GCOE Discussion Paper Series
Global COE Program
Human Behavior and Socioeconomic Dynamics

Discussion Paper No.238

A Theory of Multidimensional Information Disclosure

Wataru Tamura

January 2012

GCOE Secretariat
Graduate School of Economics
OSAKA UNIVERSITY
1-7 Machikaneyama, Toyonaka, Osaka, 560-0043, Japan

A Theory of Multidimensional Information Disclosure*

Wataru Tamura†

January 16, 2012

Abstract

We study disclosure of information about the multidimensional state of the world when uninformed receivers' actions affect the sender's utility. Given a disclosure rule, the receivers form an expectation about the state following each message. Under the assumption that the sender's expected utility is written as the expected value of a quadratic function of those conditional expectations, we identify conditions under which full and no disclosure is optimal for the sender and show that a linear transformation of the state is optimal if it is normally distributed. We apply our theory to advertising, political campaigning, and monetary policy.

(JEL: D83, L15, M37, D72, E52)

Keywords: information disclosure, semidefinite programming, linear transformation.

*I am very grateful to Masaki Aoyagi for his most insightful comments and invaluable suggestions. This paper (an earlier version of which had the title "Information Disclosure in Strategic Environments") was presented at YNU, Okayama University, Waseda University, Osaka Prefecture University, the Summer Workshop on Economic Theory 2010, Sapporo Applied Microeconomics Workshop, Applied Microeconomic Theory Workshop, and the 2011 Japanese Economic Association Spring Meeting. I have benefited from comments and suggestions by the participants of these seminars and the workshops, and by Junichiro Ishida, Shigehiro Serizawa, and Takashi Ui. I especially thank Takehito Masuda and Keiichi Morimoto for helpful discussions which have significantly improved earlier drafts of this paper. This paper was awarded the 14th Moriguchi Prize by the ISER at Osaka University.

†Graduate School of Economics, Osaka University, 1-7, Machikaneyama, Toyonaka, Osaka 560-0043, Japan. e-mail: jge005tw@mail2.econ.osaka-u.ac.jp.

1 Introduction

Controlling "market expectations" about the state of the world is important in various situations. For example, a central bank has to control market expectations in order to stabilize inflation and the output gap around the desired values. A manufacturing firm needs to build a good reputation for the quality of its product. A ruling political party wants to maintain a high approval rating by being sensitive to voters' expectations about its policy stance and competence. In order to control the information available to market participants, a central bank designs a communication strategy (Blinder et al. (2008), Woodford (2005), and others), a firm an advertising strategy (Anderson and Renault (2006), Johnson and Myatt (2006)), and a political party a campaign strategy (Prat (2002), Polborn and Yi (2006)).

In this paper, we analyze a model in which a privately informed sender discloses information about the realization of the state to uninformed receivers, who then engage in economic activities that affect the sender's utility. Through the choice of a *disclosure rule* that specifies the information available to the receivers for each state of the world, the sender influences the receivers' belief. The question we address in this paper is, given the prior distribution of the state, what disclosure rule maximizes the sender's expected utility.

Formally, a disclosure rule assigns to each realization of the state a probability distribution over *messages*. As such, the disclosure rule determines the joint distribution of the state and the message, which in turn determines the distribution of the receivers' belief. We assume that the sender's expected utility, which is originally a function of the joint distribution of the receivers' action profile and the state, is reduced in equilibrium to the expected value of a quadratic function of the receivers' expectation of the state.[1] Under this assumption, the sender's problem is to control the distribution (more precisely, its second moments) of the receivers' expectation of the state. A sufficient condition on the underlying preferences for this assumption is that both the sender and the receivers have quadratic utility functions over the receivers' action profile and the state. Such a specification is common in models of oligopoly, network externalities, and so on, and in the recent studies of transparency policy including Morris and Shin (2007) and Cornand and Heinemann (2008) among others. We show through our applications presented in Section 6 that this assumption is satisfied in a number of applications such as monopoly advertising, political campaigning, and monetary policy making.

[1]Chakraborty and Harbaugh (2010) study multidimensional cheap talk in a setting in which the sender's preferences are described by a continuous (but not necessarily quadratic) utility function over the receivers' conditional expectation. As illustrated in the next section, our formulation does not necessarily imply that the sender has state-independent preferences.

Our first result identifies when full and no disclosure is optimal (Theorem 1). In order to investigate the optimality of partial disclosure rule, we begin by establishing an upper bound of the sender's expected utility. We show that the upper bound depends only on the second moments of the state and is obtained by solving a *semidefinite programming* problem (Theorem 2). To the best of our knowledge, the approach based on semidefinite programming is novel in the context of information disclosure.[2] With this preparation, we show that the optimal disclosure rule is given by a linear transformation of the state when it is normally distributed (Theorem 3). We should emphasize that this is the first result that presents a complete and systematic derivation of the optimal disclosure rule, whether partial or full, in a continuous state space.

We next examine the implications of our results in three applications. In Subsection 6.1, we consider the optimal advertising strategy of a monopoly firm privately informed of its product quality and marginal cost, and show that its optimal advertising policy is to reveal less information about its product quality than the socially optimal level. In Subsection 6.2, we examine in a model of electoral competition the incentives of a political party to reveal information about its candidate and show that incumbency advantage leads to a socially inefficient amount of information revelation to voters. In Subsection 6.3, we formulate a two-period model of monetary policy and characterize the optimal disclosure rule and stabilization policy.

Optimal disclosure of information has been studied in a number of different contexts, including auctions (Milgrom and Weber (1982), Bergemann and Pesendorfer (2007), Board (2009), Ganuza and Penalva (2010)), corporate finance (Admati and Pfleiderer (2000), Boot and Thakor (2001)), interim performance evaluation (Aoyagi (2010), Goltsman and Mukherjee (2011)), transparency in policymaking (Gavazza and Lizzeri (2009), Prat (2005)), *etc.*

This paper is closely related to recent studies of Kamenica and Gentzkow (2011) and Rayo and Segal (2010), who also investigate the optimal disclosure rule under alternative specifications of the state space and the sender's utility function. Kamenica and Gentzkow (2011) study a general setting in which the sender needs to control the distribution of posterior distributions and find general properties of posteriors induced by the optimal disclosure rule. They characterize the optimal disclosure rule in some simple settings, including when the state space is binary. Rayo and Segal (2010) characterize the optimal (randomized) disclosure rule for the discrete state space when the sender has certain preferences. The main contribution of our analysis is to identify optimal disclosure in the case where the state is continuously distributed. A detailed discussion of our contribution is provided in Section 3.

[2]The same idea is also found in other applications of semidefinite programming such as minimal trace factor analysis and optimal experiment design (see, for example Vandenberghe and Boyd (1996)).

In line with much of the literature including Kamenica and Gentzkow (2011) and Rayo and Segal (2010), we assume that the sender can commit to her disclosure rule. While this is a strong assumption, it is justifiable in situations where the sender's information is verifiable *ex post* in the form of survey data, estimation results, experts' reports, and so on. In such situations, reputational and legal concerns would stop the principal from deviating from the pre-announced disclosure rule for a short-run gain. It is worth noting that the optimal disclosure rule for the normally distributed state is a linear transformation, making it easy to match the disclosed information with the private information. In this sense, it is more credible than other complex rules. Although some recent papers assume that the sender can commit to any disclosure rule (e.g., Goltsman and Mukherjee (2011)), most applied papers assume a limited ability to commit to a disclosure rule and restrict the class of disclosure rules the sender can choose from. For example, some papers including Shapiro (1986) and Ederer (2010) examine when full disclosure is superior to no disclosure while Gal-Or (1986) and Admati and Pfleiderer (2000) among others assume that the sender is able to choose only the precision of messages the receivers observe so that a closed-form solution for the sender's expected utility is obtained.

This paper also contributes to the growing literature on the social value of information.[3] In applications, we discuss the divergence between private and social incentives to disclose information in terms of informativeness of the messages generated by each disclosure rule. One advantage of our multidimensional analysis is that it allows us to examine not only the level but also the type of information that is revealed in equilibrium and at the social optimum.[4] For example, the optimal disclosure rule for a political party may reveal less information about its general competence and too much about its policy stance than the socially optimal disclosure rule.

The paper is organized as follows. Section 2 presents a motivating example. In Section 3, we set up the model and discuss our key assumptions. Section 4 identifies conditions under which full and no disclosure is optimal and characterizes an upper bound of the sender's expected utility. In Section 5, the optimal disclosure rule is explicitly obtained when the state is normally distributed. Section 6 provides applications, and Section 7 concludes the paper.

[3]For recent literature, see Morris and Shin (2002) and Angeletos and Pavan (2007) among others.

[4]For the measure of informativeness of disclosure rules, we follow Ganuza and Penalva (2010), who propose precision criteria based on the variability of conditional expectations. Especially, if the conditional expectation induced by a disclosure rule is normally distributed, its variance can measure the precision of the message generated.

2 Motivating Example

To provide a concrete example of what we analyze and what our assumption does and does not mean, we begin with a simple example in which a sender discloses information to two receivers who then take actions. This example illustrates how our reduced-form formulation arises in settings with quadratic preferences over the receivers' actions and the state. This motivating example is also useful in identifying the key issues that the theorems presented below resolve.

An organization consists of a principal with private information and two agents ($i = 1, 2$) who are hired by the principal to sell her products. The agents simultaneously choose the target consumers $a_i \in \mathbb{R}$ and their profits depend on agent-specific market condition $x_i \in \mathbb{R}$ as well as on the action profile (a_1, a_2). In particular, we suppose that agent i has a utility function

$$u_i(a_1, a_2, x_1, x_2) = -(a_i - x_i)^2 - \gamma(a_i - a_{-i})^2$$

and that the principal has

$$v(a_1, a_2, x_1, x_2) = -\sum_{i=1}^{2}(a_i - x_i)^2 - \delta(a_1 - a_2)^2$$

where $\gamma \geq 0$ and $\delta \geq 0$ measure the relative importance of coordination between two agents.[5] Each agent has incentives to adapt to the state in order to reduce the *adaptation loss*, $(a_i - x_i)^2$, and to choose a similar target in order to reduce the *coordination loss*, $(a_i - a_{-i})^2$, which may arise due to network externalities, reputations, economics of scale, and so on. The principal is privately informed about (x_1, x_2), interpreted as the information about her products and market conditions which may be estimated from past records. The prior distribution of the state is common knowledge.[6] How should the principal disclose her private information? When is full disclosure optimal?

A (deterministic) disclosure rule is a mapping $g : \mathbb{R}^2 \to M$ that assigns to each realization of the state a message $m \in M$ where the message space M is also chosen by the principal. Although we consider a more general class of (possibly randomized) disclosure rules in the following sections, we begin by comparing the following three disclosure rules; *full disclosure* $g^f(x_1, x_2) = (x_1, x_2) \in \mathbb{R}^2$, which discloses full information; *no disclosure* $g^n(x_1, x_2) = 0 \in \mathbb{R}$,

[5]Use of a coordination game with quadratic preferences is common in organization economics. Our model follows Alonso et al. (2008), Calvó-Armengol and de Martí (2009), Calvó-Armengol et al. (2009) and Dessein and Santos (2006).

[6]Our formulation allows any correlation between x_1 and x_2.

which discloses a completely uninformative message; and *average disclosure* $g^a(x_1, x_2) = (x_1 + x_2)/2 \in \mathbb{R}$, which discloses a sample mean of two variables.

The timing of the game is as follows. First, the principal commits to a disclosure rule $g \in \{g^f, g^n, g^a\}$. Second, the state of the world is realized, and a message $m = g(x_1, x_2)$ is disclosed according to the disclosure rule. Third, given the disclosure rule and the message, the agents form a posterior belief and choose actions.

First, we derive the agents' equilibrium strategies and the principal's expected utility. Given g and $m \in M$, two agents play a game whose unique Nash equilibrium is, for $i = 1, 2$,

$$a_i^*(m) =(1-\psi)\mathbb{E}[x_i|m] + \psi\mathbb{E}[x_{-i}|m]$$

where $\psi = \gamma/(1+2\gamma)$. The equilibrium strategy is linear in the conditional expectations. Therefore we denote the equilibrium strategy profile by $a^*(\hat{x}_1, \hat{x}_2) = (a_1^*(\hat{x}_1, \hat{x}_2), a_2^*(\hat{x}_1, \hat{x}_2))$ where $\hat{x}_j \equiv \mathbb{E}(x_j|m)$ is the agents' *estimate* of the state. Notice that a prior distribution of the state and a disclosure rule specify the joint distribution of $(x_1, x_2, \hat{x}_1, \hat{x}_2)$, which determines the joint distribution of (x_1, x_2, a_1^*, a_2^*) in equilibrium. We now rewrite the principal's expected utility. The adaptation loss is written as

$$\begin{aligned}
\mathbb{E}(a_i^* - x_i)^2 =& \mathbb{E}_{(\hat{x}_1,\hat{x}_2)}\left[\mathbb{E}_{x_i}\left[(a_i^*)^2 - 2a_i^* x_i + x_i^2|\hat{x}_1, \hat{x}_2\right]\right] \\
=& \mathbb{E}_{(\hat{x}_1,\hat{x}_2)}\left[(a_i^*)^2 - 2a_i^*\hat{x}_i\right] + \mathbb{E}_{(\hat{x}_1,\hat{x}_2)}\left[\mathbb{E}\left(x_i^2|\hat{x}_1, \hat{x}_2\right)\right] \\
=& \mathbb{E}_{(\hat{x}_1,\hat{x}_2)}\left[-\hat{x}_i^2 + \psi^2\left(\hat{x}_1^2 - 2\hat{x}_1\hat{x}_2 + \hat{x}_2^2\right)\right] + \mathbb{E}x_i^2.
\end{aligned}$$

Similarly, the coordination loss is

$$\mathbb{E}(a_1^* - a_2^*)^2 = \mathbb{E}_{(\hat{x}_1,\hat{x}_2)}\left[(1-2\psi)\left(\hat{x}_1^2 - 2\hat{x}_1\hat{x}_2 + \hat{x}_2^2\right)\right].$$

Thus, the principal's expected utility is written as

$$\begin{aligned}
&\mathbb{E}_{(x_1,x_2,\hat{x}_1,\hat{x}_2)} v(a^*(\hat{x}_1, \hat{x}_2), x_1, x_2) \\
=& -\sum_{i=1}^{2}\mathbb{E}(a_i^*(\hat{x}_1, \hat{x}_2) - x_i)^2 - \delta\mathbb{E}(a_1^*(\hat{x}_1, \hat{x}_2) - a_2^*(\hat{x}_1, \hat{x}_2))^2 \\
=& \mathbb{E}_{(\hat{x}_1,\hat{x}_2)}\left[(1-\phi)(\hat{x}_1^2 + \hat{x}_2^2) + 2\phi\hat{x}_1\hat{x}_2\right] - \mathbb{E}x_1^2 - \mathbb{E}x_2^2
\end{aligned}$$

where $\phi \equiv (2\gamma^2+\delta)/(1+2\gamma)^2$. Since $\mathbb{E}x_1^2$ and $\mathbb{E}x_2^2$ are independent of the disclosure rule, the principal's problem is reduced to the maximization of $\mathbb{E}\hat{v}(\hat{x}_1, \hat{x}_2) + c$ where c is a constant

Table 1: Comparison among the three rules.

Rule		No	Full	Average
$\mathrm{var}(\hat{x}_1)$		0	1/12	1/24
$\mathrm{var}(\hat{x}_2)$		0	1/12	1/24
$\mathrm{cov}(\hat{x}_1, \hat{x}_2)$		0	0	1/24
	for $\delta < \underline{\delta}$	3rd	1st	2nd
Rank of $\mathbb{E}v$	for $\underline{\delta} < \delta < \overline{\delta}$	3rd	2nd	1st
	for $\delta > \overline{\delta}$	2nd	3rd	1st

The thresholds are $\underline{\delta} = (1 + 4\gamma)/2$ and $\overline{\delta} = 1 + 4\gamma + 2\gamma^2$.

and $\hat{v}$ is a quadratic function defined by

$$\hat{v}(\hat{x}_1, \hat{x}_2) = (1 - \phi)(\hat{x}_1^2 + \hat{x}_2^2) + 2\phi\hat{x}_1\hat{x}_2. \tag{1}$$

Note that in general the disclosure rule cannot affect the expected value of the estimates, that is $\mathbb{E}\hat{x}_i = \mathbb{E}\mathbb{E}(x_i|m) = \mathbb{E}x_i$. Note also that the second moments are given by $\mathbb{E}\hat{x}_i^2 = \mathrm{var}(\hat{x}_i) + (\mathbb{E}x_i)^2$ and $\mathbb{E}\hat{x}_1\hat{x}_2 = \mathrm{cov}(\hat{x}_1, \hat{x}_2) + (\mathbb{E}x_1)(\mathbb{E}x_2)$. From these observations, we have

$$\begin{aligned}\mathbb{E}v(a^*, x_1, x_2) =& (1 - \phi)(\mathrm{var}(\hat{x}_1) + \mathrm{var}(\hat{x}_2)) + 2\phi\mathrm{cov}(\hat{x}_1, \hat{x}_2) \\ & - \mathrm{var}(x_1) - \mathrm{var}(x_2) - \phi\left(\mathbb{E}x_1 - \mathbb{E}x_2\right)^2.\end{aligned}$$

Since the second line is independent of the disclosure rule, the principal's expected utility is also expressed as a linear function of variance-covariances of the estimates (plus a constant term).

It is worth noting that the expected utility conditional on each message is not necessarily written as $\hat{v}(\hat{x}_1, \hat{x}_2) + c$. The former is expressed as

$$\mathbb{E}\left[v(a^*, x_1, x_2)|m\right] = \hat{v}(\hat{x}_1, \hat{x}_2) - \mathbb{E}(x_1^2 + x_2^2|m).$$

The second term in the right-hand side may change as what message is disclosed while its average is determined by the prior distribution but not the disclosure rule. For any disclosure rule, the expectation of the conditional expectation becomes the unconditional expectation so that $\mathbb{E}_m[\mathbb{E}[(x_1^2 + x_2^2)|m]] = \mathbb{E}(x_1^2 + x_2^2)$.

We now evaluate the performances of the three disclosure rules. For ease of exposition, we assume that x_1 and x_2 are independent and uniformly distributed on $[-1/2, 1/2]$. Table

1 gives the characteristics of the three deterministic disclosure rules. The first three rows report the variances and covariance of the estimates while the last three rows report the ranking of the principal's expected utility. Given $\gamma > 0$, there are two thresholds for δ; $\underline{\delta} \equiv (1+4\gamma)/2$ and $\overline{\delta} \equiv 1 + 4\gamma + 2\gamma^2$. Full disclosure is superior to average disclosure for $\delta < \underline{\delta}$ while the reverse holds for $\delta > \underline{\delta}$. Intuitively, when the principal puts a lower weight on coordination ($\delta < \underline{\delta}$), the principal should disclose full information in order to induce adaptation even if it may cause mis-coordination. In contrast, when the coordination is important ($\delta > \underline{\delta}$), the principal induces similar decisions by disclosing the average state.

Several questions arise. What is the optimal disclosure rule? Is a higher covariance between $\hat{x}_1$ and $\hat{x}_2$ incompatible with a higher variance of each $\hat{x}_i$? What determines the limit of information disclosure as a means of controlling expectations? In the following sections, we will answer these questions and find that in the above example full and average disclosure is indeed optimal among the general class of disclosure rules when the state is uniformly distributed.[7]

3 The Model

A sender privately observes the multidimensional state $\mathbf{x} = (x_1, \ldots, x_k)' \in \mathbb{R}^k$ where $\mathbf{x}$ has a density over a convex support in $\mathbb{R}^k$ with a non-empty interior, zero mean and a positive definite variance matrix Σ.[8] The sender publicly discloses information about the realization of the state to uninformed receivers who then engage in economic activities such as consumption, investment, *etc.* that affect the sender's utility.

For the inducement of preferred actions, the sender controls what information to make available to the receivers by choosing a *disclosure rule* (α, M), which consists of a measurable set M of messages and a family of conditional probability distributions $\{\alpha(\cdot|\mathbf{x})\}_{\mathbf{x}}$ over M. We assume that the disclosure rule (α, M) is such that the conditional expectation $\mathbb{E}[\mathbf{x}|m]$ exists for every $m \in M$ and so does its second moment $\text{var}(\mathbb{E}[\mathbf{x}|m])$.[9] Let $\hat{\mathbf{x}} \equiv \mathbb{E}[\mathbf{x}|m]$ be the conditional expectation given m, call the *estimates.* A disclosure rule induces a joint distribution of $(\mathbf{x}, m)$, and hence a joint distribution of $(\mathbf{x}, \hat{\mathbf{x}})$. This definition includes the following communication strategies that are common in the literature of information economics; *full disclosure*, that reveals the realization of the state; *no disclosure*, that reveals no information; *noisy communication*, that adds white noises to the sender's observation; *partition*, that reveals an element of the partition over the state space that contains the

[7] In Section 5, we also characterize the optimal disclosure rule when the state is normally distributed.

[8] We denote by A' the transpose of matrix A. Throughout the paper, all untransposed vectors are column vectors.

[9] A sufficient condition is that the support of $\mathbf{x}$ is a compact set in $\mathbb{R}^k$.

state. The disclosure rule is *deterministic* if there exists a function $g : \mathbb{R}^k \to M$ such that $m = g(\mathbf{x})$ almost surely.[10] As we have seen in the previous section, disclosing the average value of the state, $m = g(\mathbf{x}) = \sum x_i/k$, is an example of a deterministic disclosure rule. In contrast, disclosing noisy messages, $m_i = x_i + \varepsilon_i$ where $\varepsilon_i \sim N(0, \sigma_i^2)$, is a typical example of a stochastic disclosure rule.

As discussed in the introduction, we consider situations in which the sender controls the receivers' actions, denoted by a, through information disclosure. Let $v(a, \mathbf{x})$ be the sender's utility function. In what follows, we make the following two assumptions. First, we suppose that the receivers' behavior is simply a continuous function of their conditional expectation of the state. We denote their actions given $\hat{\mathbf{x}} = \mathbb{E}[\mathbf{x}|m]$ by $a^*(\hat{\mathbf{x}})$. Under this assumption, the sender's problem is to control the joint distribution of $\mathbf{x}$ and $\hat{\mathbf{x}}$ so as to maximize her expected utility $\mathbb{E}_{(\mathbf{x},\hat{\mathbf{x}})} v(a^*(\hat{\mathbf{x}}), \mathbf{x})$. Second, there exists a $k \times k$ symmetric matrix V such that for any disclosure rule,

$$\mathbb{E}_{(\mathbf{x},\hat{\mathbf{x}})} v(a^*(\hat{\mathbf{x}}), \mathbf{x}) = \mathbb{E}_{\hat{\mathbf{x}}}\left[\hat{\mathbf{x}}' V \hat{\mathbf{x}}\right] + c \tag{2}$$

where c is a constant that is independent of the disclosure rule. Let $\hat{v}(\hat{\mathbf{x}}) = \hat{\mathbf{x}}' V \hat{\mathbf{x}}$. A sufficient condition is that the receivers' equilibrium strategies are given by an affine function of their conditional expectations and the sender has a quadratic utility function over a and $\mathbf{x}$. We call $\mathbb{E}\hat{v}(\hat{\mathbf{x}})$ the *gain* from a disclosure rule.[11]

As briefly discussed in the previous section, the law of iterated expectations plays a key role in deriving such a representation. First, the expected value of the product $a^*(\hat{\mathbf{x}}) \cdot x_i$ is expressed as the expected value of a function of the estimates. That is, $\mathbb{E}_{(\mathbf{x},\hat{\mathbf{x}})}[a^*(\hat{\mathbf{x}}) x_i] = \mathbb{E}_{\hat{\mathbf{x}}}[a^*(\hat{\mathbf{x}}) \hat{x}_i]$. Second and more importantly, the expected utility conditional on the message, $\mathbb{E}[v(a^*(\hat{\mathbf{x}}), \mathbf{x})|m]$, is *not* necessarily written as $\hat{v}(\hat{\mathbf{x}}) + c$. For example, suppose that the sender has $v(a, x) = -(a - x)^2$ and the receiver has $u(a, x) = -(a - x)^2$ so that the receiver chooses $a^* = \hat{x}$. Then the sender's expected utility conditional on m is $\mathbb{E}[v(a^*, x)|m] = \hat{x}^2 - \mathbb{E}[x^2|m]$ while $\hat{v}(\hat{x}) = \hat{x}^2$. Although the disclosure rule affects the distribution of $\mathbb{E}_x[x^2|m]$, it cannot affect its average value $\mathbb{E}_m[\mathbb{E}_x[x^2|m]] = \mathbb{E}x^2$. Thus, our formulation may apply when the sender's conditional expected utility given each m cannot be written as a quadratic function of the estimates.

An important consequence of our assumptions is that the sender's expected utility can

[10] Note that every deterministic disclosure rule can be represented by a partition and vice versa. For example, a message m under a deterministic rule g is equivalent to disclosing its inverse image $g^{-1}(m) = \{\mathbf{x} \in \mathbb{R}^k : g(\mathbf{x}) = m\}$.

[11] This terminology is due to Kamenica and Gentzkow (2011), who define it to be the difference between the sender's expected utilities under a disclosure rule and no disclosure. In this paper, no disclosure induces $\hat{\mathbf{x}} = \mathbf{0}$ and hence $\mathbb{E}v = c$.

be expressed as $\mathbb{E}_{\hat{\mathbf{x}}}[\hat{\mathbf{x}}'V\hat{\mathbf{x}}] + c = \mathrm{tr}(V\hat{\Sigma}) + c$ where $\hat{\Sigma} \equiv \mathbb{E}[\hat{\mathbf{x}}\hat{\mathbf{x}}']$ denotes the second moment of the estimates.[12] There are two key features of the sender's problem that immediately follow from this equality: (i) if two disclosure rules induce the same $\hat{\Sigma}$, they yield the same expected utility, and (ii) given $\hat{\Sigma}$, it is easy to compute the sender's expected utility, $\mathrm{tr}(V\hat{\Sigma}) + c$.[13]

Although such a reduced-form formulation can be generated in different ways from the underlying preferences of the receivers and their equilibrium behaviors as illustrated in the previous section, there are important cases that cannot be reduced to our model. The first case is where the receivers' action space is discrete. In the above simple example, if the action space is given by $\{-1, 1\}$, then the receiver's action is not continuous in his expectation of the state.[14] Second, even when the receiver's behavior is given by a continuous function of the estimates, the sender's utility function should not be too complex. For example, suppose that $a^*(\hat{x}) = \hat{x}$ and $v(a, x) = -(a+1)^2(a-1)^2$. Then the sender's expected utility cannot be expressed in the form of (2).

Rayo and Segal (2010) study optimal information disclosure where the sender's expected utility is written as $\mathbb{E}\hat{x}_1\hat{x}_2 + c$, and characterize the optimal *randomized* disclosure rule for a finite state space. Since we assume that the state has a continuous distribution, our analysis is based on different techniques and applicable to common distributions such as uniform and normal. In Section 5, we will find that the disclosure of a weighted average becomes a solution to their problem when the state is normally distributed.[15]

Kamenica and Gentzkow (2011) consider a general problem where the sender's expected utility is expressed as the expected value of a *function of the receiver's posterior belief* and characterize the posterior beliefs induced by the optimal disclosure rule. They provide the optimal (partial) disclosure rule in simple settings, especially when the state space is binary. They also analyze whether no disclosure is suboptimal for the setting in which the sender's utility depends only on the receiver's conditional expectation of the state. Although we make a stronger assumption on the sender's preferences, we provide a simple and complete characterization of the optimal disclosure rule in the continuous state space, which is useful for applied research.

[12]For any rule (α, M), we have $\mathbb{E}[\hat{\mathbf{x}}'V\hat{\mathbf{x}}] = \mathbb{E}[\mathrm{tr}(\hat{\mathbf{x}}'V\hat{\mathbf{x}})] = \mathbb{E}[\mathrm{tr}(V\hat{\mathbf{x}}\hat{\mathbf{x}}')] = \mathrm{tr}(V\mathbb{E}[\hat{\mathbf{x}}\hat{\mathbf{x}}']) = \mathrm{tr}(V\hat{\Sigma})$ where $\mathrm{tr}(A)$ is the trace of matrix A.

[13]Even though this operation itself does not rely on the assumption that $\hat{\mathbf{x}}$ is the receivers' conditional expectation of $\mathbf{x}$, we often use it when we derive matrix V. In the above simple example, we use it when we compute $\mathbb{E}_{(x,\hat{x})}xa^*(\hat{x}) = \mathbb{E}\hat{x}^2$.

[14]In such a case, the receiver takes $a = 1$ $(= -1)$ if $\hat{x} \geq (<) 0$, respectively. That is, the first assumption is not satisfied. In this case, the sender's expected utility is expressed as $\mathbb{E}_{(x,\hat{x})}v(a^*, x) = 1 + 2b\mathbb{E}(|\hat{x}|) - \mathbb{E}x^2$.

[15]One can also show that, by applying Theorem 2 presented below, the average disclosure is optimal if (x_1, x_2) is uniformly distributed over $[0, 1]^2$.

4 Optimal Disclosure and Semidefinite Programming

First, we identify when full and no disclosure is optimal and when partial disclosure yields a higher expected utility than full and no disclosure. Detailed proof is in Appendix A

Theorem 1 (i) *Full disclosure is optimal if and only if $\hat{v}$ is a convex function or, equivalently, V is positive semidefinite.* (ii) *No disclosure is optimal if and only if $\hat{v}$ is a concave function or, equivalently, V is negative semidefinite.*

The *if parts*, which simply follow from Jensen's inequality and the law of iterated expectations, are already known in the literature. The *only if part* of (ii) is also found in Kamenica and Gentzkow (2011)[Proposition 3], who analyze a setting in which the sender has a continuous but not necessarily quadratic utility function over conditional expectations. In contrast, the *only if part* of (i) is novel and relies on the quadraticity of $\hat{v}$. To illustrate the main idea behind the proof, consider the following example: suppose that $V = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, which is not a positive semidefinite matrix so that there is a vector $\mathbf{x}_- = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ that satisfies $\mathbf{x}_-' V \mathbf{x}_- < 0$. Then any $\mathbf{x}$ can be expressed as $\mathbf{x} = \beta \mathbf{x}_+ + \gamma \mathbf{x}_-$ where $\mathbf{x}_+ = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. We argue that a partial disclosure rule that reveals only β yields a higher gain than full disclosure that reveals both β and γ. When only β is disclosed, the estimate is given by

$$\hat{\mathbf{x}} = \beta \mathbf{x}_+ + \mathbb{E}(\gamma|\beta)\mathbf{x}_-,$$

and then the realization of the state is written as

$$\mathbf{x} = \hat{\mathbf{x}} + (\gamma - \mathbb{E}(\gamma|\beta))\,\mathbf{x}_-. \tag{3}$$

Indeed, if the sender reveals the realization of γ in addition to β, an additional variation in the receivers' expectation, $(\gamma - \mathbb{E}(\gamma|\beta))\mathbf{x}_-$, is generated. Using the expression of (3), the sender's gain from full disclosure can be written as follows:

$$\begin{aligned}
\mathbb{E}\left[\mathbf{x}' V \mathbf{x}\right] =& \mathbb{E}\left[[\hat{\mathbf{x}} + (\gamma - \mathbb{E}(\gamma|\beta))\mathbf{x}_-]' \, V \left[\hat{\mathbf{x}} + (\gamma - \mathbb{E}(\gamma|\beta))\mathbf{x}_-\right]\right] \\
=& \mathbb{E}\left[\hat{\mathbf{x}}' V \hat{\mathbf{x}}\right] + 2\mathbb{E}_\beta\left[\mathbb{E}_\gamma\left[(\gamma - \mathbb{E}(\gamma|\beta))\mathbf{x}_-' V \hat{\mathbf{x}}|\beta\right]\right] + \mathbb{E}\left[(\gamma - \mathbb{E}(\gamma|\beta))^2 \mathbf{x}_-' V \mathbf{x}_-\right] \\
=& \mathbb{E}\left[\hat{\mathbf{x}}' V \hat{\mathbf{x}}\right] + \left(\mathbf{x}_-' V \mathbf{x}_-\right) \mathbb{E}_\beta\left[\mathrm{var}(\gamma|\beta)\right].
\end{aligned}$$

Since $\mathbb{E}\left[\mathrm{var}(\gamma|\beta)\right] > 0$ and $\mathbf{x}_-' V \mathbf{x}_- < 0$, we have $\mathbb{E}[\mathbf{x}' V \mathbf{x}] < \mathbb{E}[\hat{\mathbf{x}}' V \hat{\mathbf{x}}]$. Intuitively, compared with full disclosure, the partial disclosure specified above reduces an unfavorable variability of conditional expectations. Hence, full disclosure is suboptimal whenever V is not positive semidefinite.

The proof of the *only if* part of (ii) in Appendix is based on a reverse argument that, compared with no disclosure, the sender can generate a favorable variability of conditional expectations if V is not negative semidefinite. In particular, the sender controls information so that the conditional expectations lie on a vector $\mathbf{x}_+$ such that $\mathbf{x}_+' V \mathbf{x}_+ > 0$.

For the purpose of use in applications in Section 6, we restate Theorem 1 for the case of $k \leq 2$.

Corollary 1 *For the one-dimensional state space ($k = 1$), full disclosure is optimal if and only if $V \geq 0$ and no disclosure is optimal if and only if $V \leq 0$. For the two-dimensional state space ($k = 2$), full disclosure is optimal if and only if $V_{11}, V_{22}, \det(V) \geq 0$ and no disclosure is optimal if and only if V_{11}, $V_{22}, -\det(V) \leq 0$.*

In the motivating example, V is positive semidefinite if and only if $\delta \leq (1+4\gamma)/2$, which coincides with the condition for the optimality of full disclosure among the three simple rules. So far we know little about what disclosure rule is optimal when $\delta > (1+4\gamma)/2$ in the example.

To investigate the optimality of partial disclosure, we begin by establishing an upper bound of the sender's expected utility (or equivalently the gain) attainable through information control. Recall that in general $\mathbb{E}[\hat{\mathbf{x}}' V \hat{\mathbf{x}}] = \mathrm{tr}(V\hat{\Sigma})$ where $\hat{\Sigma} = \mathbb{E}[\hat{\mathbf{x}}\hat{\mathbf{x}}'] = \mathrm{var}(\hat{\mathbf{x}})$. Hereafter, we interpret the sender's problem as the choice of a variance matrix of $\hat{\mathbf{x}}$ by choosing a disclosure rule.

We now investigate conditions on $\hat{\Sigma}$ that can be induced by a disclosure rule. First we know that (i) $\hat{\Sigma}$ must be positive semidefinite since it is a variance matrix. Furthermore, for any joint distribution of $(\mathbf{x}, m)$, the *law of total variance* holds;

$$\mathrm{var}(\mathbf{x}) = \mathbb{E}[\mathrm{var}(\mathbf{x}|m)] + \mathrm{var}(\mathbb{E}(\mathbf{x}|m))$$

where $\mathrm{var}(\mathbf{x}|m) = \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}|m])(\mathbf{x} - \mathbb{E}[\mathbf{x}|m])'|m]$. Since every variance matrix is positive semidefinite, so is its expectation $\mathbb{E}[\mathrm{var}(\mathbf{x}|m)]$.[16] This implies that (ii) $\Sigma - \hat{\Sigma}$ must be positive semidefinite. Let $\succeq$ denote the Löwner partial ordering on the set of $k \times k$ symmetric positive semidefinite matrices. That is, for two symmetric positive semidefinite matrices A and B, $A \succeq B$ if $A - B$ is positive semidefinite.[17] Let O and I denote the zero and the identity matrix,

[16] For all non-zero vector $\mathbf{z} \in \mathbb{R}^k$, we have $\mathbf{z}'\mathbb{E}[\mathrm{var}(\mathbf{x}|m)]\mathbf{z} = \mathbb{E}[\mathbf{z}'\mathrm{var}(\mathbf{x}|m)\mathbf{z}] \geq 0$ since $\mathrm{var}(\mathbf{x}|m)$ is positive semidefinite for any $m \in M$.

[17] For the standard notation in matrix algebra and some basic properties of matrices, see Horn and Johnson (1985) and Boyd and Vandenberghe (2004).

respectively. Then, every $\hat{\Sigma}$ must satisfy $\Sigma \succeq \hat{\Sigma} \succeq O$. Roughly speaking, variance matrices of the estimates induced by the disclosure rule are (partially) ordered by the matrix inequality according to which full (no) disclosure attains the greatest (least) element. Therefore, an upper bound on the gain is given by solving the following *semidefinite programming*:

$$\max_{\Sigma \succeq \hat{\Sigma} \succeq O} \operatorname{tr}(V\hat{\Sigma}).$$

To simplify the notation in the problem and its solution, it is useful to change the variable in the above program. Let $Z \equiv \Sigma^{-\frac{1}{2}}\hat{\Sigma}\Sigma^{-\frac{1}{2}}$ and $W \equiv \Sigma^{\frac{1}{2}}V\Sigma^{\frac{1}{2}}$. Since the variance matrix Σ is nonsingular, Z must be a symmetric positive semidefinite matrix. Then the condition $\Sigma \succeq \hat{\Sigma} \succeq O$ is equivalent to $I \succeq Z \succeq O$. It is straightforward to see that $Z = \Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{1}{2}} = I$ for full disclosure and $Z = \Sigma^{-\frac{1}{2}}O\Sigma^{-\frac{1}{2}} = O$ for no disclosure. The gain is also written in terms of Z and W as $\operatorname{tr}(V\hat{\Sigma}) = \operatorname{tr}(WZ)$. Thus, an upper bound of the gain is characterized as follows:

Lemma 1 *Let $W = \Sigma^{\frac{1}{2}}V\Sigma^{\frac{1}{2}}$. Then the upper bound of the gain is given by solving the following semidefinite programming:*

$$\text{(SDP)} \qquad \max_{I \succeq Z \succeq O} \operatorname{tr}(WZ).$$

Before presenting the solution to SDP, it may be helpful to give an intuition of the constraint $\Sigma \succeq \hat{\Sigma} \succeq O$ in the context of the motivating example in Section 2. Recall that matrix V has entries $V_{11} = V_{22} = 1 - \phi$ and $V_{12} = V_{21} = \phi$ where $\phi = (2\gamma^2 + \delta)/(1 + 2\gamma)^2$ (see (1)). Moreover, we have assumed that x_1 and x_2 are independent and uniformly distributed over $[-1/2, 1/2]$ so that Σ has $\sigma_{11} = \sigma_{11} = 1/12$ and $\sigma_{12} = 0$. It immediately follows from $\hat{\Sigma} \succeq O$ (i.e., $\hat{\Sigma}$ is positive semidefinite) that $\hat{\sigma}_{11} \geq 0$, $\hat{\sigma}_{22} \geq 0$ and $\hat{\sigma}_{11}\hat{\sigma}_{22} \geq \hat{\sigma}_{12}^2$. That is, the variances of the estimates are nonnegative and the correlation between two estimates, $\operatorname{corr}(\hat{x}_1, \hat{x}_2) = \sqrt{\hat{\sigma}_{12}^2/(\hat{\sigma}_{11}\hat{\sigma}_{22})}$, is in $[-1, 1]$. Similarly, from $\Sigma \succeq \hat{\Sigma}$ (i.e., $\Sigma - \hat{\Sigma}$ is positive semidefinite), we have $\sigma_{11} \geq \hat{\sigma}_{11}$, $\sigma_{22} \geq \hat{\sigma}_{22}$, and

$$(\sigma_{12} - \hat{\sigma}_{12})^2 \leq (\sigma_{11} - \hat{\sigma}_{11})(\sigma_{22} - \hat{\sigma}_{22}). \tag{4}$$

The first two conditions imply that the variance of the estimate cannot exceed that of the underlying state. It may make sense that any message about the state cannot be more informative than revealing the state itself. The condition (4) implies that to generate a certain covariance of the estimates that differs from that of the underlying state ($\operatorname{cov}(\hat{x}_1, \hat{x}_2) \neq \operatorname{cov}(x_1, x_2)$), the sender must induce lower variances of the estimates than that under full disclosure ($\operatorname{var}(\hat{x}_1) < \operatorname{var}(x_1)$ and $\operatorname{var}(\hat{x}_2) < \operatorname{var}(x_2)$). In the context of the

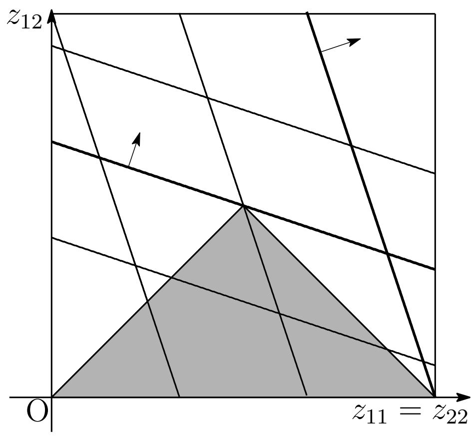

Figure 1: Solution to the semidefinite programming

motivating example, this condition turns out to be an important trade-off for the principal. Recall that both the adaptation and coordination losses are decreasing in $\hat{\sigma}_{12} = \mathbb{E}(\hat{x}_1\hat{x}_2)$. To induce a higher covariance between the two estimates, the sender must reduce the variance of the estimates, which is also valuable to the principal whenever $\delta < \overline{\delta}$. Roughly speaking, to facilitate coordination between the two agents, the principal needs to withhold some information and reduce the degree of adaptation.

We now apply Lemma 1 to the problem in the motivating example and obtain an upper bound of the gain. Since $\Sigma = \frac{1}{12}I$, we have $Z = 12\hat{\Sigma}$ and $W = \frac{1}{12}V$. Moreover, the constraint $I \succeq Z \succeq O$ is expressed as (i) $z_{11}, z_{22} \geq 0$, $z_{12}^2 \leq z_{11}z_{22}$, and (ii) $z_{11}, z_{22} \leq 1$, $z_{12}^2 \leq (1 - z_{11})(1 - z_{22})$. Thus, SDP for the example is written as

$$\begin{aligned} \text{(SDP)} \qquad & \max_{z_{11}, z_{12}, z_{22}} \frac{1}{12}[(1-\phi)(z_{11}+z_{22}) + 2\phi z_{12}] \\ & \text{subject to } z_{11}, z_{22} \in [0,1] \\ & \qquad\qquad z_{12}^2 \leq \min\{z_{11}z_{22},\ (1-z_{11})(1-z_{22})\}. \end{aligned}$$

Notice that in order to relax the constraint on z_{12}, we have to choose $z_{11} = z_{22}$.[18] Then the inequality constraint is reduced to $z_{12} \leq \min\{z_{11}, (1 - z_{11})\}$. Figure 1 depicts the feasible set of (z_{11}, z_{12}) as a shaded area with the level curves of the objective function $\mathrm{tr}(WZ)$. Recall that the solution Z to SDP corresponds to the variance matrix of the estimates as $\hat{\Sigma} = \Sigma^{\frac{1}{12}} Z \Sigma^{\frac{1}{12}} = \frac{1}{12}Z$. If the slope $w_{11}/w_{12} = (1 - \phi)/\phi$ of the level curves is greater than 1, the solution is $(z_{11}, z_{22}, z_{12}) = (1, 1, 0)$, or equivalently $(\hat{\sigma}_{11}, \hat{\sigma}_{22}, \hat{\sigma}_{12}) = (1/12, 1/12, 0)$, which is achieved by full disclosure (see the first three rows in Table 1). On the other

[18]For any (z_{11}, z_{22}) with $z_{11} \neq z_{22}$, consider $\tilde{z}_{11} = \tilde{z}_{22} = (z_{11} + z_{22})/2$. Since $w_{11} = w_{22} = (1 - \phi)/12$, we have $w_{11}z_{11} + w_{22}z_{22} = w_{11}\tilde{z}_{11} + w_{22}\tilde{z}'_{22}$. Moreover, $z_{11}z_{22} < \tilde{z}_{11}\tilde{z}_{22}$ and $(1 - z_{11})(1 - z_{22}) < (1 - \tilde{z}_{11})(1 - \tilde{z}_{22})$. Thus we can relax the inequality constraint without altering the objective value.

hand, if w_{11}/w_{12} is less than 1, the solution is $(z_{11}, z_{12}, z_{22}) = (1/2, 1/2, 1/2)$, or equivalently $(\hat{\sigma}_{11}, \hat{\sigma}_{22}, \hat{\sigma}_{12}) = (1/24, 1/24, 1/24)$, which is, indeed, what average disclosure achieves. Thus, we find that average disclosure attains the upper bound for $\delta > (1 + 4\gamma)/2$.

Corollary 2 *In the motivating example with a prior $x_i \overset{\text{iid}}{\sim} U[-\frac{1}{2}, \frac{1}{2}]$ for $i = 1, 2$, full disclosure is optimal if $\delta \leq (1 + 4\gamma)/2$ and average disclosure is optimal if $\delta > (1 + 4\gamma)/2$.*

Although Lemma 1 provides a key insight into the control of conditional expectations, we need some knowledge in matrix algebra to solve the program. Here we present a solution to SDP and relegate its derivation to Appendix A.[19]

Theorem 2 *Let $Q_+ = [\mathbf{q}_1, \ldots, \mathbf{q}_r]$ consist of the eigenvectors associated with the nonnegative eigenvalues of $W = \Sigma^{\frac{1}{2}} V \Sigma^{\frac{1}{2}}$. Then a projection matrix $Z = P_{Q_+} = Q_+(Q'_+Q_+)^{-1}Q'_+$ is a solution to SDP. Moreover, the upper bound achieved equals the sum of all positive eigenvalues of W.*

An important implication for $k = 2$ is that the two estimates $\hat{x}_1$ and $\hat{x}_2$ must be perfectly correlated when x_1 and x_2 are independent. To see this, suppose that $\sigma_i^2 = \text{var}(x_i)$ for $i = 1, 2$ and that V (and hence W) is neither positive nor negative semidefinite so that there exists exactly one positive eigenvalue. Suppose also that $V_{12} \neq 0$ so that the optimal disclosure rule is nontrivial.[20] Let (q_1, q_2) be the eigenvector associated with the unique positive eigenvalue. Then the solution to SDP is given by

$$Z = \begin{pmatrix} \frac{q_1^2}{q_1^2+q_2^2} & \frac{q_1 q_2}{q_1^2+q_2^2} \\ \frac{q_1 q_2}{q_1^2+q_2^2} & \frac{q_2^2}{q_1^2+q_2^2} \end{pmatrix},$$

and the second moment of the estimates is

$$\hat{\Sigma} = \Sigma^{\frac{1}{2}} Z \Sigma^{\frac{1}{2}} = \begin{pmatrix} \frac{\sigma_1^2 q_1^2}{q_1^2+q_2^2} & \frac{\sigma_1 q_1 \sigma_2 q_2}{q_1^2+q_2^2} \\ \frac{\sigma_1 q_1 \sigma_2 q_2}{q_1^2+q_2^2} & \frac{\sigma_2^2 q_2^2}{q_1^2+q_2^2}. \end{pmatrix}$$

Thus, the correlation between $\hat{x}_1$ and $\hat{x}_2$ equals $q_1 q_2/|q_1 q_2|$, which is either 1 or -1 whenever both $\hat{x}_1$ and $\hat{x}_2$ have positive variances.[21] In other words, to achieve the upper bound, the two estimates must satisfy a linear restriction $\hat{x}_2 = \beta \hat{x}_1$ where the coefficient is given by $\beta = \sigma_2 q_2/\sigma_1 q_1$. Another implication is that the solution to SDP satisfies $z_{11} + z_{22} = 1$. This

[19]If all eigenvalues of W is negative, the zero matrix $Z = O$ achieves the upper bound as established in Theorem 1.

[20]If x_1 and x_2 are independent and $V_{12} = 0$, then the optimal disclosure is $g(x_1, x_2) = x_1$ when $V_{11} > 0 > V_{22}$ and $g(x_1, x_2) = x_2$ when $V_{22} > 0 > V_{11}$.

[21]For mutually dependent states, we have $\text{corr}(\hat{y}_1, \hat{y}_2) \in \{-1, 1\}$ where $\begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \end{pmatrix} = \Sigma^{-\frac{1}{2}} \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix}$.

implies that two variables, $z_{11} \in [0,1]$ and $\rho \in \{-1,1\}$, determine the other two variables z_{22} and z_{12} as $z_{22} = 1 - z_{11}$ and $z_{12} = \rho\sqrt{z_{11}(1-z_{11})}$. From these observations, the search for the upper bound is characterized as a simple maximization problem that does not need to compute the positive eigenvalue: let $\rho \in \{-1,1\}$ and $z \in [0,1]$ be parameters and $h(\rho, z)$ be a function on $\{-1,1\} \times [0,1]$ defined by

$$h(\rho, z) = \sigma_1^2 V_{11} z + 2\sigma_1\sigma_2 V_{12}\rho\sqrt{z(1-z)} + \sigma_2^2 V_{22}(1-z).$$

Then we have $\max h(\rho, z) = \max_{\Sigma \succeq \hat{\Sigma} \succeq O} \operatorname{tr}(V\hat{\Sigma})$.

Although Theorem 2 tells us what distribution of the estimates should be induced by the optimal disclosure rule, it tells little about how to construct a disclosure rule that induces such a distribution of the estimates. While the upper bound characterized in Theorem 2 depends only on the second moment of the underlying distribution of the state, the optimal rule may depend on the entire distribution of $\mathbf{x}$ since we have to obtain the conditional expectations for each m. In general, there is little hope of finding a disclosure rule that attains the upper bound. This leaves us with two choices: one is to investigate necessary conditions for the optimal disclosure rule as in Kamenica and Gentzkow (2011) and Rayo and Segal (2010), while the other is to characterize the optimal rules for some class of state distributions. We take the second approach and identify optimal disclosure rules under the assumption that the state has a normal distribution.

5 Normally Distributed State and Linear Disclosure Rule

In this section, we characterize the optimal disclosure rule when the state has a normal distribution. Specifically, we suppose that $\mathbf{x} \sim N(\mathbf{0}, \Sigma)$ where Σ is symmetric and positive definite.

A *linear rule of rank l* is a deterministic disclosure rule such that $m = g(\mathbf{x})$ is a linear transformation of rank l.[22] For $l \geq 1$, we can represent a linear rule by $g(\mathbf{x}) = A'\mathbf{x} \in \mathbb{R}^l$ where A is a $k \times l$ matrix of rank l.[23] We can interpret a linear rule of rank l as a rule publicizing l variables $(m_1, \ldots, m_l)$ none of which is redundant. Note that full disclosure is a linear rule of rank k such as $g(\mathbf{x}) = \mathbf{x}$ and no disclosure rank zero such as $g(\mathbf{x}) = 0$. Since

[22] That is, the dimension of the range of g equals l.

[23] If g maps each realization of the state into $\mathbb{R}^L$ with $L > l$, there are $L - l$ redundant variables, say $(m_{l+1}, \ldots, m_L)$, in the sense that $\mathbb{E}[\mathbf{x}|m_1, \ldots, m_L] = \mathbb{E}[\mathbf{x}|m_1, \ldots, m_l]$. Thus we can represent the linear rule of rank l by a $k \times l$ matrix A without loss of generality.

we know that no disclosure is optimal if and only if V is negative semidefinite, so we suppose that V is not negative semidefinite.

When the state is normally distributed and a linear rule of rank $l \geq 1$ is chosen, the message $\mathbf{m} = A'\mathbf{x} \in \mathbb{R}^l$ has a normal distribution with zero mean. The standard result of the normal distribution gives the conditional expectation $\hat{\mathbf{x}} = \mathbb{E}[\mathbf{x}|\mathbf{m}] = \Sigma A(A'\Sigma A)^{-1}\mathbf{m}$.[24] Since $\mathbf{m} = A'\mathbf{x}$, we have

$$\hat{\mathbf{x}} = \Sigma A(A'\Sigma A)^{-1}A'\mathbf{x}.$$

Intuitively, a linear rule of rank l projects the realization of the state onto an l dimensional subspace in which the estimates are distributed. It follows from the analogue of Theorem 1 that the optimal disclosure rule must induce a distribution of conditional expectations such that V is positive semidefinite on its support. That is, $\hat{\mathbf{x}}'V\hat{\mathbf{x}} \geq 0$ for every $\hat{\mathbf{x}}$ in its support. Otherwise, we can reduce unfavorable variability of conditional expectations in a similar manner to the *only if* part of Theorem 1 (i). This necessary condition effectively narrows the set of potential solutions.

Let $B \equiv \Sigma^{\frac{1}{2}}A$ and $P_B = B(B'B)^{-1}B'$. The matrix P_B is an orthogonal projection matrix that maps vectors in $\mathbb{R}^k$ onto the column space of B. Then the estimates are $\hat{\mathbf{x}} = \Sigma^{\frac{1}{2}}P_B\Sigma^{-\frac{1}{2}}\mathbf{x}$, and hence the second moment $\hat{\Sigma}$ is written as

$$\hat{\Sigma} = \mathbb{E}\hat{\mathbf{x}}\hat{\mathbf{x}}' = \Sigma^{\frac{1}{2}}P_B\Sigma^{\frac{1}{2}}.$$

Thus the gain of the disclosure rule is written as $\mathrm{tr}(V\hat{\Sigma}) = \mathrm{tr}(V\Sigma^{\frac{1}{2}}P_B\Sigma^{\frac{1}{2}}) = \mathrm{tr}(WP_B)$ where $W = \Sigma^{\frac{1}{2}}V\Sigma^{\frac{1}{2}}$.

Consider a linear rule $\mathbf{m} = A'\mathbf{x}$ such that $A = \Sigma^{-\frac{1}{2}}Q_+$ where, as denoted in Theorem 2, $Q_+ = [\mathbf{q}_1, \ldots, \mathbf{q}_r]$ is the eigenvectors associated with the nonnegative eigenvalues of W. Then $B = \Sigma^{\frac{1}{2}}A = Q_+$ and $\mathrm{tr}(WP_B) = \mathrm{tr}(WP_{Q_+})$, which equals the upper bound identified in Theorem 2. Thus, we find a linear rule that is optimal among the general class of disclosure rules.

Theorem 3 *Suppose that the state is normally distributed, that is $\mathbf{x} \sim N(\mathbf{0}, \Sigma)$. Then a linear rule $g(\mathbf{x}) = Q_+'\Sigma^{-\frac{1}{2}}\mathbf{x}$ is optimal where $Q_+ = [\mathbf{q}_1, \ldots, \mathbf{q}_r]$ is the eigenvectors associated*

[24] Consider the joint distribution $(\mathbf{x}, \mathbf{m}) \sim N(\tilde{\mu}, \tilde{\Sigma})$ where the mean vector and variance matrix can be partitioned as

$$\tilde{\mu} = \begin{pmatrix} \tilde{\mu}_{\mathbf{x}} \\ \tilde{\mu}_{\mathbf{m}} \end{pmatrix}, \quad \tilde{\Sigma} = \begin{pmatrix} \tilde{\Sigma}_{\mathbf{x},\mathbf{x}} & \tilde{\Sigma}_{\mathbf{x},\mathbf{m}} \\ \tilde{\Sigma}_{\mathbf{m},\mathbf{x}} & \tilde{\Sigma}_{\mathbf{m},\mathbf{m}} \end{pmatrix}.$$

It follows (see, for example Vives (2008)) then the conditional density of $\mathbf{x}$ given $\mathbf{m}$ is normal with conditional mean $\tilde{\mu}_{\mathbf{x}} + \tilde{\Sigma}_{\mathbf{x},\mathbf{m}}\tilde{\Sigma}_{\mathbf{m},\mathbf{m}}^{-1}(\mathbf{m} - \tilde{\mu}_{\mathbf{m}})$ and variance matrix $\tilde{\Sigma}_{\mathbf{x},\mathbf{x}} - \tilde{\Sigma}_{\mathbf{x},\mathbf{m}}\tilde{\Sigma}_{\mathbf{m},\mathbf{m}}^{-1}\tilde{\Sigma}_{\mathbf{m},\mathbf{x}}$. Now apply to $\mathbf{m} = A'\mathbf{x}$, we have $\tilde{\Sigma}_{\mathbf{x},\mathbf{m}} = \mathbb{E}[\mathbf{x}\mathbf{x}'A] = \Sigma A$ and $\tilde{\Sigma}_{\mathbf{m},\mathbf{m}} = \mathbb{E}[A'\mathbf{x}\mathbf{x}'A] = A'\Sigma A$.

with the nonnegative eigenvalues of the symmetric matrix $W = \Sigma^{\frac{1}{2}} V \Sigma^{\frac{1}{2}}$.

The optimal linear rule in Theorem 3 is interpreted as the following information processing. First, the sender adjusts the variance of the state as $\mathbf{y} \equiv \Sigma^{-\frac{1}{2}}\mathbf{x}$ so that $\text{var}(\mathbf{y}) = I_k$. Second, r variables $(m_1, \ldots, m_r)$ is disclosed each of which is a linear combination $m_i = q_{1i}y_1 + \cdots + q_{ki}y_k$ of $\mathbf{y}$ where the weights $(q_{1i}, \ldots, q_{ki})$ constitute an eigenvector associated with a nonnegative eigenvalue of W.[25]

Here we briefly discuss the role of two key properties in Theorem 3: the normality of the state and the linearity of the disclosure rule. The first remark is that there exists $Z \in \{\tilde{Z} : I \succeq \tilde{Z} \succeq O\}$ such that it cannot be induced by the linear rule. Recall that for any linear rule A (translated into $B = \Sigma^{\frac{1}{2}}A$), we have $Z = P_B = B(B'B)^{-1}B'$. A property of orthogonal projection matrices (that is, symmetric and idempotent) is that every eigenvalue is either zero or one. Therefore, the linear rule cannot induce Z such that it has an eigenvalue in $(0, 1)$. Second, when the normality fails, the linear rule is no longer able to achieve the upper bound identified in Theorem 2. For example, consider $V = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ and $x_i \overset{\text{iid}}{\sim} U[-1/2, 1/2]$ for $i = 1, 2$ so that $\Sigma = \frac{1}{12}I$ and $W = \frac{1}{12}V$. Then the upper bound is given by the positive eigenvalue of W, which equals $(1+\sqrt{5})/24 \approx 0.1348$, while the linear rule yields at most $\frac{86}{27} \cdot \frac{1}{24} \approx 0.1327$ under $g(x_1, x_2) = 3x_1 + 2x_2$.

5.1 The Two-dimensional Normal State

We now apply the theory developed above to obtain the optimal disclosure rule under the normally distributed state for $k = 2$. Let $\hat{v}(\hat{\mathbf{x}}) = \hat{\mathbf{x}}'V\hat{\mathbf{x}}$ and $(x_1, x_2) \sim N((0, 0), \Sigma)$. From Corollary 1, we focus on the case where V is indefinite (i.e., neither positive nor negative semidefinite) and find the optimal linear rule A of rank 1 that maximizes $\text{tr}(V\hat{\Sigma})$.[26]

We normalize the state and the estimates by $\mathbf{y} = \Sigma^{-\frac{1}{2}}\mathbf{x}$ and $\hat{\mathbf{y}} = \Sigma^{-\frac{1}{2}}\hat{\mathbf{x}}$. First we see how an orthogonal projection P_B determines the distribution of $\hat{\mathbf{y}}$. For $B = (b_1, b_2)' \in \mathbb{R}^2$, we have $\hat{\mathbf{y}} = P_B\mathbf{y}$, or equivalently

$$\begin{cases} \hat{y}_1 = \frac{b_1 y_1 + b_2 y_2}{b_1^2 + b_2^2} b_1 \\ \hat{y}_2 = \frac{b_1 y_1 + b_2 y_2}{b_1^2 + b_2^2} b_2. \end{cases}$$

[25] It is worth noting that if the rank of W is less than k, say $k - n$, then the rank of optimal linear rule is indeterminate. Formally, letting $Q_{++} = [\mathbf{q}_1, \ldots, \mathbf{q}_{r-n}]$ be the eigenvectors associated with positive eigenvalues, any linear rule $A = \Sigma^{-\frac{1}{2}}B$ such that B contains every column of Q_{++} but orthogonal to Q_- attains the same value. Hence the rank of an optimal linear rule may be $r - n$ or more but must be less than or equal to r.

[26] If V is positive (negative) semidefinite, full (no) disclosure which corresponds to a linear rule of rank 2 (0, respectively) is optimal.

Thus we find that $(\hat{y}_1, \hat{y}_2)$ are distributed on the line $b_2\hat{y}_1 = b_1\hat{y}_2$. Moreover the variance matrix of $(\hat{y}_1, \hat{y}_2)$ is $\text{var}(\hat{\mathbf{y}}) = \mathbb{E}[P_B \mathbf{y}\mathbf{y}' P_B'] = P_B$, or equivalently

$$\text{var}(\hat{y}_1, \hat{y}_2) = \begin{pmatrix} c_1^2 & c_1c_2 \\ c_1c_2 & c_2^2 \end{pmatrix}$$

where $(c_1, c_2) = (\frac{b_1}{\sqrt{b_1^2+b_2^2}}, \frac{b_2}{\sqrt{b_1^2+b_2^2}})$ is a point on the unit sphere in $\mathbb{R}^2$. This implies that the sender's choice variable is essentially one-dimensional, and hence the optimization problem can be solved by standard calculus.

Corollary 3 *Suppose that V is indefinite and $W = \Sigma^{\frac{1}{2}} V \Sigma^{\frac{1}{2}}$. Then the optimal linear rule is $A = \Sigma^{-\frac{1}{2}} B$ where $B = (b_1, b_2)' \in \mathbb{R}^2$ is such that:* (i) *if $w_{12} = 0$, then $B = (1, 0)'$ for $w_{22} < 0 < w_{11}$, and $B = (0, 1)'$ for $w_{11} < 0 < w_{22}$;* (ii) *if $w_{12} = w_{21} \neq 0$, then*

$$\frac{b_1}{b_2} = \frac{(w_{11} - w_{22}) + \sqrt{(w_{11} - w_{22})^2 + 4w_{12}^2}}{2w_{12}}.$$

Corollary 3 characterizes the optimal linear rule when $k = 2$ as in the motivating example and in Rayo and Segal (2010). In contrast to the finite state space case, the optimal disclosure rule is deterministic and linear if the state has a bivariate normal distribution. For example, for $x_i \sim N(0, \sigma_i^2)$ for $i = 1, 2$, the solution to the Rayo and Segal (2010)'s problem is the linear rule $g(x_1, x_2) = \sigma_1^{-1} x_1 + \sigma_2^{-1} x_2$.

6 Applications

6.1 Optimal Advertising Policy

Suppose that a monopoly firm chooses an information disclosure policy about its new product. In particular, the firm observes its cost shock x_c and quality shock x_a. Assume that random variables x_a and x_c are independent and normally distributed with means zero and variances σ_a^2 and σ_c^2, respectively. A disclosure rule (α, M) determines information m available to consumers. For example, a computer manufacturer discloses various information about its product including display resolution, battery life, processing speed, and so on. A production company releases on-line free music/movie clips. By controlling information revealed, the firm can induce a preferred distribution of the consumers' conditional expectations about the product quality and production cost.

Suppose that a representative consumer has a quadratic utility function $u(q, x_a) = \left((a + x_a)q - \frac{1}{2}q^2\right) - pq$ where q is the quantity consumed and p is the unit price. It follows that the inverse demand function given message m is $q = a + \hat{x}_a - p$ where $\hat{x}_a = \mathbb{E}[x_a|m]$ is the consumer's conditional expectation about the quality shock. The firm's profit function is $v(q, x_c) = pq - (c + x_c)q$. For simplicity, we assume that the price is exogenously fixed.[27] The timing of the game is as follows. First, the firm commits to a disclosure rule (α, M). Second, the firm observes the realization of the state (x_a, x_c) and discloses information m. Third, the consumer estimates the product quality and determines the demand quantity.

The firm's expected profit is written as

$$\begin{aligned}\mathbb{E}v =& \mathbb{E}\left[(p - c - x_c)(a + \hat{x}_a - p)\right] \\ =& -\mathbb{E}\hat{x}_a\hat{x}_c + (p - c)(a - p).\end{aligned}$$

Thus, we have $\hat{v}(\hat{x}_a, \hat{x}_c) = -\hat{x}_a\hat{x}_c$. An immediate implication is that the firm's expected profit is a decreasing function of the covariance between $\hat{x}_a$ and $\hat{x}_c$. Intuitively, the firm is better off increasing the probability that demand expand when it has a lower cost. The expected value of social welfare (i.e., consumer surplus plus the firm's profit) is

$$\begin{aligned}\mathbb{E}w =& \mathbb{E}\left[\left((a + x_a)q - \frac{1}{2}q^2\right) - (c + x_c)q\right] \\ =& \mathbb{E}\left[\frac{1}{2}\hat{x}_a^2 - \hat{x}_a\hat{x}_c\right] + (a - p)\left(\frac{a + p}{2} - c\right)\end{aligned}$$

so that the socially optimal disclosure rule maximizes the expected value of a quadratic function $\hat{w}(\hat{x}_a, \hat{x}_c) = \frac{1}{2}\hat{x}_a^2 - \hat{x}_a\hat{x}_c$.

From Corollary 3, the optimal disclosure rule for the firm is $g^P(x_a, x_c) = \sigma_c x_a - \sigma_a x_c$ while the socially optimal disclosure rule is $g^S(x_a, x_c) = \kappa x_a - \sigma_a x_c$ where

$$\kappa = \frac{\sigma_a + \sqrt{\sigma_a^2 + 4\sigma_c^2}}{2} > \sigma_c.$$

The coefficient on x_a is interpreted as the amount of information revealed about the product quality.

Proposition 1 *The monopoly advertisements contain less information about the product quality than the socially optimal advertisements.*

Lewis and Sappington (1994) examine the amount of information a monopoly firm might

[27]Milgrom and Roberts (1986) analyze a model in which the firm chooses its price and consumers draw product-quality inferences from price as well as advertisement. Although such a "signaling effect" of action is important in a number of different contexts such as monetary policy (see, Baeriswyl and Cornand (2010)), it requires different techniques and is beyond the scope of the present paper.

provide to potential buyers. In their setting, each buyer privately observes an imperfect signal about his valuation and the firm controls the precision of the signals. Anderson and Renault (2006) and Johnson and Myatt (2006) distinguish the informational content of advertisements (e.g., price vs. attributes and hype vs. real information, respectively). In our model, the firm should disclose a one-dimensional index constructed from its product quality and marginal cost and control the variances and covariance of the "market expectations."

6.2 Campaign Advertising and Incumbency Advantage

The population consists of two groups of voters, indexed by $i \in \{1, 2\}$. These groups differ in their policy preferences over a one-dimensional policy space. Let $q_1 = -\frac{1}{2}$ and $q_2 = \frac{1}{2}$ be the preferred policies of groups 1 and 2, respectively. As in Prat (2002), voters also judge candidates in another dimension, say *valence*, which represents the general competence of a candidate such as negotiating ability, leadership, and integrity. Unlike policy preferences, all voters' preferences are the same in the valence dimension.

Two parties compete against each other in an election. An incumbent runs from the ruling party and a challenger from the opposition party. The ruling party is privately informed about the characteristics of the incumbent and makes campaign advertising that may reveal information about him. The incumbent is characterized by two parameters (x, y) where $x \in [-\frac{1}{2}, \frac{1}{2}]$ represents his policy stance and $y \in [-\frac{1}{2}, \frac{1}{2}]$ represents his valence. Assume that x and y have zero means and are independent of each other. The *ex ante* distribution of (x, y) is common knowledge, but its realization is observed only by the ruling party. We address the optimal campaign policy for the ruling party that maximizes the probability of reelection in the absence of the opposition party's campaign.[28] A possible interpretation of the campaign strategy is such that the party chooses topics discussed in a meeting and in candidates' speeches.

When the incumbent of type (x, y) is elected, voters in group $i \in \{1, 2\}$ receive utility $u_i = -|x - q_i| + y$. On the other hand, when the challenger is elected, they receive $u_i = -|0 - q_i| + t_i$ where t_i is a private information of group i that represents an ideological bias toward the challenger. We assume that t_i is independent of t_{-i} and is uniformly distributed over $[-\frac{1}{2h}, \frac{1}{2h}]$ for a sufficiently small $h > 0$.[29] Thus, given m and t_i, voters in group i vote for the ruling party if $\mathbb{E}[u_i(x, y)|m] \geq -|0 - q_i| + t_i$.

The timing of the game is summarized as follows. The ruling party commits to a disclosure rule. The ruling party observes the incumbent's type (x, y) and publicly discloses a message

[28] Information disclosure by multiple senders raises a new issue and is beyond the scope of the paper. A recent work of Gentzkow and Kamenica (2011) tackles such a problem.

[29] For $h \in (0, \frac{1}{2})$, the interior solution is guaranteed.

m according to the disclosure rule. The noise variables (t_1, t_2) are realized. Given m and t_i, each voter votes for the candidate he prefers. If both groups vote for the same candidate, he wins with probability one. If two groups disagree, then the incumbent wins with probability $\psi \in [0, 1]$.

Let $\hat{x} = \mathbb{E}[x|m]$ and $\hat{y} = \mathbb{E}[y|m]$. The voters' expected utility from the incumbent conditional on m is written as

$$\mathbb{E}[u_i(x, y)|m] = \begin{cases} -(\frac{1}{2} + \hat{x}) + \hat{y} & \text{for group 1} \\ -(\frac{1}{2} - \hat{x}) + \hat{y} & \text{for group 2} \end{cases}$$

The probability P_1 (P_2) that voters in group 1 (group 2) vote for the incumbent is given by $P_1 = \frac{1}{2} + h(\hat{y} - \hat{x})$ ($P_2 = \frac{1}{2} + h(\hat{y} + \hat{x})$, respectively). The conditional probability $P(m)$ that the incumbent wins given message m is

$$\begin{aligned} P(m) =& P_1 P_2 + \psi(1 - P_1)P_2 + \psi P_1(1 - P_2) \\ =& \frac{1 + 2\psi}{4} + h\hat{y} + (1 - 2\psi)h^2(\hat{y} + \hat{x})(\hat{y} - \hat{x}). \end{aligned}$$

Thus, we have $\hat{v}(\hat{x}, \hat{y}) = (1 - 2\psi)h^2(\hat{y}^2 - \hat{x}^2)$. Since we assume that x and y are independent, we find the following result.[30]

Proposition 2 *When the incumbent has an advantage ($\psi \geq \frac{1}{2}$), then it is optimal for the ruling party to reveal only the incumbent's policy stance ($g(x, y) = x$).*

Intuitively, when the incumbent has an advantage, the ruling party has an incentive to increase the probability that at least one group prefers the incumbent to the challenger even though it decreases the probability of unanimity. Consequently, the incumbency advantage impairs the selection of a competent candidate through the campaign strategy that reveals no information about the valence characteristics. We also predict that the opposition party needs to attract both groups of voters and has an incentive to reveal the valence dimension of the candidates; for example, revealing scandals involving the incumbent and emphasizing his inconsistent statements.

Similar situations arise in different contexts. For example, in a criminal court, a defense attorney who needs to persuade only a part of juries is better off making an emotional appeal to them while a prosecutor who needs to avoid a conflict among juror is better off gathering objective evidence of guilt. In this case, the voting procedure determines the incentives of information revelation by the prosecutor and the attorney.

[30]Note that the result does not depend on the marginal distributions of x and y.

Polborn and Yi (2006) analyze information revelation in a political campaign assuming that each candidate must truthfully reveal either positive or negative information. Coate (2004) and Galeotti and Mattozzi (2011) analyze informative campaign which perfectly reveals the candidate's policy position to a fraction of voters while Prat (2002) analyzes campaign advertising when the campaign expenditures signal the candidate's valence.

6.3 Central Bank Transparency

We examine how central bank transparency affects the volatility of the output gap and inflation and characterize the optimal disclosure rule and monetary policy. As in Geraats (2002) and Jensen (2002), we consider a simple two-period model where period 1 is regarded as the present and period 2 as the future.

The private sector behavior is summarized by a standard Phillips curve[31]

$$l_t = \pi_t - \mathbb{E}^P_{t-1}\pi_t + \varepsilon_t$$

where l_t is (log) employment in period t, π_t is the inflation rate in period t (the change in the log price level between period $t-1$ and t), and ε_t is an employment shock (a supply shock). The expectation operator $\mathbb{E}^P_{t-1}[\cdot]$ denotes the market expectation formed in period $t-1$.

The central bank has perfect control over inflation $\pi_t = i_t$ where i_t is the central bank's intended inflation.[32] We assume that the central bank can commit to a contingent monetary policy in the short-run, but cannot commit to the future policy. For example, career concerns of the policymakers may prevent discretionary policymaking in the short-run, but in the future the composition of the policymaker board may alter and an alternative policy plan may be chosen. Moreover, an unpredictable change in economic and political conditions may make the initial plan totally inadequate.

The central bank's loss function is $\mathbb{E}[L_1 + \beta L_2]$ where $\beta \in (0, 1)$ is the discount factor and L_t is the period t loss function

$$L_t = \pi_t^2 + \lambda(l_t - l_t^*)^2$$

for some $\lambda > 0$. The employment target l_t^* can be interpreted as a demand shock due to stochastic preferences of the representative household or as the central bank's preference shock due to a change in the degree of central bank independence.

[31]The description of the economy is based on Faust and Svensson (2001).

[32]Faust and Svensson (2001) assume that the central bank has *imperfect* control over inflation so that $\pi_t = i_t + \eta_t$ where η_t is a control error.

We assume that the supply shock and the employment target are independent and normally distributed with mean zero. In particular, we assume that $l_1^* \sim N(0, \sigma_l^2)$ and $\varepsilon_1 \sim N(0, \sigma_\varepsilon^2)$, and that these shocks evolve according to $\varepsilon_{t+1} = \rho_\varepsilon \varepsilon_t + \xi_{t+1}$ and $l_{t+1}^* = \rho_l l_t^* + \zeta_{t+1}$ where ξ_{t+1} and ζ_{t+1} are independent shocks with mean zero and $\rho_l, \rho_\varepsilon \in (-1, 1)$.

The timing of events is as follows. The central bank chooses a disclosure rule (α, M) and a short-run monetary policy plan $i_1 : \mathbb{R}^2 \times M \to \mathbb{R}$ which depends on the realization of the supply and demand shocks and the message disclosed. The state of nature (l_1^*, ε_1) is realized and a message $m \in M$ is publicly announced according to (α, M). The central bank sets a short-run monetary policy $i_1(l_1^*, \varepsilon_1, m)$, and the inflation rate π_1 is determined. Given m and $i_1(l_1^*, \varepsilon_1, m)$, the private sector forms an expectation about the future inflation rate $\mathbb{E}_1^P[\pi_2]$. In period 2, the central bank chooses a policy i_2 given the realization of (l_2^*, ε_2) and the market expectation.

In period 2, the central bank's problem is given by

$$\begin{aligned} &\min_{i_2} \ \pi_2^2 + \lambda(l_2 - l_2^*)^2 \\ &\text{subject to } l_2 = \pi_2 - \mathbb{E}_1^P \pi_2 + \varepsilon_2 \\ &\pi_2 = i_2. \end{aligned}$$

From the first-order condition and the rational expectation, we have

$$\begin{aligned} \pi_2 =& \frac{\lambda}{1+\lambda} \left[\lambda(\hat{l}_2^* - \hat{\varepsilon}_2) + (l_2^* - \varepsilon_2) \right] \\ l_2 - l_2^* =& -\frac{1}{1+\lambda} \left[\lambda(\hat{l}_2^* - \hat{\varepsilon}_2) + (l_2^* - \varepsilon_2) \right] \end{aligned}$$

where $\hat{y} \equiv \mathbb{E}_1^P y$ denotes the market expectation of a random variable y formed in period 1. The loss in period 2 is given by

$$L_2 = \frac{\lambda}{1+\lambda} \left[\lambda(\hat{l}_2^* - \hat{\varepsilon}_2) + (l_2^* - \varepsilon_2) \right]^2 .$$

We now consider the short-run monetary policy and optimal disclosure rule, which solve the following problem

$$\begin{aligned} &\min_{i_1(\cdot),(\alpha,M)} \mathbb{E}[\pi_1^2 + \lambda(l_1 - l_1^*)^2] + \beta \mathbb{E} L_2(\hat{l}_2^*, \hat{\varepsilon}_2, l_2^*, \varepsilon_2) \\ &\text{subject to } l_1 = \pi_1 - \mathbb{E}_0^P \pi_1 + \varepsilon_1 \\ &\pi_1 = i_1(l_1^*, \varepsilon_1, m). \end{aligned}$$

Note that we can restrict our search for optimal short-run policy plans to the class of m-measurable functions without loss of generality. To see this, consider a disclosure rule (α, M) and a short-run policy plan $i_1 : \mathbb{R}^2 \times M \to \mathbb{R}$. The private sector observes m and $i_1(l_1^*, \varepsilon_1, m)$, and forms an expectation $(\hat{l}_2^*, \hat{\varepsilon}_2)$. Now consider a disclosure rule $(\tilde{\alpha}, M \times \mathbb{R})$ where the message is given by $\tilde{m} = (m, i_1(l_1^*, \varepsilon_1, m))$, and a $\tilde{m}$-measurable policy $\tilde{i}_1(\tilde{m}) = i_1(l_1^*, \varepsilon_1, m)$. This pair of a disclosure rule and a policy plan is essentially identical to the initial pair $((\alpha, M), i_1)$ in the sense that the information revealed to the private sector and the short-run monetary policy in period 1 are the same almost surely. Therefore we first characterize the optimal short-run policy plan given each disclosure rule, and then find the optimal disclosure rule.

Fix a disclosure rule (α, M). From the first-order condition, the optimal short-run policy $i_1(\cdot)$ is given by

$$i_1(\hat{l}_1^*, \hat{\varepsilon}_1) = \frac{\lambda}{1+\lambda}(\hat{l}_1^* - \hat{\varepsilon}_1),$$

and the *ex ante* expected loss is

$$\begin{aligned}\mathbb{E}L_1 + \beta\mathbb{E}L_2 =& \lambda\mathbb{E}(l_1^* - \varepsilon_1)^2 - \frac{\lambda^2}{1+\lambda}\mathbb{E}(\hat{l}_1^* - \hat{\varepsilon}_1)^2 \\ &+ \beta\left[\frac{\lambda}{1+\lambda}(l_2^* - \varepsilon_2)^2 + \frac{\lambda^2(2+\lambda)}{1+\lambda}\mathbb{E}(\hat{l}_2^* - \hat{\varepsilon}_2)^2\right].\end{aligned}$$

Recall that $\hat{l}_2^* = \mathbb{E}_1^P[l_2^*] = \rho_l \hat{l}_1^*$ and $\hat{\varepsilon}_2^* = \mathbb{E}_1^P[\varepsilon_2^*] = \rho_\varepsilon \hat{\varepsilon}_1^*$. Then we have

$$V = \frac{\lambda^2}{1+\lambda}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} - \frac{\beta\lambda^2(2+\lambda)}{1+\lambda}\begin{pmatrix} \rho_l^2 & -\rho_l\rho_\varepsilon \\ -\rho_l\rho_\varepsilon & \rho_\varepsilon^2 \end{pmatrix}. \tag{5}$$

Since $\det(V) = -\frac{\beta(2+\lambda)\lambda^4}{(1+\lambda)^2}(\rho_l - \rho_\varepsilon)^2 \leq 0$, the following statement holds.

Proposition 3 *A linear rule of rank 1 is optimal whenever $\rho_l \neq \rho_\varepsilon$.*

The central bank needs to respond to the shocks (l_1^*, ε_1) to stabilize the output gap. On the other hand, it should avoid information revelation about the future policy since it weakens the policy effectiveness in the future. This trade-off makes partial revelation optimal in the generic case where $\rho_l \neq \rho_\varepsilon$.

To make this point clear, suppose that $\rho_l = 1$ and $\rho_\varepsilon = 0$. Then

$$W = \Sigma^{\frac{1}{2}} V \Sigma^{\frac{1}{2}} = \frac{\lambda^2}{1+\lambda}\begin{pmatrix} \sigma_l^2(1 - \beta(2+\lambda)) & -\sigma_l\sigma_\varepsilon \\ -\sigma_l\sigma_\varepsilon & \sigma_\varepsilon^2 \end{pmatrix}.$$

From Corollary 3, the optimal linear rule $g(l_1^*, \varepsilon_1) = \kappa l_1^* - \varepsilon_1$ where $\kappa \in (0,1)$ is a decreasing function of $\beta(2+\lambda)$ and $(\sigma_\varepsilon/\sigma_l)$.[33] As β increases, the information revelation about l_1^* becomes more costly and then the amount of information contained in the message, which can be measured by κ, should decrease.[34]

Given the optimal linear rule, the short-run policy is written as

$$i_1 = \frac{\lambda}{1+\lambda}(\hat{l}_1^* - \hat{\varepsilon}_1) = \frac{\lambda}{1+\lambda}\frac{\kappa\sigma_l^2 + \sigma_\varepsilon^2}{\kappa^2\sigma_l^2 + \sigma_\varepsilon^2}(\kappa l_1^* - \varepsilon_1). \tag{6}$$

As κ increases, the stabilization policy becomes more responsive to l_1^*, which increases $\mathbb{E}\pi_1^2$ and decreases $\mathbb{E}(l_1 - l_1^*)^2$. Intuitively, as the central bank becomes more myopic, the cost from the output stabilization due to information revelation becomes less important. This comparative statics is summarized as follows.[35]

Proposition 4 *As the central bank becomes myopic, the output gap is stabilized while the inflation becomes volatile.*

As illustrated above, our framework is useful to characterize the optimal monetary policy that plays a signaling role as well as the stabilization role.[36] Note that the optimal short-run policy plan i_1 is a linear function of the message $m = \kappa l_1^* - \varepsilon_1$ (see (6)). This implies that the policy outcome $i_1(m)$ contains the same information as the message, and hence the optimal disclosure rule and monetary policy are also implemented by committing to the optimal short-run policy plan and making the policy outcome transparent.

A number of papers (e.g., Faust and Svensson (2001) and Jensen (2002) among others) investigate the welfare effect of central bank transparency and the optimal monetary policy in different transparency regimes. Unlike these papers which restrict communication strategies available to the central bank to noisy communications, we characterize the optimal disclosure rule and monetary policy plan in the general class of policies. Our analysis suggests that the

[33] An exact expression for κ is

$$\begin{aligned}\kappa =& 1 - \frac{1}{2}(1 + (\sigma_\varepsilon/\sigma_l)^2 + \beta(2+\lambda)) \\ &+ \frac{1}{2}\sqrt{(1 + (\sigma_\varepsilon/\sigma_l)^2 + \beta(2+\lambda))^2 - 4\beta(2+\lambda)}.\end{aligned}$$

[34] The amount of information about the supply shock revealed is measured by the variability of the conditional expectation of l_1^*, $\mathrm{var}(\mathbb{E}[l_1^*|m]) = \kappa^2\sigma_l^4/(\kappa^2\sigma_l^2 + \sigma_\varepsilon^2)$, which is increasing in $\kappa \in (0,1)$.

[35] In a similar manner, one can examine how the policy maker's preferences, parameterized by λ, affect the monetary policy and central bank transparency.

[36] Recently, Baeriswyl and Cornand (2010) study this problem in the setting where the monetary policy is imperfectly observed by the private sector.

central bank should control the covariance of market expectations rather than the variance of private sector forecast errors of each variable, written as $\mathrm{var}(\varepsilon - \hat{\varepsilon})$ and $\mathrm{var}(l^* - \hat{l}^*)$.

7 Conclusion

We study multidimensional information disclosure where the sender's expected utility is expressed as the expected value of a function of the receivers' expectations of the state. The semidefinite programming is applied to identifying necessary conditions for the second moment of the conditional expectations that can be induced by the disclosure rule and characterizing an upper bound of the sender's expected utility. We characterize the optimal disclosure rule among the general class of (possibly randomized) rules as a linear transformation of the state when it is normally distributed. Based on such a simple and tractable characterization, we study several applications and provide interesting implications. Possible directions for future work include studying settings with multiple senders and with receivers' private information.

Appendix

A Proofs

Proof of Theorem 1. The *if* parts (the optimality of full/no disclosure when V is definite) follow from Jensen's inequality and the *only if* parts (the optimality of partial disclosure when V is indefinite) are shown by constructing a rule that yields a higher gain than full/no disclosure.

(i): *if part.* Suppose that V is positive semidefinite. We will show that full disclosure is optimal. If V is positive semidefinite, or equivalently if $\hat{v}$ is a convex function, then, for any disclosure rule (α, M), we have from Jensen's inequality that $\mathbb{E}\hat{v}(\hat{\mathbf{x}}) \leq \mathbb{E}[\mathbb{E}[\hat{v}(\mathbf{x})|m]] = \mathbb{E}\hat{v}(\mathbf{x})$ but the last is equal to the gain under full disclosure. Hence full disclosure is optimal.

(ii): *if part.* Suppose that V is negative semidefinite. We will show that no disclosure is optimal. If V is negative semidefinite, or equivalently if $\hat{v}$ is a concave function, then, for any disclosure rule (α, M), we have $\mathbb{E}\hat{v}(\hat{\mathbf{x}}) \leq \hat{v}(\mathbb{E}\hat{\mathbf{x}}) = \hat{v}(\mathbb{E}\mathbf{x})$ but the last is equal to the gain under no disclosure. Hence no disclosure is optimal.

(i): *only if part.* Suppose that V is indefinite. We will construct a partial disclosure that attains a higher gain than full disclosure. Since V is not positive semidefinite, there exists a vector $\mathbf{x}_- \in \mathbb{R}^k$ such that $\mathbf{x}_-' V \mathbf{x}_- < 0$. Let $L_{\mathbf{x}_-}$ be the linear subspace in $\mathbb{R}^k$ spanned by a vector $\mathbf{x}_-$ and $L^\perp_{\mathbf{x}_-}$ be the orthogonal complement of $L_{\mathbf{x}_-}$. Any point in $\mathbb{R}^k$ is represented by the sum of vectors in $L_{\mathbf{x}_-}$ and $L^\perp_{\mathbf{x}_-}$, as $\mathbf{x} = \mathbf{y} + \gamma\mathbf{x}_-$ for some $\gamma \in \mathbb{R}$ and $\mathbf{y} \in L^\perp_{\mathbf{x}_-}$.

Now consider a disclosure rule (α, M) such that the sender discloses vector $\mathbf{y} \in L^\perp_{\mathbf{x}_-}$ for each realization of $\mathbf{x}$. Under this rule, the receivers know on which line $\mathbf{x}$ is realized, but they are still uninformed about $\gamma \in \mathbb{R}$. Thus $\mathbf{x}|m$ is distributed over a line through $\hat{\mathbf{x}}$ parallel to $\mathbf{x}_-$, and hence we can write $\mathbf{x}|m - \hat{\mathbf{x}}$ as $\gamma\mathbf{x}_-$ where $\gamma \in \mathbb{R}$ is a corresponding random variable.

Now compare the gain $\mathbb{E}\hat{v}(\mathbf{x})$ under full disclosure with the gain $\mathbb{E}\hat{v}(\hat{\mathbf{x}})$ under rule (α, M):

$$\begin{aligned}
\mathbb{E}[\mathbf{x}'V\mathbf{x}] - \mathbb{E}[\hat{\mathbf{x}}'V\hat{\mathbf{x}}] =& \mathbb{E}[\mathbb{E}[\mathbf{x}'V\mathbf{x}|m] - \mathbb{E}[\hat{\mathbf{x}}'V\hat{\mathbf{x}}]] \\
=& \mathbb{E}_m\left[\mathbb{E}[(\mathbf{x}-\hat{\mathbf{x}})'V(\mathbf{x}-\hat{\mathbf{x}})|m]\right] \\
=& \mathbb{E}_m\left[\mathbb{E}[\gamma^2\mathbf{x}_-'V\mathbf{x}_-|m]\right] \\
=& \mathbf{x}_-'V\mathbf{x}_-\mathbb{E}\gamma^2 < 0
\end{aligned}$$

where the inequality holds since $\gamma \neq 0$ almost surely.[37]

(ii): *only if part.* Suppose that V is indefinite. We will construct a partial disclosure that attains a higher gain than no disclosure, which equals zero. Let $\mathbf{s} \in \mathcal{S}^{k-1}$ be a point in the unit sphere in $\mathbb{R}^k$ and consider a disclosure rule under which the sender discloses the sign of $\mathbf{s}'\mathbf{x} \in \mathbb{R}$.[38] Let $\hat{\mathbf{x}}_p = \mathbb{E}[\mathbf{x}|\mathbf{s}'\mathbf{x} \geq 0]$ and $\hat{\mathbf{x}}_n = \mathbb{E}[\mathbf{x}|\mathbf{s}'\mathbf{x} < 0]$.

Since V is not negative semidefinite, there exists $\mathbf{x}_+ \in \mathbb{R}^k$ such that $\mathbf{x}_+'V\mathbf{x}_+ > 0$. We want to show that there exists $\mathbf{s} \in \mathcal{S}^{k-1}$ such that $\hat{\mathbf{x}}_p = \gamma_p^s\mathbf{x}_+$ and $\hat{\mathbf{x}}_n = \gamma_n^s\mathbf{x}_+$ for some $\gamma_p^s, \gamma_n^s \in \mathbb{R}$.

Let B be a $k \times (k-1)$ matrix such that all its column vectors are orthogonal to $\mathbf{x}_+$. Consider a function $g : \mathbb{R}^{k-1} \to \mathbb{R}^{k-1}$ defined by $g(\mathbf{s}) = B'[\mathbb{E}(\mathbf{x}|\mathbf{s}'\mathbf{x} \geq 0) - \mathbb{E}(\mathbf{x}|\mathbf{s}'\mathbf{x} < 0)]$.[39] Since g is a continuous function from an $(k-1)$-sphere into Euclidean $(k-1)$-space, from the Borsuk-Ulam theorem, there exists $\mathbf{s} \in \mathcal{S}^{k-1}$ such that $g(\mathbf{s}) = g(-\mathbf{s})$. However we know that $g(-\mathbf{s}) = -g(\mathbf{s})$.[40] Therefore there exists $\mathbf{s} \in \mathcal{S}^{k-1}$ such that $g(\mathbf{s}) = -g(\mathbf{s})$, which must be equal to zero. Using the fact that $\Pr(\mathbf{s}'\mathbf{x} \geq 0)\mathbb{E}[\mathbf{x}|\mathbf{s}'\mathbf{x} \geq 0] + \Pr(\mathbf{s}'\mathbf{x} < 0)\mathbb{E}[\mathbf{x}|\mathbf{s}'\mathbf{x} < 0] = \mathbb{E}\mathbf{x} = \mathbf{0}$, we have $B'\mathbb{E}[\mathbf{x}|\mathbf{s}'\mathbf{x} \geq 0] = B'\mathbb{E}[\mathbf{x}|\mathbf{s}'\mathbf{x} < 0] = \mathbf{0}$. This implies that they are proportional to $\mathbf{x}_+$; so we can write $\hat{\mathbf{x}} = \gamma\mathbf{x}_+$ where $\gamma \in \{\gamma_p^s, \gamma_n^s\}$ is a corresponding random variable. Then $\mathbb{E}\hat{v}(\hat{\mathbf{x}}) = \mathbb{E}[\gamma^2\mathbf{x}_+'V\mathbf{x}_+] > 0 = \hat{v}(\mathbb{E}\mathbf{x})$. ■

Proof of Theorem 2. First, we give a necessary condition for the solution, and then

[37] Note also that the second equality holds since $\mathbb{E}_m[\mathbb{E}[\hat{\mathbf{x}}'V\mathbf{x}|m]] = \mathbb{E}_m[\mathbb{E}[\mathbf{x}'V\hat{\mathbf{x}}|m]] = \mathbb{E}_m[\hat{\mathbf{x}}'V\mathbb{E}[\mathbf{x}|m]] = \mathbb{E}_m[\hat{\mathbf{x}}'V\hat{\mathbf{x}}]$.

[38] Formally, for each $\mathbf{s} \in \mathcal{S}^{k-1} \subset \mathbb{R}^k$, define $(\alpha^s, \{m_+, m_-\})$ by $\alpha^s(m_+|\mathbf{s}'\mathbf{x} \geq 0) = 1$, $\alpha^s(m_-|\mathbf{s}'\mathbf{x} \geq 0) = 0$, $\alpha^s(m_+|\mathbf{s}'\mathbf{x} < 0) = 0$, and $\alpha^s(m_-|\mathbf{s}'\mathbf{x} < 0) = 1$.

[39] Note that for all $\mathbf{s} \in \mathcal{S}^{k-1}$, $\mathbb{E}[\mathbf{x}|m] \neq 0$.

[40] This follows from $B'[\mathbb{E}(\mathbf{x}| - \mathbf{s}'\mathbf{x} \geq 0) - \mathbb{E}(\mathbf{x}| - \mathbf{s}'\mathbf{x} < 0)] = B'[\mathbb{E}(\mathbf{x}|\mathbf{s}'\mathbf{x} \leq 0) - \mathbb{E}(\mathbf{x}|\mathbf{s}'\mathbf{x} > 0)]$. Note that $\{\mathbf{x} : \mathbf{s}'\mathbf{x} = 0\}$ has measure zero.

establish the result.

Step 1: We will show that Z is an orthogonal projection matrix whenever Z is a solution to SDP. Since $Z \in \mathbb{S}_+^k$ is a symmetric matrix, we have the eigenvalue decomposition $Z = C\Lambda C'$ where C is an orthogonal matrix and Λ is a diagonal matrix with real entries.[41] $Z \succeq O$ implies that all eigenvalues are nonnegative ($\lambda_i \geq 0$ for every i), and $I - Z = C(I - \Lambda)C' \succeq O$ implies that all eigenvalues must satisfy $1 - \lambda_i \geq 0$ for all i. From a property of the trace operator, we have $\mathrm{tr}(WZ) = \mathrm{tr}(WC\Lambda C') = \mathrm{tr}(\Lambda C'WC) = \sum_{i=1}^k \lambda_i \delta_i$ where δ_i is the i-th diagonal entry of $C'WC$. Now consider a matrix $\tilde{Z} = C\tilde{\Lambda}C'$ where $\tilde{\Lambda}$ is a diagonal matrix with each entry $\tilde{\lambda}_i$ being equal to 0 if $\delta_i < 0$ and 1 if $\delta_i \geq 0$. By construction, we have $\mathrm{tr}(WZ) = \sum \lambda_i \delta_i \leq \sum \tilde{\lambda}_i \delta_i = \mathrm{tr}(W\tilde{Z})$. Since $\tilde{Z}$ is a symmetric positive semidefinite matrix and furthermore is idempotent,[42] the solution to SDP must be an orthogonal projection matrix.[43]

Step 2: We now show that for any orthogonal projection matrix Z of rank l, there exists a $k \times l$ matrix D such that $Z = QD(D'D)^{-1}D'Q'$ where $Q = [\mathbf{q}_1, \ldots, \mathbf{q}_k]$ consists of all eigenvectors of W. Since W is symmetric, we have the eigenvalue decomposition $W = Q\Omega Q' = \sum \omega_i \mathbf{q}_i \mathbf{q}_i'$.

Note that every orthogonal projection matrix is characterized by its target subspace in $\mathbb{R}^k$. Fix an arbitrary subspace in $\mathbb{R}^k$ and suppose that it is spanned by column vectors of some $k \times l$ matrix B. Then the orthogonal projection matrix onto this subspace is written as $P_B = B(B'B)^{-1}B'$.[44] Let $D = Q'B$. Then $B = QD$, and hence

$$\begin{aligned}
P_B =& QD[(QD)'(QD)]^{-1}(QD)' \\
=& QD[D'D]^{-1}D'Q' \\
=& QP_DQ'.
\end{aligned}$$

Step 3: Assume, without loss of generality, that each eigenvalue ω_i is nonnegative for $i = 1, \ldots, r$, and negative for $i = r, \ldots, k$. Let $Q_+ \equiv [\mathbf{q}_1, \ldots, \mathbf{q}_r]$ and $Q_- \equiv [\mathbf{q}_{r+1}, \ldots, \mathbf{q}_k]$. Note that $Q'Q = I_k$ implies $Q_+'Q_+ = I_r$ and $Q_+'Q_- = O_{r,k-r}$. We will show that $\mathrm{tr}(WP_{Q_+}) \geq \mathrm{tr}(WP_B)$ for any $k \times k$ orthogonal projection matrix $P_B = QP_DQ'$. Recall that every diagonal entry of P_D satisfies $0 \leq (P_D)_{ii} \leq 1$ since both P_D and $I - P_D$ are positive semidefinite.

[41] A matrix A is orthogonal if $AA' = I$.

[42] A matrix A is idempotent if $A^2 = A$. Note that $\tilde{Z}^2 = C\tilde{\Lambda}C'C\tilde{\Lambda}C' = C\tilde{\Lambda}^2C' = \tilde{Z}$ since $\tilde{\Lambda}^2 = \tilde{\Lambda}$.

[43] A matrix A is an orthogonal projection matrix if it is symmetric and idempotent.

[44] For any $k \times l$ matrix B of rank l, P_B is symmetric and idempotent. Check $P_B' = [B(B'B)^{-1}B']' = B(B'B)^{-1}B'$ and $P_B^2 = B(B'B)^{-1}B'B(B'B)^{-1}B' = B(B'B)^{-1}B'$.

Then

$$
\begin{aligned}
\operatorname{tr}(WP_B) =& \operatorname{tr}(Q\Omega Q'QP_DQ') \\
=& \operatorname{tr}(\Omega P_D) \\
=& \sum_{i=1}^{k} \omega_i (P_D)_{ii} \\
\leq& \omega_1 + \cdots + \omega_r.
\end{aligned}
$$

Finally we check $\operatorname{tr}(WP_{Q_+}) = \sum_{i=1}^{r} \omega_i$.

$$
\begin{aligned}
\operatorname{tr}(WP_{Q_+}) =& \operatorname{tr}(\Omega Q'Q_+(Q_+'Q_+)^{-1}Q_+'Q) \\
=& \operatorname{tr}\left(\Omega \begin{pmatrix} Q_+' \\ Q_-' \end{pmatrix} Q_+(Q_+'Q_+)^{-1}Q_+' \begin{pmatrix} Q_+ & Q_- \end{pmatrix}\right) \\
=& \operatorname{tr}\left(\Omega \begin{pmatrix} I_r \\ O_{k-r,r} \end{pmatrix} \begin{pmatrix} I_r & O_{r,k-r} \end{pmatrix}\right) \\
=& \operatorname{tr}\left(\begin{pmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_k \end{pmatrix} \begin{pmatrix} I_r & O_{r,k-r} \\ O_{k-r,r} & O_{k,k} \end{pmatrix}\right) \\
=& \omega_1 + \cdots + \omega_r.
\end{aligned}
$$

Thus we conclude that $\operatorname{tr}(WP_{Q_+}) \geq \operatorname{tr}(WZ)$ for every orthogonal projection matrix Z. ■

References

Admati, Anat R. and Paul Pfleiderer (2000), "Forcing firms to talk: Financial disclosure regulation and externalities." *Review of Financial Stludies*, 13, 479–519.

Alonso, Ricardo, Wouter Dessein, and Niko Matouschek (2008), "When does coordination require centralization?" *American Economic Review*, 98, 145–179.

Anderson, Simon P. and Régis Renault (2006), "Advertising content." *American Economic Review*, 96, 93–113.

Angeletos, George-Marios and Alessandro Pavan (2007), "Efficient use of information and social value of information." *Econometrica*, 75, 1103–1142.

Aoyagi, Masaki (2010), "Information feedback in a dynamic tournament." *Games and Economic Behavior*, 70, 242–260.

Baeriswyl, Romain and Camille Cornand (2010), "The signaling role of policy actions." *Journal of Monetary Economics*, 57, 682–695.

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

Blinder, Alan S., Michael Ehrmann, Marcel Fratzscher, Jakob De Haan, and David-Jan Jansen (2008), "Central bank communication and monetary policy: A survey of theory and evidence." *Journal of Economic Literature*, 46, 910–945.

Board, Simon (2009), "Revealing information in auctions: the allocation effect." *Economic Theory*, 38, 125–135.

Boot, Arnoud W. A. and Anjan V. Thakor (2001), "The many faces of information disclosure." *Review of Financial Studies*, 14, 1021–1057.

Boyd, Stephen and Lieven Vandenberghe (2004), *Convex Optimization*. Cambridge University Press, Cambridge.

Calvó-Armengol, Antoni and Joan de Martí (2009), "Information gathering in organizations: Equilibrium, welfare, and optimal network structure." *Journal of the European Economic Association*, 7, 116–161.

Calvó-Armengol, Antoni, Joan de Martí, and Andrea Prat (2009), "Endogenous communication in complex organizations." Working Paper.

Chakraborty, Archishman and Rick Harbaugh (2010), "Persuasion by cheap talk." *American Economic Review*, 100, 2361–2382.

Coate, Stephen (2004), "Political competition with campaign contributions and informative advertising." *Journal of the European Economic Association*, 2, 772–804.

Cornand, Camille and Frank Heinemann (2008), "Optimal degree of public information dissemination." *Economic Journal*, 118, 718–742.

Dessein, Wouter and Tano Santos (2006), "Adaptive organizations." *Journal of Political Economy*, 114, 956–995.

Ederer, Florian (2010), "Feedback and motivation in dynamic tournaments." *Journal of Economics and Management Strategy*, 19, 733–769.

Faust, Jon and Lars E. O. Svensson (2001), "Transparency and credibility: Monetary policy with unobserved goals." *International Economic Review*, 42, 369–397.

Gal-Or, Esther (1986), "Information transmission – Cournot and Bertrand equilibria." *Review of Economic Studies*, 53, 85–92.

Galeotti, Andrea and Andrea Mattozzi (2011), ""Personal influence": Social context and political competition." *American Economic Journal: Microeconomics*, 3, 307–327.

Ganuza, Juan-Jose and Jose S. Penalva (2010), "Signal orderings based on dispersion and the supply of private information in auctions." *Econometrica*, 78, 1007–1030.

Gavazza, Alessandro and Alessandro Lizzeri (2009), "Transparency and economic policy." *Review of Economic Studies*, 76, 1023–1048.

Gentzkow, Matthew and Emir Kamenica (2011), "Competition in persuasion." Working paper.

Geraats, Petra M. (2002), "Central bank transparency." *Economic Journal*, 112, F532–F565.

Goltsman, Maria and Arijit Mukherjee (2011), "Interim performance feedback in multistage tournaments: The optimality of partial disclosure." *Journal of Labor Economics*, 29, 229–265.

Horn, Goger A. and Charles R. Johnson (1985), *Matrix Analysis*. Cambridge University Press, Cambridge.

Jensen, Henrik (2002), "Optimal degree of transparency in monetary policymaking." *Scandinavian Journal of Economics*, 104, 399–422.

Johnson, Justin P. and David P. Myatt (2006), "On the simple economics of advertising, marketing, and product design." *American Economic Review*, 96, 756–784.

Kamenica, Emir and Matthew Gentzkow (2011), "Bayesian persuasion." *American Economic Review*, 101, 2590–2615.

Lewis, Tracy R. and David E. M. Sappington (1994), "Supplying information to facilitate price discrimination." *International Economic Review*, 35, 309–327.

Milgrom, Paul R. and John Roberts (1986), "Price and advertising signals of product quality." *Journal of Political Economy*, 94, 796–821.

Milgrom, Paul R. and Robert J. Weber (1982), "A theory of auctions and competitive bidding." *Econometrica*, 50, 1089–1122.

Morris, Stephen and Hyun Song Shin (2002), "Social value of public information." *American Economic Review*, 92, 1521–1534.

Morris, Stephen and Hyun Song Shin (2007), "Optimal communication." *Journal of the European Economic Association*, 5, 594–602.

Polborn, Mattias K. and David T. Yi (2006), "Informative positive and negative campaigning." *Quarterly Journal of Political Science*, 1, 351–371.

Prat, Andrea (2002), "Campaign advertising and voter welfare." *Review of Economic Studies*, 69, 997–1017.

Prat, Andrea (2005), "The wrong kind of transparency." *American Economic Review*, 95, 862–877.

Rayo, Luis and Ilya Segal (2010), "Optimal information disclosure." *Journal of Political Economy*, 118, 949–987.

Shapiro, Carl (1986), "Exchange of cost information in oligopoly." *Review of Economic Studies*, 53, 433–446.

Vandenberghe, Lieven and Stephen Boyd (1996), "Semidefinite programming." *SIAM Review*, 38, 49–95.

Vives, Xavier (2008), *Information and Learning in Markets: The Impact of Market Microstructure*. Princeton University Press.

Woodford, Michael (2005), "Central-bank communication and policy effectiveness." In *The Greenspan Era: Lessons for the Future*, 399–474, Federal Reserve Bank of Kansas City, Kansas City.