta}$ could be optimal.

   (b) Is independence of θ_1 and θ_2 sufficient for the main result?

3. In general, if $b_1 - a_1 \neq b_2 - a_2$, I do not have much idea that what distributions of θ would work except uniform distribution. However, if we restrict $b_1 - a_1 = b_2 - a_2$, any pair of distributions would work as long as $\mathbf{E}[\theta|\theta_1 + \theta_2 = \hat{\theta}]$ stays on the diagonal of the square support. For example, if θ_1 and θ_2 have "identical" distribution subject to a parallel shift, i.e. $\phi_1(x) \equiv \phi_2(x + b_1 - b_2)$.

7 Concluding remarks

We conclude the paper by discussing two modelling assumptions of the paper.

7.1 Mechanisms contingent on the principal's ex post messages

From Section 3 on, we have assumed that a mechanism depends only on the principal's ex ante message m, and the agents report v. If m fully disclose the principal's information θ, of course it would be without loss of generality. However, in case m is not fully revealing, the principal may prefer to affect the allocation by sending another message, perhaps simultaneously with the agents or even after that.

Although we excluded such a possibility for simplicity, the assumption is crucial for Theorem 1, and we use Theorem 1 for Theorem 2 and 3. Thus, it would be useful to discuss how the results would be affected if we allow for such ex post messages of the principal.

It turns out that, for the two applications we consider in this paper, the conclusions do not qualitatively change. To see this, we first consider the auction problem as in Section 4.

To simplify the notation, here we only consider the case of revenue maximization. However, the similar result holds even if the principal's objective has a more general objective. In the following, we first solve for the optimal mechanism assuming that the principal can commit to her reporting even in the ex post stage. By the *inscrutability principle* of Myerson (1983), the maximum revenue is achieved by the following strategy. For the ex ante disclosure, she does not announce any informative signal, and hence, the agents play the mechanism in the second stage having the prior F_Θ over Θ. For the ex post reporting (which each agent would not observe before his reporting), the principal is truthful. However, we show that this maximum expected revenue is precisely the same level as under ex ante full disclosure that we have discusses in Section 4. This implies that ex ante full disclosure is at least one of the optimal strategies. We discuss later when the principal may in reality prefer ex ante disclosure rather than the other optimal strategies.

For each i and v, let $x_i(v) = E_\theta(p_i(v,\theta))$, $y_i(v) = E_\theta(\theta_i p_i(v,\theta))$, and $z_i(v) = E_\theta(t_i(v,\theta))$. The principal's objective is then

$$\int_v \int_\theta \sum_i t_i(v,\theta) dF_\Theta dF_V = \int_v \sum_i z_i(v) dF_V,$$

where F_V is the (joint) distribution over V, and F_Θ is over Θ.

Each i's IC and IR are given by

$$\begin{aligned} v_i x_i(v) + y_i(v) - z_i(v) &\geq v_i x_i(v_i', v_{-i}) + y_i(v_i', v_{-i}) - z_i(v_i', v_{-i}), \\ v_i x_i(v) + y_i(v) - z_i(v) &\geq 0, \end{aligned}$$

which implies, by the envelope condition (and setting the lowest value's expected utility to be zero),

$$v_i x_i(v) + y_i(v) - z_i(v) = \int_0^{v_i} x_i(\tilde{v}_i, v_{-i}) d\tilde{v}_i.$$

Thus, the principal's problem is, if we ignore the monotonicity constraints, to maximize

$$\begin{aligned}
\int_v \sum_i z_i(v) dF_V &= \int_v \sum_i [v_i x_i(v) + y_i(v) - \int_0^{v_i} x_i(\tilde{v}_i, v_{-i}) d\tilde{v}_i] dF_V \\
&= \int_v \sum_i [x_i(v)\phi_i(v_i) + y_i(v)] dF_V \\
&= \int_v \int_\theta \sum_i [p_i(v,\theta)(c_i(v_i) + \theta_i)] dF_\Theta dF_V.
\end{aligned}$$

where $c_i(v_i) = v_i - \frac{1-F_i(v_i)}{f_i(v_i)}$ is the virtual value of v_i, and does not depend on θ if v, θ are independent.

Therefore, the optimal mechanism assigns the object to an agent whose "adjusted virtual surplus" $c_i(v_i) + \theta_i$ is highest whenever that is positive (if no agent has a positive adjusted virtual surplus, then the principal keeps the object).

Notice that this allocation rule is achieved by running Myerson's optimal auction *after* disclosing θ to the agents in the ex ante stage. Thus, the ex ante full disclosure is one of the optimal alternatives. Moreover, if the principal cannot commit to her ex post reporting, then she always have an incentive to misreport high θ in this mechanism.

Given that there are multiple optimal strategies, a natural question would be which ones may be more reasonable than others. One argument that prefers ex ante full disclosure may be the following. As we have discussed before, one instance where the principal may be able to commit to her ex ante disclosure is when she has some evidence. Our model represents a situation where the principal can costlessly show the associated evidence with ex ante announcement, while showing a false evidence is impossible.

In addition to it, imagine a situation where this evidence is costly for a court to verify it. Such an assumption would be reasonable if it requires some expertise to correctly interpret the evidence (and the principal and each agent has such expertise, while the court does not). Then, if the principal has an opportunity to report her message in the ex post stage, she may have an incentive to misreport θ (without the associated evidence). This means that, as opposed to ex ante disclosure where the principal can convey information to the agents without frictions, ex post report may need to satisfy the principal's incentive compatibility.

The analysis above implies that, when it is even infinitesimally costly for verification by a court, the principal may prefer to commit herself to reveal all the information ex ante.

Now we consider the bilateral trade problem. In this case, it is easy to see that our qualitative result, suboptimality of full disclosure, is not affected even if we add the ex post revelation of the principal. This is because, if we allow for such an additional opportunity for the principal, of course she can do at least weakly better. Because full ex ante disclosure is suboptimal before, it continues to be suboptimal.

7.2 Linearity

Although the linearity assumption imposed in this paper is restrictive, it may be worth mentioning that we can relax the assumption at least to some extent. For example, for revenue maximization in auction, instead of having $\tilde{v}_1 = v_1 + \theta$, we can have $\tilde{v}_1 = v_1 + \phi(\theta)$, where $v_1 \in [0, 1]$ and $\phi(\theta)$ is a function of θ. In this case, we can define $\tilde{\theta} = \phi(\theta)$ and treat $\tilde{\theta}$ as the principal's information. Then the model is linear in $\tilde{\theta}$ (rather than in θ).

References

Alonso, R. and O. Câmara (2014): "Persuading Voters," Working paper.

Angeletos, G. and A. Pavan (2007): "Efficient Use of Information and Social Value of Information," *Econometrica*, 75, 1103–1142.

Azrieli, Y. and S. Kim (2013): "Pareto efficiency and weighted majority rules," *International Economic Review*, forthcoming.

Bergemann, D. and M. Pesendorfer (2007): "Information structures in optimal auctions," *Journal of Economic Theory*, 137, 580–609.

Courty, P. and H. Li (2000): "Sequential Screening," *Review of Economic Studies*, 67, 697–717.

Eső, P. and B. Szentes (2007): "Optimal Information Disclosure in Auctions and the Handicap Auction," *Review of Economic Studies*, 74, 705–731.

Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

Kaplan, T. R. and S. Zamir (2002): "The Strategic Use of Seller Information in Private-Value Auctions," Working paper.

Landsberger, M., J. Rubinstein, E. Wolfstetter, and S. Zamir (2001): "First-price auctions when the ranking of valuations is common knowledge," *Review of Economic Design*, 6, 461–480.

Li, M., T. Mylovanov, and A. Zapechelnyuk (2014): "Information Monopoly," Working paper.

Milgrom, P. and R. J. Weber (1982): "The Value of Information in A Sealed-Bid Auction," *Journal of Mathematical Economics*, 10, 105–114.

Morris, S. and H. S. Shin (2002): "Social Value of Public Information," *American Economic Review*, 92, 1521–1534.

Myerson, R. B. (1981): "Optimal Auction Design," *Mathematics of Operation Research*, 6, 58–73.

——— (1983): "Mechanism Design by An Informed Principal," *Econometrica*, 51, 1767–1797.

Myerson, R. B. and M. A. Satterthwaite (1983): "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, 29, 265–281.

Skreta, V. (2011): "On the informed seller problem: optimal information disclosure," *Review of Economic Design*, 15, 1–36.

Belief Convergence and Divergence with Rationally Inattentive Learning

Tong Su*

September 24, 2015

Abstract

In standard Bayesian models, agents' heterogeneous beliefs will always converge upon arrival of new information, which is not consistent with many daily observations. In this paper I show that belief divergence may occur if agents' learning is rationally inattentive. When attention is costly, agents optimally choose to acquire potentially new information which they believe most likely to come, leading to a conformism learning. Hence, agents whose initial beliefs are far from the truth will react less often compared to agents whose beliefs are closer to the truth, leading to a divergence in agents' beliefs in expectation. I characterize the condition for belief divergence and show that it is more likely to happen when the truth is more extreme and the attention cost is moderate.

1 Introduction

Standard Bayesian models predict that even though agents hold heterogenous beliefs ex ante, they will asymptotically agree with the arrival of common new information, as their postier beliefs put more and more weight in new information and their prior beliefs just "wash out". However, in reality this prediction seems not always true. For example, in 17th century, when pioneer astronomists established the theory of Heliocentrism based on new scientific findings, instead of being accepted by all people, the new theory brought huge controversy to the society. People's beliefs on the "right model of the universe" diverged but not converge with those new findings. In financial market, traders' perspectives can also diverge when they observe a common public announcement by the central bank.

*Toulouse School of Economics. email:tong.su@tse-fr.eu I am deeply grateful to my supervisor Christian Hellwig, for his generous guidance and encouragement for this project. I also thank Jacques Crémer, Georgy Lukyanov and Takuro Yamashita, for helpful comments.

In this paper, I resolve the "disagreement puzzle" by introducing a rational inattention framework, where agents are still rational and Bayesian but learn new information subject to an attention cost. In particular, before agents observe new information and update their beliefs accordingly, they must decide what they want to observe, at a cost, in the first place. The more new information they choose to observe, the more attention cost they have to pay. The standard Bayesian learning corresponds to my model when all agents learn with costless full attention.

When attention cost exists and under certain model specifications, I show that rational agents learn with conformism, meaning that agents optimally choose to observe new information that is more likely to confirm their priors. The intuition behind this result is that in my model the value "per new information" is constant due to quadratic loss utilities and Gaussian signal structure, hence an agent chooses to observe new information that is most relevant, which is exactly signal realizations that most consistant with his prior belief.

As a result of conformism learning, agents with different prior beliefs who face the same information source generally learn differently in expectation as their "observation windows" do not coincide. If the truth lies in between of two agents' prior beliefs, we always have beliefs convergence as the "left" agent in expectation updates his belief towards right and the "right" agent in expectation updates towards left. However, things are different if the truth lies at, for example the right, against both agents' beliefs. Compare to the "far left" agent, the "left" agent updates his belief more frequently because his observe window is more close to the truth, which I call the *inattention effect*, but at a slower pace as his belief is already close to the truth, with I call the *adjusting effect*. In this case, as both agents' beliefs update towards right in expectation, belief divergence, meaning that the "left" agent updates faster than the "far left" agent, can happen if the inattention effect dominates the adjusting effect.

The rest paper organizes as follows. Section 2 presents the one-agent version of my inattentive learning model and solve agents' optimal learning problem. Section 3 introduces belief heterogeneity and characterizes the condition for belief convergence/divergence. Section 4 discusses my model on several aspects and compares it to the literature. Section 5 concludes the paper.

2 A Learning Model with Attention Cost

We first consider an one-agent version of the inattentive learning model. There is an unknown parameter θ, which I refer as "the truth", is distributed from a given support $\theta \in R$. The agent believes that θ follows a cumulative density $F(\theta)$, which I call the agent's prior

belief. The prior belief takes Gaussian form with mean μ and variance β^{-1}.

The agent will take an action a in the end of the game and his utility takes the quadratic loss form,

$$u(a, \theta) = -(a - \theta)^2 \tag{2.1}$$

There is a Gaussian signal S reveals information on the truth θ. In particular, I assume that the signal realization $s \in S$ reveals θ perfectly expect a white noise,

$$s = \theta + \varepsilon \quad \text{and} \quad \varepsilon \sim \mathcal{N}(0, \alpha^{-1}) \tag{2.2}$$

Before the agent takes action, he can update his belief on θ by learning the signal S with an attention cost. Agent's attention cost is proportional to his observation window. The larger range of signal realization he wants to observe the higher attention he has to pay. I also allow the agent to learn signal by random. Formally, the agent's learning strategy is described by a function $\lambda(s)$, that the agent will observe the realized signal,

$$\mathcal{I} = \begin{cases} s & \text{with probability} \quad \lambda(s) \in [0, 1] \\ \varnothing & o.w. \end{cases} \tag{2.3}$$

And the attention cost is given by $c \cdot \kappa = c \cdot \int \lambda(s) ds$, where κ describes how much attention the agent devotes on learning new information and c is the cost per unit attention.

Now I solve agent's optimal learning problem with attention cost. Due to the quadratic loss utility function, conditional on agent's information $\mathcal{I} \in S \cup \{\varnothing\}$, his optimal action and utility are given by,

$$a^*(\mathcal{I}) = \mathbf{E}[\theta|\mathcal{I}] \quad \text{and} \quad v(\mathcal{I}) = \mathbf{E}[u(a^*, \theta)|\mathcal{I}] = -\text{Var}[\theta|\mathcal{I}] \tag{2.4}$$

Hence agent's utility with and without an informative observation are given by,[1]

$$v(s) = -(\alpha + \beta)^{-1} \quad \text{for } \forall s \in S \quad \text{and} \quad v(\varnothing) = -\beta^{-1} \tag{2.5}$$

Proposition 1. *Agent's optimal attention is given by,*

$$\lambda^*(s) = \begin{cases} 1 & when \quad s \in [\mu - \kappa^*/2, \mu + \kappa^*/2] \\ 0 & o.w. \end{cases} \tag{2.6}$$

[1]Here I assume that agent is naive and do not update his belief without an informative observation. I will discuss this assumption in Section 4.

where κ^* is the optimal attention that is decreasing in the unit cost of attention c, and

$$\kappa^* \begin{cases} > 0 & when \quad c < \frac{\beta^{-1}-(\alpha+\beta)^{-1}}{\sqrt{2\pi(\alpha^{-1}+\beta^{-1})}} \\ = 0 & o.w. \end{cases} \tag{2.7}$$

where κ^*, if not 0, is pinned down by $\frac{\beta^{-1}-(\alpha+\beta)^{-1}}{\sqrt{\beta^{-1}+\alpha^{-1}}}\phi(\frac{\kappa}{2\sqrt{\beta^{-1}+\alpha^{-1}}}) = c$.

Proof. In the Appendix □

Proposition 1 describes the amount of agent's optimal attention as well as its allocation. It says that the agent should put full attention on observing certain signal realizations while spend no attention at all on other realizations. And when attention cost, c, is too high, the agent will optimally choose not to learn at all and take action only according to his prior belief.

More importantly, this optimal attention strategy results in *conformism* learning, that the agent will ultimately learn new information that is more likely to confirm his prior belief. The intuition behind this result is that, due to quadratic lose utility function, the extra value of an informative observation is constant. Hence, agent allocates his attention only on signal realizations that he believes are most relevant, which exactly are those more close to his prior mean of θ.

2.1 Inattentive Learning in Multiple Periods

In this subsection, I generalize the inattentive learning model into a simple dynamic version. Consider that the agent lives multiple periods and in each period $t = 1, 2, \ldots,$ he learns from a signal S_t and takes action a_t. Suppose signals in different periods are conditionally independent,

$$s_t = \theta + \varepsilon_t \quad \text{and} \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \alpha^{-1}) \tag{2.8}$$

The prior belief of the agent before period t is given by a Normal distribution with mean μ_{t-1} and variance β_{t-1}, which will be updated, with his learning in S_t, into his posterior belief that characterized with (μ_t, β_t). The belief dynamic is given by,

$$(\mu_t, \beta_t) = \begin{cases} (\frac{\alpha s_t + \beta_{t-1}\mu_{t-1}}{\alpha+\beta_{t-1}}, \alpha + \beta_{t-1}) & \text{when} \quad s_t \in [\mu_{t-1} - \kappa_t^*/2, \mu_{t-1} + \kappa_t^*/2] \\ (\mu_{t-1}, \beta_{t-1}) & o.w. \end{cases} \tag{2.9}$$

Agent's optimal attention κ_t^*, if not equals to 0, is pinned down by,

$$\underbrace{\left(\beta_{t-1}^{-1} - (\alpha + \beta_{t-1})^{-1}\right)}_{\text{marginal benefit per observation}} \cdot \underbrace{\frac{1}{\sqrt{\beta_{t-1}^{-1} + \alpha^{-1}}} \phi\left(\frac{\kappa_t^*}{2\sqrt{\beta_{t-1}^{-1} + \alpha^{-1}}}\right)}_{\text{probability of marginal observation}} = c \tag{2.10}$$

where $\phi(\cdot)$ is p.d.f of standard normal distribution.

Proposition 2. *Agent gradually reduces his attention over time and eventually stops learning.*

Proof. By Equation 2.9, we know that $\frac{\beta_{t-1}^{-1} - (\alpha+\beta_{t-1})^{-1}}{\sqrt{\beta_{t-1}^{-1}+\alpha^{-1}}}$ is weakly decreasing in t and $\sqrt{\beta_{t-1}^{-1} + \alpha^{-1}}$ is weakly increasing in t. So we have agent's optimal attention κ_t is weakly decreasing in t. In the limit, the agent stops learning because,

$$\frac{\beta_{t-1}^{-1} - (\alpha + \beta_{t-1})^{-1}}{\sqrt{\beta_{t-1}^{-1} + \alpha^{-1}}} \to 0 \quad \text{as} \quad t \to +\infty \tag{2.11}$$

□

3 Belief Convergence and Divergence

Now I extend the model in Section 2 with two agents with different ex ante priors and study belief convergence and divergence as a result of inattentive learning. For the sake of simplicity, here I assume the model has only 1 period. Agent i's ($i \in \{a, b\}$) prior belief on θ is given by a normal distribution with mean $\mu_{i,0}$ and variance $\beta_{i,0}^{-1}$. We assume that two agents have the same *confidence* on their prior belief, that is $\beta_{a,0} = \beta_{b,0} = \beta$. Without loss of generality, we assume that $\mu_{a,0} > \mu_{b,0}$.

Since two agents are only different in their prior expectation on θ, by Proposition 1, their attention is the same, $\kappa_a^* = \kappa_b^* = \kappa^*$. Call agent i's *observation window* $W_i = [\mu_{i,0} - \kappa^*/2, \mu_{i,0} + \kappa^*/2]$. In general two agents' observation windows do not coincide. Even though two agents spend the same effort on learning new information, due to their different prior estimation on θ, they allocate their attention differently because they disagree on what signal realizations are more relevant.

I call the *disagreement* as the difference between two agents' expected value of θ. The ex ante disagreement is $\Delta_0 = \mu_{a,0} - \mu_{b,0}$ and the ex post disagreement is $\Delta_1 = \mu_{a,1} - \mu_{b,1}$. I define belief convergence and divergence in the following way,

Definition 1. Suppose an outsider knows the truth θ and the initial disagreement Δ_0 and

predicts Δ_1, we call (in expectation) beliefs *converge*, if $\mathbf{E}[\Delta_1|\theta] < \Delta_0$; and beliefs *diverge*, if $\mathbf{E}[\Delta_1|\theta] > \Delta_0$.

Before drawing conclusions about belief convergence/divergence, I first show that the agents' expected posterior mean $\mu_{i,1}$ is a weighted average between his prior mean $\mu_{i,0}$ and his expected observation $\mathbf{E}[s_i|s_i \in W_i, \theta]$. Knowing θ, the outsider predicts agent i's posterior mean as follows,

$$\begin{aligned}\mathbf{E}[\mu_{i,1}|\theta] &= \int_{s\in W_i} \frac{\alpha s + \beta\mu_{i,0}}{\alpha+\beta} f(s|\theta)ds + (1 - \text{Prob}\{s \in W_i, \theta\})\mu_{i,0} \\ &= p_i(1-q)\mu_{i,0} + p_i q\mathbf{E}[s|s \in W_i, \theta] + (1-p_i)\mu_{i,0} \\ &= (1 - p_i q)\mu_{i,0} + p_i q\mathbf{E}[s|s \in W_i, \theta]\end{aligned} \tag{3.1}$$

where p_i is the probability that agent i makes an informative observation and $q \equiv \frac{\alpha}{\alpha+\beta}$ is the weight that agents put on new information when they update their beliefs.

Hence, the expected change in agent i's belief can be decomposed as follows,

$$\mathbf{E}[\mu_{i,1} - \mu_{i,0}|\theta] = qp_i(\mathbf{E}[s|s \in W_i, \theta] - \mu_{i,0}) \tag{3.2}$$

From the above equation, there are two things that affect agent i's belief dynamic. First, it depends on how much the agent will update his belief conditional on he makes an informative observation. Second, it also depends on how often he can make an informative observation. Note that in a benchmark model where agents have no attention costs $c = 0$, agents always make informative observation so that $p_i \equiv 1$. And since the signal S is unbiased and agents' observation window is full set, $\mathbf{E}[s] \equiv \theta$. Hence the belief adjustment in expectation is $q(\theta - \mu_{i,0})$.

By Equation 3.2, the expected change in agents' belief disagreement can be also decomposed as follows,

$$\begin{aligned}\mathbf{E}[\Delta_1 - \Delta_0|\theta] &= \mathbf{E}[(\mu_{a,1} - \mu_{b,1}) - (\mu_{a,0} - \mu_{b,0})|\theta] \\ &= \mathbf{E}[\mu_{a,1} - \mu_{a,0}|\theta] - \mathbf{E}[\mu_{b,1} - \mu_{b,0}|\theta] \\ &= q\big(p_a(\mathbf{E}[s|s \in W_a, \theta] - \mu_{a,0}) - p_b(\mathbf{E}[s|s \in W_b, \theta] - \mu_{b,0})\big)\end{aligned} \tag{3.3}$$

Similar to the analysis of single agent's belief dynamic, there are also two effects present in the disagreement dynamic. The expected change in belief disagreement depends firstly on the relative change in agents' beliefs when they make informative observations. And secondly,

it also depends on the relative frequency that agents can make informative observation. I call the first effect as the *adjusting effect* and the second effect as the *inattention effect*. Roughly speaking, an agent whose initial belief is more close to the truth will make informative observation more often, however, will adjust his belief more slowly, compared to an agent who is learning from a initial belief that is more far from the truth.

Again we can use the benchmark model, where $c = 0$, to illustrate the interaction between the adjusting effect and the inattention effect. When agents' attention is costless, both agents always make informative observation, so inattention effect plays no role in the benchmark model. Hence, agents' beliefs always converge due to the adjusting effect.

Proposition 3. *In the benchmark model where agents have costless attention, $c = 0$, agents' beliefs always converge in expectation.*

Proof. When attention is costless, $p_a = p_b = 1$. By Equation 3.3, we have for $\forall\theta$,

$$\mathbf{E}[\Delta_1 - \Delta_0|\theta] = q\big((\theta - \mu_{a,0}) - (\theta - \mu_{b,0})\big) = q(\mu_{b,0} - \mu_{a,0}) < 0 \tag{3.4}$$

□

In the general model, agents' belief can either converge or diverge depending on which of the two effects dominates, which is formally shown by the following main result of this paper.

Proposition 4. *When agents have costly attention, there exists a pair of thresholds $\overline{\theta}$ and $\underline{\theta}$ with $\underline{\theta} < \mu_{b,0} < \mu_{a,0} < \overline{\theta}$, such that agents' beliefs converge if the truth is moderate, $\theta \in (\underline{\theta}, \overline{\theta})$; and diverge if the truth is extreme, $\theta > \overline{\theta}$ or $\theta < \underline{\theta}$.*

Proof. In the Appendix □

Now I provide two types of learning to illustrate the intuition behind above result. The first type, which I call *consensus learning*, describes a situation that the truth lies between agents initial beliefs, $\mu_{b,0} < \theta < \mu_{a,0}$. For example, when traders do their research make predictions on a firm's long-term stock price, in most cases the true answer lies in between those predictions among the most optimistic trader and the most pessimistic trader. The second type, which I call *pioneer learning*, describe a situation that the truth lies far beyond both agents initial guesses, $\mu_{b,0} < \mu_{a,0} << \theta$. When scientists make pioneering breakthroughs, like what Nicolaus Copernicus, Giordano Bruno and Galileo Galilei had done in the 16-17th century for astronomy, usually the newly established scientific frontier lies beyond what all people had believed. Note that there is no fine lines between examples for the two types of learning. All traders' predictions may fail when an unexpected economic recession comes, it is

then more like a pioneer learning. When a new scientific theory becomes publicly known, the debate between advocates for the old and the new theories look more like consensus learning.

When learning is consensus type, Proposition 4 says that agents' belief will converge. Even though in general two may learn new information with different frequency because their observation window are located differently, they will always adjust their beliefs towards the truth whenever they make informative observation. Since the truth lies in between their beliefs, in expectation their posterior beliefs will become more closers and hence their disagreement converges.

A more interesting case is pioneer learning, that agents are learning from a radical truth that lies beyond both their prior belief. Figure 1 is given below for illustration purpose.

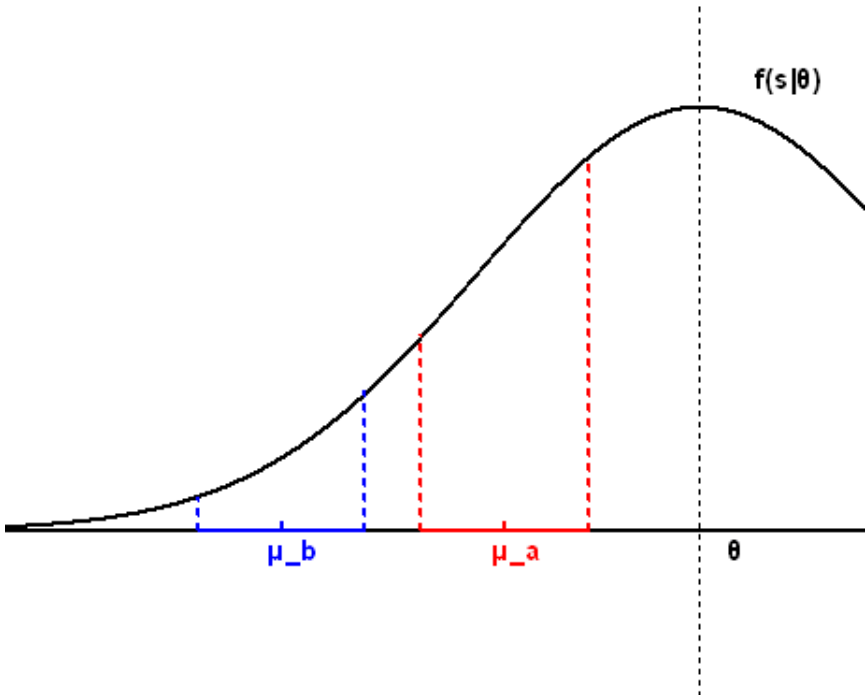


Figure 1: An Illustration of Pioneer Learning

In Figure 1, the black bell-curve, surrounding θ, is the probability density for signal realizations. The red and blue zones are observation windows for agent a and agent b respectively. The size of the area that one agent's observation zone intersects the density function represents the probability that the agent will make an informative observation. Due to conformism, both agents choose to learn new information that is close to their prior knowledge. Since the truth lies on the right of both agents' prior means, agent b, or the *far-left* agent will learn new information less frequently compared to agent a, the *left* agent, as captured in the inattention effect. However, conditional on both agents have informative observation, the far-left agent will adjust his belief towards the truth faster because the conditional probability density

function for the signal is more right-skewed, compared to the left agent, as captured in the adjusting effect. Hence, belief divergence will occur if the inattention effect dominates the adjusting effect, which is more likely to happen if θ is more extreme.

4 Discussion

4.1 Related Literature

The idea that agents who hold different ideas will eventually agree with each other if they continuously learn new information was first introduce by Blackwell and Dubins (1962). There is a small literature, among others Acemoglu et al. (2015), Cripps et al. (2008) and Miller and Sanchirico (1999), that studies under what circumstance that agents fail to agree in the long-run with common information. The main approach of those papers is to show disagreement can be persistence for some special types of public signals. Instead, in my paper the public signal is standard Gaussian, and the disagreement is driven by agents' rational inattention. My paper shares more spirits with Schwartzstein (2014), in which a Bayesian agent faces sequential new information, but can not remember all of them due to his memory capacity, leading to possible failure in making the right prediction in the long term.

My paper also belongs to the literature on rational inattention and endogenous information acquisition. The pioneering paper in this field is Sims (2003), followed by many other papers that uses the framework of rational inattention to study how agents learn information flexibly. For example, Yang (2015) studies security design problem when investors' attention is limited. My model, though not uses the exact setting as in Sims (2003) or Yang (2015), shares the same intuition that agents should endogenize what they learn by considering their attention costs.

The concept that agents may not agree more with each other with new common information has also be studied in applied theories, including politics (Dixit and Weibull (2007)), medias (Baron (2006)), trade (Banerjee and Kremer (2010)) and contract theory (Adrian and Westerfield (2009)).

4.2 The Role of Attention Cost c

This paper shows that the dynamic of belief disagreement changes qualitatively if agents learn new information with attention cost. One interesting question is that how does belief convergence/divergence is affected by the parameter of the cost of attention, namely the c in my model. Without formal mathematics, we can see belief divergence is more likely to

happen for moderate c. Because when c is small, agents are learning nearly with full attention, making belief divergence never possible; and when c is large, ultimately agents stops learning and belief disagreement becomes static.

This result actually captures the fact that people do not necessarily agree more often when they are more capable of learning, for example as a consequence of the technology progress. One vivid example for the policy implication of this result is that, in the late 1980s when the former USSR leader Mikhail Gorbachev's liberalization reform granted people more information, such as removing censorships and promoting free press, the society on the contrary became less stable.

4.3 Agent's Response on Uninformative Observation

In my model, I assumed that whenever an agent makes an uninformative observation, he does not update his belief. In other words, the agents are naive and can not use Bayesian inference to learn from uninformative message given their learning strategy. I use this assumption mainly for the sake of tractability, which is soon disappear when agents become sophisticated. This naiveness assumption is also used in recent papers that study inattentive learning problem, such as Schwartzstein (2014) and Mullainathan (2002).

However, the main result should not qualitatively depend on the naiveness assumption. Since the main driving force for belief disagreement is inattention effect, as long as agents with different priors use the same (sophisticated) strategy to determine their observation windows, the logic behind Proposition 4 does not change.

5 Conclusion

In this paper I present a simple rational inattention framework to show that the heterogenous beliefs held by different agent can fail to converge even though all agents are learning from a common information source. The intuition behind this result is as follows. Since agents' attentions are costly, they only pay attention to new information that they believe most likely to occur. As a result, those agents whose initial belief is further from the truth will learn less frequently, compare to agents whose initial belief is more close to the truth, even though they adjust their beliefs faster once they have learnt from the new information.

In general, the dynamic of belief agreement is driven by two effects: adjusting effect and inattention effect. Belief divergence occurs if the the second effect dominates. I also show that belief divergence is more likely to occur if the truth lies outside and far from all agents' initial

beliefs. This model can be used to explain phenomena such as scientific break-throughs in the history tend to raise controversy rather than consensus, and in financial market, dissent among market participant usually increase when the economy has experienced a big shock.

References

Acemoglu, D., V. Chernozhukov, and Yildiz (2015): "Fragility of Asymptotic Agreement under Bayesian Learning," *Theoretical Economics*, forthcoming.

Adrian, T. and M. M. Westerfield (2009): "Disagreement and Learning in a Dynamic Contracting Model," *Review of Financial Studies*, 22, 3873–3906.

Banerjee, S. and I. Kremer (2010): "Disagreement and Learning: Dynamic Patterns of Trade," *Journal of Finance*, 65, 1269–1302.

Baron, D. P. (2006): "Persistent Media Bias," *Journal of Public Economics*, 90, 1–36.

Blackwell, D. and L. Dubins (1962): "Merging of Opinions with Increasing Information," *The Annals of Mathematical Statistics*, 33, 882–886.

Cripps, M. W., J. C. Ely, G. J. Mailath, and L. Samuelson (2008): "Common Learning," *Econometrica*, 76, 909–933.

Dixit, A. and J. Weibull (2007): "Political Polarization," *Proceedings of the National Academy of Sciences*, 104, 7351–7356.

Miller, R. I. and C. W. Sanchirico (1999): "The Role of Absolute Continuity in "Merging of Opinions" and "Rational Learning"," *Games and Economic Behavior*, 29, 170–190.

Mullainathan, S. (2002): "A Memory-Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 117, 735–774.

Schwartzstein, J. (2014): "Selective Attention And Learning," *Journal of the European Economic Association*, 12, 1423–1452.

Sims, C. A. (2003): "Implications of Rational Inattention," *Journal of Monetary Economics*, 50, 665–690.

Yang, M. (2015): "Optimality of Debt under Flexible Information Acquisition," Working paper.

A Appendix

A.1 Proof of Proposition 1

The agent chooses his observation window W optimally to maximize his expected value before he actually learns the signal realization. Due to quadratic loss utility function, the agent's utility with an informative observation dose not depends on the exact value of signal realization that he observes. His expected utility is given by,

$$\begin{aligned} V(W) =& -(\alpha+\beta)^{-1}\int_{s\in W}\lambda(s)f(s|\mu)ds - \beta^{-1}\int_{s\in W}(1-\lambda(s))f(s|\mu)ds \\ & -\beta^{-1}\int_{s\notin W}f(s|\mu)ds - c\int_{s\in W}\lambda(s)ds \\ =&\beta^{-1}+(\beta^{-1}-(\alpha+\beta)^{-1})\int_{s\in W}\lambda(s)f(s|\mu)ds - c\int_{s\in W}\lambda(s)ds \\ =&\beta^{-1}+(\beta^{-1}-(\alpha+\beta)^{-1})p - c\kappa \end{aligned} \tag{A.1}$$

From the above equation, given a fixed attention $\kappa = \int_{s\in W} ds$, the agent should maximize the total probability p that he makes informative observations. In particular, since the conditional density function for signal realizations $f(s|\mu)$ is a Normal distribution with mean μ, the highest p is obtained when,

$$\lambda(s) = \begin{cases} 1 & \text{when} \quad s\in[\mu-\kappa^*/2, \mu+\kappa^*/2] \\ 0 & \text{o.w.} \end{cases} \tag{A.2}$$

Now taking derivative with respect to κ for $V(W)$ gives the FOC on the optimal attention,

$$\frac{\beta^{-1}-(\alpha+\beta)^{-1}}{\sqrt{\beta^{-1}+\alpha^{-1}}}\phi\left(\frac{\kappa}{2\sqrt{\beta^{-1}+\alpha^{-1}}}\right) = c \tag{A.3}$$

Note that the agent should pay positive attention if and only if the maximal marginal utility from his attention exceeds the marginal cost, which is given by $\frac{\beta^{-1}-(\alpha+\beta)^{-1}}{\sqrt{2\pi(\beta^{-1}+\alpha^{-1})}} > c$

A.2 Proof of Proposition 4

Without loss of generality, we can assume that $\mu_{a,0} > \mu_{b,0}$ and $\theta > \mu_{b,0}$. Note that since $W_i = [\mu_{i,0}-\kappa/2, \mu_{i,0}+\kappa/2]$, we have $\mathbf{E}[s|s\in W_i,\theta]-\mu_{i,0} > 0$ if and only if $\theta > \mu_{i,0}$.

By Equation 3.3, agents beliefs converge if

$$p_a(\mathbf{E}[s|s \in W_a, \theta] - \mu_{a,0}) - p_b(\mathbf{E}[s|s \in W_b, \theta] - \mu_{b,0}) < 0 \tag{A.4}$$

or

$$\frac{p_a(\mathbf{E}[s|s \in W_a, \theta] - \mu_{a,0})}{p_b(\mathbf{E}[s|s \in W_b, \theta] - \mu_{b,0})} < 1 \tag{A.5}$$

Since p_a, p_b and $\mathbf{E}[s|s \in W_b, \theta] - \mu_{b,0} > 0$, the inequality in A.5 satisfies if $\mathbf{E}[s|s \in W_a, \theta] - \mu_{a,0} < 0$, or equivalently $\theta < \mu_{a,0}$. Now we need to show that $\exists \overline{\theta} > \mu_{a,0}$ that inequality in A.5 fails, that is to say agents beliefs diverge when the truth is extreme.

We first show that $\frac{\mathbf{E}[s|s \in W_a, \theta] - \mu_{a,0}}{\mathbf{E}[s|s \in W_b, \theta] - \mu_{b,0}}$ is increasing in θ but bounded below 1 for all $\theta > \mu_{a,0}$. The truncated mean is given by,

$$\mathbf{E}[s|\mu_{i,0} - \kappa/2 < s < \mu_{i,0} + \kappa/2] = \theta + \frac{\phi(\frac{\mu_{i,0} - \kappa/2 - \theta}{\alpha^{-1/2}}) - \phi(\frac{\mu_{i,0} + \kappa/2 - \theta}{\alpha^{-1/2}})}{\Phi(\frac{\mu_{i,0} + \kappa/2 - \theta}{\alpha^{-1/2}}) - \Phi(\frac{\mu_{i,0} - \kappa/2 - \theta}{\alpha^{-1/2}})} \tag{A.6}$$

So we have

$$\frac{\mathbf{E}[s|s \in W_a, \theta] - \mu_{a,0}}{\mathbf{E}[s|s \in W_b, \theta] - \mu_{b,0}} = \frac{\theta - \mu_{a,0} + \frac{\phi(\frac{\mu_{a,0} - \kappa/2 - \theta}{\alpha^{-1/2}}) - \phi(\frac{\mu_{a,0} + \kappa/2 - \theta}{\alpha^{-1/2}})}{\Phi(\frac{\mu_{a,0} + \kappa/2 - \theta}{\alpha^{-1/2}}) - \Phi(\frac{\mu_{a,0} - \kappa/2 - \theta}{\alpha^{-1/2}})}}{\theta - \mu_{b,0} + \frac{\phi(\frac{\mu_{b,0} - \kappa/2 - \theta}{\alpha^{-1/2}}) - \phi(\frac{\mu_{b,0} + \kappa/2 - \theta}{\alpha^{-1/2}})}{\Phi(\frac{\mu_{b,0} + \kappa/2 - \theta}{\alpha^{-1/2}}) - \Phi(\frac{\mu_{b,0} - \kappa/2 - \theta}{\alpha^{-1/2}})}} \tag{A.7}$$

Note that $\frac{\theta - \mu_{a,0}}{\theta - \mu_{b,0}}$ is increasing in θ. For the expression $\frac{\phi(\frac{\mu_{i,0} - \kappa/2 - \theta}{\alpha^{-1/2}}) - \phi(\frac{\mu_{i,0} + \kappa/2 - \theta}{\alpha^{-1/2}})}{\Phi(\frac{\mu_{i,0} + \kappa/2 - \theta}{\alpha^{-1/2}}) - \Phi(\frac{\mu_{i,0} - \kappa/2 - \theta}{\alpha^{-1/2}})}$, it is increasing in θ and decreasing in μ. So $\frac{\frac{\phi(\frac{\mu_{a,0} - \kappa/2 - \theta}{\alpha^{-1/2}}) - \phi(\frac{\mu_{a,0} + \kappa/2 - \theta}{\alpha^{-1/2}})}{\Phi(\frac{\mu_{a,0} + \kappa/2 - \theta}{\alpha^{-1/2}}) - \Phi(\frac{\mu_{a,0} - \kappa/2 - \theta}{\alpha^{-1/2}})}}{\frac{\phi(\frac{\mu_{b,0} - \kappa/2 - \theta}{\alpha^{-1/2}}) - \phi(\frac{\mu_{b,0} + \kappa/2 - \theta}{\alpha^{-1/2}})}{\Phi(\frac{\mu_{b,0} + \kappa/2 - \theta}{\alpha^{-1/2}}) - \Phi(\frac{\mu_{b,0} - \kappa/2 - \theta}{\alpha^{-1/2}})}}$ is also increasing in θ.

Also note that $\mathbf{E}[s|s \in W_a, \theta] - \mu_{a,0} < \mathbf{E}[s|s \in W_b, \theta] - \mu_{b,0}$ and as $\theta \to +\infty$, we have $\mathbf{E}[s|s \in W_a, \theta] - \mu_{a,0} = \mathbf{E}[s|s \in W_b, \theta] - \mu_{b,0} = \kappa/2$. Hence $\frac{\mathbf{E}[s|s \in W_a, \theta] - \mu_{a,0}}{\mathbf{E}[s|s \in W_b, \theta] - \mu_{b,0}}$ is increasing in θ and bounded below 1.

It can be easily checked that, $p_a > p_b$ and as $\theta \to +\infty$, $p_a/p_b \to +\infty$. So there exists a threshold $\overline{\theta}$ such that for all $\theta > \overline{\theta}$ we have $\frac{p_a(\mathbf{E}[s|s \in W_a, \theta] - \mu_{a,0})}{p_b(\mathbf{E}[s|s \in W_b, \theta] - \mu_{b,0})} > 1$.

By symmetry, a lower threshold $\underline{\theta}$ can be determined by the same way. This completes the proof.