Journal of Economic Theory and Econometrics, Vol. 33, No. 3, Sep. 2022, 11–32

A Note on Reputation in Noisy Cheap Talk *

Yong-Ju Lee†

Abstract We revisit Morris (2001) two-period advice game but with communication error, and investigate how communication error affects the advisor's reputational incentives and thus her information transmission. Reputational incentives differ from Morris (2001) and critically depend on the structure of noise and specific equilibrium strategies. We borrow the notion of "plausible deniability" by Blume *et al.* (2019) and explain the effects of communication error on information transmission and welfare. We show that the weakened reputational incentives reduce the good type advisor's incentive to lie, compared to Morris (2001).

Keywords Advice Game, Cheap Talk, Reputation, Communication Error, Plausible Deniability

JEL Classification C72, D82, D83

*We are grateful to Yong-Gwan Kim, In-Uck Park, and Eun Jeong Heo for their valuable comments. We also deeply thank the two anonymous reviewers for their constructive comments and suggestions. We thank Jae-seok Sean Lee, a graduate student of University of Missouri, for SymPy computation. This study was supported by the 2021 Yeungnam University Research Grant.

†School of Economics and Finance, Yeungnam University, 280 Daehak-ro, Gyeongsan 38541, South Korea. Email: yongjulee@ynu.ac.kr

Received June 20, 2022, Revised September 14, 2022, Accepted September 22, 2022

1. INTRODUCTION

There are several types of noise at each step of communication. Mainly, two types of noise have been considered in the literature on strategic information transmission: a *noisy signal* in the information acquisition stage and a *communication error* in the information transmission stage.

Consider the two types of noise in a cheap talk model between a sender and a receiver. Given that state and message spaces are binary and that the probabilities of each noise are the same, the two noises lead to an identical result if the sender is non-strategic and hence, truthfully delivers whatever information she has observed. However, the two types of noise seem quite different from an epistemological aspect: delivering uncertain information without error and delivering certain information with error. Furthermore, in many social relationships, the sender and the receiver are strategic in the game. Then, different incentives may arise for the sender's purpose between the two situations; accordingly, the receiver must prepare different countermeasures for his purpose. This interaction of responses may result in different outcomes. In addition, if the other elements such as reputational concern are incorporated, the combined effects might be significant.

In his two-period advice game between a decision maker and an advisor, Morris (2001) investigates the reputational incentive of the advisor who receives a *noisy signal* about the state of the world. Reputational concern emerges as an important factor of advice in the repeated relationship since the decision maker's decision depends on the extent to which he trusts the advisor. Thus, even if the advisor does not attach any intrinsic value to her reputation, she may have instrumental reputational concerns simply because she does not wish to be considered to be a biased bad type and wishes her advice to be heeded by the decision maker in his future decision. Morris (2001) shows that even the *good* advisor (whose interests are perfectly aligned with the decision maker's) may lie for her reputation building (i.e., the political correctness effect in his words).

In this note, we slightly modify advice game of Morris (2001) with a different form of noise, a *communication error*. The advisor perfectly observes the state of the world but, instead, the advisor's message is subject to communication error. That is, with some probability, the decision maker receives the message sent by the advisor, but with a complementary probability, the decision maker receives a wrong message. Then, we investigate how communication error affects the advisor's reputational incentives and thus her information transmission.

We show that the advisor still cares about her reputation instrumentally in the repeated interaction but reputational incentives differ from Morris (2001) and

critically depend on the structure of noise and specific equilibrium strategies, and when compared to Morris (2001), the weakened reputational incentives reduce the good type advisor's incentive to lie. We show that the results can be well explained by the notion of "plausible deniability" by Blume *et al.* (2019)

The remainder of this paper is organized as follows. The rest of the section briefly reviews related literature. Section 2 explains the environment of Morris (2001) and its main results, while Section 3 characterizes the main properties of all informative equilibria of the game with communication error and presents an intuitive example in which the good advisor always tells the truth. Section 4 concludes.

2. RELATED LITERATURE

The main insight obtained by Crawford and Sobel (1982) is that informative communication between a privately informed expert and an uninformed decision maker is possible even when the two parties have non-aligned interests. Strategic communication in the repeated environment drastically differs from a one-shot game and truthful information transmission might be possible. Within the framework of Crawford and Sobel (1982), Golosov *et al.* (2014) show that equilibria exist that achieve full revelation of the state of the world in finite time. Other variations of dynamic communication game investigate conditions under which truthful information transmission is possible (e.g., Renault *et al.*, 2013; Foerster, 2019).

If players interact repeatedly, reputational concerns may arise and play a role in strategic communication. This paper belongs to the literature on repeated cheap talk that analyzes reputational concerns arising endogenously, even though there is no intrinsic reputational value (e.g., Sobel, 1985; Benabou and Laroque, 1992; Morris, 2001). The literature shows that reputational concern may lead to information loss in the repeated cheap talk.

Sobel (1985) first introduces the repeated cheap-talk model with reputation, and Benabou and Laroque (1992) extend Sobel (1985) to the case of noisy private signals.[1] Given the *assumption* that a good advisor tells the truth, both studies demonstrate that a bad advisor, whose interests are opposed to those of the decision maker, sometimes has an incentive to tell the truth for reputation building and sometimes to lie for exploiting that reputation.

[1] In Sobel (1985), at the beginning of each stage both the advisor and the decision maker observe the value of a parameter that measures the importance of that period's play of the game. This value has an important role in reputation building and in exploiting that reputation.

Morris (2001) considers a similar environment to that of Benabou and Laroque (1992), but defines the bad advisor as one who is biased toward an extreme decision. Morris (2001) endogenizes the behavior of the good advisor in Benabou and Laroque (1992), and further shows that, in some informative equilibria, even the good advisor who has identical preferences to the decision maker sometimes lies for reputation building (i.e., "political correctness" in his context), and that such reputational incentives may lead to a loss of information.

Unlike the reputational concern of the advisor regarding her *type* (or *preferences*) as in Sobel (1985), Benabou and Laroque (1992), and Morris (2001), some consider reputational concerns regarding *making a correct prediction* or *having accurate information* (e.g., Ottaviani and Sørensen, 2006; Guembel and Rosseto, 2009). Ottaviani and Sørensen (2006) show that when the expert's reputation is updated on the basis of the report as well as the realized state, the expert typically does not wish to truthfully reveal the signal observed. Guembel and Rosseto (2009) consider a situation in which a receiver may misunderstand a sender from two sources, the transmission noise inherent in the communication but privately observed by the sender or the publicly observable quality of the information channel.

The implication of communication error illustrated in the literature on static cheap talk (e.g., Blume *et al.*, 2007; Myerson, 1991) is that the presence of noise in the communication channel may encourage the sender to send more informative message because of the sender's concern regarding misinformation due to the error.[2] Blume *et al.* (2019) analyze a simple garbling procedure, randomized response, as a game and implement it as an experiment and find that garbling in communication increases truth-telling and does so in instances where being truthful adversely affects posterior beliefs. In our model of repeated cheap talk with reputational concern, however, communication error can be a good excuse to help the advisor lie for her current utility maximization.

3. MORRIS (2001)'S ADVICE GAME WITH NOISY SIGNAL

A decision maker (he, R) interacts with an advisor (she or S) in a twice-repeated cheap talk game. The decision maker's optimal decision depends on the state of the world, and he seeks advice from the privately but imperfectly informed advisor. The decision maker is uncertain about the objectives (or types) of the advisor. At the beginning of the game, the advisor is "good" with probability λ_1 and has a utility function that is identical to that of the decision maker.

[2]Blume *et al.* (2007) and Myerson (1991) do not feature reputational concerns.

Similarly, the advisor is "bad" with probability $1-\lambda_1$ and is biased, and always wants as high an action as possible, independent of the state. The probability $(\lambda_t, t = 1,2)$ that the advisor is "good" is the advisor's reputation and is updated according to the initial reputation and the outcomes of the first stage game.

The timing of the game is as follows. At the beginning of the first period, the state of the world $\theta_1 \in \{0,1\}$ is realized with equal probability and the advisor observes a *noisy signal* $s_1 \in \{0,1\}$ regarding the value of θ_1. With probability γ, where $\frac{1}{2} < \gamma < 1$, this signal is equal to the true state; with probability $1-\gamma$, the advisor is misinformed about the state. After observing the signal, the advisor sends a message $m_1 \in \{0,1\}$ to the decision maker. The decision maker interprets the message received in light of the advisor's uncertain type. Given the message, the decision maker takes action $a_1 \in \mathbb{R}$. Then, the state is realized and publicly observed. The decision maker updates his belief about the advisor's type as a function of the initial belief λ_1, the message received m_1, and the realized state of the world θ_1. Therefore, the advisor's reputation at the beginning of the second period can be written as $\lambda_2 = \Lambda(\lambda_1, m_1, \theta_1)$. Then the stage game is played once again, with the same advisor but a new state θ_2, a new message m_2, and a new action a_2.

The decision maker's payoff in each period depends on the state of the world θ and his choice of action a. In particular, his period utility is given by $u_R(a,\theta) = -(a-\theta)^2$. We assume that the decision maker places different weights on period 1 and period 2 utilities. Thus, the total utility of the decision maker is given by

$$-x_1 (a_1 - \theta_1)^2 - x_2 (a_2 - \theta_2)^2,$$

where $x_1 > 0$ and $x_2 > 0$. The good advisor is assumed to have utility identical to that of the decision maker. The bad advisor's period utility is given by $u_B(a) = a$ and thus total utility is assumed to be $y_1 a_1 + y_2 a_2$, where $y_1 > 0$ and $y_2 > 0$. Thus, the bad advisor always wants a higher action, independent of the state.

Since the decision maker's reaction to messages in the second stage depends on his belief about the advisor's type, the first stage game serves to build the reputation of the advisor.

The equilibrium communication strategies in any informative equilibrium in Morris (2001) feature the following. The bad advisor prefers high action and therefore sends a high message $m_1 = 1$ more often than the good advisor. Consequently, observing message $m_1 = 0$ enhances the advisor's reputation. This effect leads the good advisor to lie for her reputation building, sending $m_1 = 0$ with positive probability even if $s_1 = 1$ is observed. To sum up, there always exists a *strict* reputational incentive for the advisor to announce 0, *regardless of*

the realized state: specifically,

$$\Lambda(\lambda_1,0,1) \geq \Lambda(\lambda_1,0,0) > \lambda_1 > \Lambda(\lambda_1,1,1) \geq \Lambda(\lambda_1,1,0).$$

4. REPUTATION IN THE ADVICE GAME WITH COMMUNICATION ERROR

Now we assume that the advisor perfectly observes the state of the world ($\gamma = 1$) and then sends a message to the decision maker. However, the advisor's message m^S is subject to communication error. With probability $1-\varepsilon$, where $0 < \varepsilon < \frac{1}{2}$, the decision maker receives the message sent by the advisor ($m^R = m^S$). With probability ε, the decision maker is misinformed about the message ($m^R \neq m^S$). We assume that the decision maker cannot distinguish between messages resulting from noise and those sent intentionally by the advisor. After the decision maker's action is taken and the state is publicly revealed, the decision maker updates his belief about the advisor's type as a function of the initial belief λ_1, the message received m_1^R, and the realized state of the world θ_1. The advisor's reputation at the beginning of the second period is written as $\lambda_2 = \Lambda(\lambda_1, m_1^R, \theta_1)$.

Babbling equilibria exist in any cheap-talk model and there also exist several forms of informative equilibria. Proposition 1 characterizes the crucial properties of *all* informative equilibria, independent of (x_1, x_2) and (y_1, y_2).

Proposition 1 Any informative equilibrium satisfies the following properties:

(i) The good advisor always announces 0 when she observes state 0 and announces 1 with positive probability when she observes state 1.

(ii) The bad advisor announces 1 more often than the good advisor.

(iii) Reputational incentives for the advisor to announce 0 critically depend on the observed state and each type's strategy. Specifically,

$$\begin{cases} \Lambda(\lambda_1,0,0) > \lambda_1 > \Lambda(\lambda_1,1,0) \text{ and } \Lambda(\lambda_1,0,1) \geq \lambda_1 \geq \Lambda(\lambda_1,1,1) \text{ if } \sigma_B(0) > 0 \\ \Lambda(\lambda_1,0,0) = \lambda_1 = \Lambda(\lambda_1,1,0) \text{ and } \Lambda(\lambda_1,0,1) > \lambda_1 > \Lambda(\lambda_1,1,1) \text{ if } \sigma_B(0) = 0 \end{cases}$$

where $\sigma_B(0)$ denotes the bad advisor's probability of announcing message $m^S = 1$ when the observed state is 0.

Proof. (See the Appendix.)

We make a few notes on the results of Proposition 1. First, the advisor still cares about her reputation instrumentally in the repeated interaction. Since the second period is the last period, the advisor has no incentive to protect her reputation and reputation concern emerges only in the first period. Thus, statements (i) and (ii) apply to the first period only. Second, if the bad advisor's equilibrium strategy is $\sigma_B(0) = 0$ after observing the state 0, reputational incentive disappears. Given the strategy $\sigma_B(0) = 0$, together with statement (i), they know for sure that $m^R = 1$ is an error, because both types tell the truth in equilibrium.

Our model with communication errors and Morris (2001) model with noisy signals share the same game structure and therefore the equilibrium structures are very similar. Due to the change of reputational incentives for the advisor under different noise structures, however, the specific equilibrium strategies differ significantly. Morris (2001) decomposes the effect of noisy signal into three constituent elements, the discipline effect, the sorting effect, and the political correctness effect. In the next subsection with an intuitive example, we will introduce the notion of "plausible deniability" by Blume *et al.* (2019), and explain how the plausible deniability from communication error affects both the discipline effect and the political correctness effect in the current model.

Since the good advisor announces her signal truthfully in any informative equilibria when she observes state 0 and the weakened reputational incentives reduce the good advisor's incentive to lie when she observes state 1, the good type's information transmission improves, compared to Morris (2001). However, the effects on the biased bad advisor are not obvious in general. That is, in some extreme cases, the bad advisor truthfully reveals her signal (for instance, if her reputational concerns are big enough). In broad range, however, the bad advisor lies more often than in Morris (2001).

4.1. AN ILLUSTRATIVE EXAMPLE: AN EQUILIBRIUM IN WHICH A GOOD ADVISOR ALWAYS TELLS THE TRUTH

Now we characterize one intuitive and implicative equilibrium to illustrate the crucial properties described in Proposition 1. Again, since the results in Proposition 1 are satisfied in all informative equilibria, they are satisfied in this specific informative equilibrium.

The game is solved by backward induction.

4.1.1 EQUILIBRIUM IN THE SECOND-PERIOD GAME

In all games, there exists a unique informative equilibrium in the second period of the game.[3] If the decision maker learns something from the message he receives and chooses a higher action after a higher message, the bad advisor will choose 1 independent of the state and the good advisor will have a strict incentive to be truthful. The advisor's strategy is summarized by the following table.

Table 1: Message $m_2^S \in \{0,1\}$ in the second period

	$\theta_2 = 0$	$\theta_2 = 1$
Good advisor	0	1
Bad advisor	1	1

After receiving m_2^R, the decision maker draws inferences about the state of the world considering each type's strategy and the noise structure. Given the unique strategy of the second period in Table 1, message $m_2^S = 0$ is only sent by the good advisor. Due to communication error, however, m_2^R can be 0 or 1, anytime . Thus, unlike Morris (2001) model, the decision maker cannot be sure whether the advisor is good or bad.

If the decision maker receives message $m_2^R = 0$, given the strategy in Table 1, there are four possible cases: (i) the good type sent $m_2^S = 0$ after observing $\theta_2 = 0$ and the decision maker receives $m_2^R = 0$ as sent by the good type (i.e., $\frac{1}{2}\lambda_2(1-\varepsilon)$), (ii) the good type sent $m_2^S = 1$ after observing $\theta_2 = 1$ but the decision maker receives $m_2^R = 0$ due to communication error (i.e., $\frac{1}{2}\lambda_2\varepsilon$), (iii) the bad type sent $m_2^S = 1$ after observing $\theta_2 = 0$ and the decision maker receives $m_2^R = 0$ due to communication error (i.e., $\frac{1}{2}(1-\lambda_2)\varepsilon$), and (iv) the bad type sent $m_2^S = 1$ after observing $\theta_2 = 1$ and the decision maker receives $m_2^R = 0$ due to communication error (i.e., $\frac{1}{2}(1-\lambda_2)\varepsilon$). Hence, If the decision maker receives message $m_2^R = 0$, the probability that the true state equals 1 is

[3] If $\varepsilon = \frac{1}{2}$, then an equilibrium exists in which both types of senders tell the truth and the decision maker chooses ε if he receives $m_2^R = 0$ and $1-\varepsilon$ if he receives $m_2^R = 1$, in both periods. In this case, no reputational concern arises. In this study, we assumed that $\varepsilon < \frac{1}{2}$.

$$\begin{aligned} p\left(\theta_2 = 1 \middle| m_2^R = 0\right) &= \frac{p(\theta_2 = 1 \cap m_2^R = 0)}{p(m_2^R = 0)} \\ &= \frac{p(\theta_2 = 1 \cap m_2^R = 0)}{p\left(\theta_2 = 0 \cap m_2^R = 0\right) + p(\theta_2 = 1 \cap m_2^R = 0)} \\ &= \frac{\frac{1}{2}\lambda_2\varepsilon + \frac{1}{2}(1-\lambda_2)\varepsilon}{\frac{1}{2}\lambda_2(1-\varepsilon) + \frac{1}{2}(1-\lambda_2)\varepsilon + \frac{1}{2}\lambda_2\varepsilon + \frac{1}{2}(1-\lambda_2)\varepsilon} \\ &= \frac{\varepsilon}{\lambda_2(1-2\varepsilon) + 2\varepsilon} \end{aligned}$$

and he chooses action

$$\frac{\varepsilon}{\lambda_2(1-2\varepsilon) + 2\varepsilon}.$$

If the decision maker receives message $m_2^R = 1$, the probability that the true state equals 1 is

$$\begin{aligned} p\left(\theta_2 = 1 \middle| m_2^R = 1\right) &= \frac{\frac{1}{2}\lambda_2(1-\varepsilon) + \frac{1}{2}(1-\lambda_2)(1-\varepsilon)}{\frac{1}{2}\lambda_2\varepsilon + \frac{1}{2}(1-\lambda_2)(1-\varepsilon) + \frac{1}{2}\lambda_2(1-\varepsilon) + \frac{1}{2}(1-\lambda_2)(1-\varepsilon)} \\ &= \frac{1-\varepsilon}{-\lambda_2(1-2\varepsilon) + 2(1-\varepsilon)} \end{aligned}$$

and he chooses action

$$\frac{1-\varepsilon}{-\lambda_2(1-2\varepsilon) + 2(1-\varepsilon)}.$$

Given the decision maker's strategy, we check if it is really the best response for the bad advisor to stick to sending 1 independent of the state. If the bad advisor sends $m_2^S = 1$ regardless of the state, her payoff in the second period is given by

$$y_2\left[\frac{\varepsilon^2}{\lambda_2(1-2\varepsilon) + 2\varepsilon} + \frac{(1-\varepsilon)^2}{-\lambda_2(1-2\varepsilon) + 2(1-\varepsilon)}\right].$$

If the bad advisor sends $m_2^S = 0$ after observing $\theta_2 = 0$ or $\theta_2 = 1$, then her payoff will be

$$y_2\varepsilon(1-\varepsilon)\left[\frac{1}{\lambda_2(1-2\varepsilon)+2\varepsilon}+\frac{1}{-\lambda_2(1-2\varepsilon)+2(1-\varepsilon)}\right].$$

The former is greater than the latter as long as $\lambda_2 > 0$ and $\varepsilon \neq \frac{1}{2}$. Therefore, it is the best response for the bad advisor to announce 1 independent of the state; hence, both players' strategies described above constitute an equilibrium in the second period of the game.

Now, the payoff function for reputation for both types of advisors entering the second period can be derived:

$$\begin{aligned} v_G(\lambda_2) = -x_2 \Bigg[& \frac{1}{2}(1-\varepsilon)\left(\frac{\varepsilon}{\lambda_2(1-2\varepsilon)+2\varepsilon}\right)^2 \\ & + \frac{1}{2}\varepsilon\left(\frac{1-\varepsilon}{-\lambda_2(1-2\varepsilon)+2(1-\varepsilon)}\right)^2 \\ & + \frac{1}{2}(1-\varepsilon)\left(\frac{-\lambda_2(1-2\varepsilon)+(1-\varepsilon)}{-\lambda_2(1-2\varepsilon)+2(1-\varepsilon)}\right)^2 \\ & + \frac{1}{2}\varepsilon\left(\frac{\lambda_2(1-2\varepsilon)+\varepsilon}{\lambda_2(1-2\varepsilon)+2\varepsilon}\right)^2 \Bigg] \end{aligned} \tag{1}$$

and

$$v_B(\lambda_2) = y_2\left[\frac{(1-\varepsilon)^2}{-\lambda_2(1-2\varepsilon)+2(1-\varepsilon)}+\frac{\varepsilon^2}{\lambda_2(1-2\varepsilon)+2\varepsilon}\right] \tag{2}$$

Both functions (1) and (2) are continuous and strictly increasing in λ_2, and hence, she has a reputational concern in the first period. In the following analysis, we assume that the informative equilibrium that gives these reputational payoffs is played in the second period.

4.1.2 EQUILIBRIUM IN THE FIRST-PERIOD GAME

The first period of the game is the same as the second period of the game except that now the advisor has reputational concerns arising from her desire to have her advice listened to in the second period.

Let $\Lambda(\lambda_1, m_1^R, \theta_1)$ denote the equilibrium posterior probability assigned to the advisor being good. Then, the good advisor's total payoff in the first period is given by

$$-x_1 (a_1 - \theta_1)^2 + v_G \left(\Lambda \left(\lambda_1, m_1^R, \theta_1\right)\right),$$

and the bad advisor's payoff is given by

$$y_1 a_1 + v_B \left(\Lambda \left(\lambda_1, m_1^R, \theta_1\right)\right).$$

Unlike the unique equilibrium strategy in the second period of the game, there exist several forms of informative equilibrium. That is, the good advisor, on observing state 1, may randomize between telling the truth for the current utility and lying to enhance her reputation.[4] In this subsection, to provide accurate intuition regarding all possible informative equilibria, we analyze one special equilibrium in which the good advisor tells the truth after observing any signal. However, all informative equilibria share the crucial properties of equilibrium analyzed here.

Suppose that a good advisor always tells the truth. In this case, it is not the best response for the bad advisor also to always tell the truth. If the bad advisor always told the truth, the decision maker would not be able to update his beliefs about the advisor's type. However, the bad advisor would like to convince the decision maker that the true state is 1 and, given the decision maker's strategy, there is no reputational cost in announcing 1. Then, the bad advisor has an incentive to always announce 1, which contradicts the assumption that she always tells the truth. In addition, it is clear that in any informative equilibrium the bad advisor must send message 1 strictly more than the good advisor. If she sent message 1 less, she would have a current and a reputational incentive to send message 1. Therefore, the bad advisor always sends 1 if she observes state 1 and sends it with some probability $\tau > 0$ if she observes state 0. This strategy is summarized in Table 2.

Table 2: Message $m_1^S \in \{0,1\}$ in the first period

	$\theta_1 = 0$	$\theta_1 = 1$
Good advisor	0	1
Bad advisor	0 with probability $1-\tau$ 1 with probability τ	1

After receiving m_1^R, the decision maker draws inferences about the advisor's type if the true state is realized at the end of the first period. Suppose that the

[4]Even in these cases, the good advisor still tells the truth when she observes state 0 for both the current utility and reputation.

decision maker receives $m_1^R = 1$ and state 1 is realized. Given the advisor's strategy and noise structure, the decision maker's posterior belief about the type of the advisor is

$$\begin{aligned}\Lambda(\lambda_1, 1,\ 1) &= \frac{p(Good \cap m_1^R = 1 \cap \theta_1 = 1)}{p\left(Good \cap m_1^R = 1 \cap \theta_1 = 1\right) + p(Bad \cap m_1^R = 1 \cap \theta_1 = 1)} \\ &= \frac{\lambda_1(1-\varepsilon)}{\lambda_1(1-\varepsilon) + (1-\lambda_1)(1-\varepsilon)} \\ &= \lambda_1\end{aligned}$$

By similar computations,

$$\begin{aligned}\Lambda(\lambda_1, 1,\ 0) &= \frac{\lambda_1\varepsilon}{\lambda_1\varepsilon + (1-\lambda_1)\{\varepsilon + \tau(1-2\varepsilon)\}} < \lambda_1, \\ \Lambda(\lambda_1, 0,\ 1) &= \frac{\lambda_1\varepsilon}{\lambda_1\varepsilon + (1-\lambda_1)\varepsilon} = \lambda_1, \\ \Lambda(\lambda_1, 0,\ 0) &= \frac{\lambda_1(1-\varepsilon)}{\lambda_1(1-\varepsilon) + (1-\lambda_1)\{(1-\varepsilon) - \tau(1-2\varepsilon)\}} > \lambda_1,\end{aligned}$$

Since $\tau > 0$, this implies that

$$\Lambda(\lambda_1, 0,\ 0) > \lambda_1 = \Lambda(\lambda_1, 1,\ 1) = \Lambda(\lambda_1, 0,\ 1) > \Lambda(\lambda_1, 1,\ 0) \tag{3}$$

There is *no* reputational concern when *state 1* is realized. Advisor's reputation varies only when state 0 is realized, and her reputation increases when the decision maker receives $m_1^R = 0$.

This result is different from Morris (2001), in which each advisor has a strict reputational incentive to announce 0, *independent of the state of* the world. Specifically, under Morris (2001) environment and each type's strategy, we would have

$$\Lambda(\lambda_1, 0,\ 1) = \Lambda(\lambda_1, 0,\ 0) > \lambda_1 > \Lambda(\lambda_1, 1,\ 1) > \Lambda(\lambda_1, 1,\ 0). \tag{M}$$

The relationship (M) from Morris (2001) has $\Lambda(\lambda_1, 0,\ 1) > \lambda_1 > \Lambda(\lambda_1, 1,\ 1)$, while (3) has $\Lambda(\lambda_1, 0,\ 1) = \Lambda(\lambda_1, 1,\ 1) = \lambda_1$. Intuitively, the key difference lies

in the belief update following an erroneous low message (i.e., a combination of $m_1^R = 0$ and $\theta_1 = 1$).[5] Recall that $m_1^S = 1$ is each type's strategy when $s_1(\text{or } \theta_1) = 1$. In Morris (2001) with a *noisy signal*, the decision maker understands that this may have occurred because the advisor observed the wrong signal $s_1 = 0$. But the bad type is biased against low actions, she will enhance her tendency toward $m_1^S = m_1^R = 1$. When the decision maker receives $m_1^R = 0$, he increases the belief that the advisor is a good type (i.e., reputation increases). Hence, a tradeoff emerges for the advisor between current and future utility because each advisor has a reputational incentive to announce 0, independent of the state of the world. In case of a *communication error,* the combination of $m_1^R = 0$ and $\theta_1 = 1$ is interpreted differently. Upon observing $\theta_1 = 1$, the decision maker knows for sure that the advisor must also have observed $\theta_1 = 1$. Since the two types of advisor agree on what is the optimal action when $\theta_1 = 1$ (i.e., $m_1^S = 1$), neither type has an incentive to lie about it. Hence, $m_1^R = 0$ must be simply due to a communication error and the advisor's reputation remains unchanged. Therefore, there is no reputational concern when the advisor observes state 1. *This kind of reputational concern plays a role in reducing the good type advisor's incentive to lie when she observes state 1*. Anyway, the derivation of different reputational concerns from the two cases with different noise structure suggests that the result in Morris (2001) may critically depend on the advisor receiving a noisy signal.

Now, we return to the decision maker's inferences about the state of the world. If the decision maker receives a message $m_1^R = 0$, the probability that the true state equals 1 is

$$p\left(\theta_1 = 1 \middle| m_1^R = 0\right) = \frac{\varepsilon}{\lambda_1(1-\varepsilon) + (1-\lambda_1)\{(1-\varepsilon) - \tau(1-2\varepsilon)\} + \varepsilon}$$

and he chooses action

$$\frac{\varepsilon}{\lambda_1(1-\varepsilon) + (1-\lambda_1)\{(1-\varepsilon) - \tau(1-2\varepsilon)\} + \varepsilon}.$$

If the decision maker receives a message $m_1^R = 1$, the probability that the true state equals 1 is

$$p\left(\theta_1 = 1 \middle| m_1^R = 1\right) = \frac{1-\varepsilon}{\lambda_1\varepsilon + (1-\lambda_1)\{\varepsilon + \tau(1-2\varepsilon)\} + 1 - \varepsilon}$$

and he chooses action

[5] This way of logic of reputation update holds true in any informative equilibrium.

$$\frac{1-\varepsilon}{\lambda_1\varepsilon+(1-\lambda_1)\{\varepsilon+\tau(1-2\varepsilon)\}+1-\varepsilon}.$$

Next, we check the bad advisor's strategy and derive τ as a function of λ_1. Suppose that the bad advisor observed state 0. Her total utility from telling the truth ($m_1^S=0$), which is the sum of her current utility and her expected value of reputation, would be

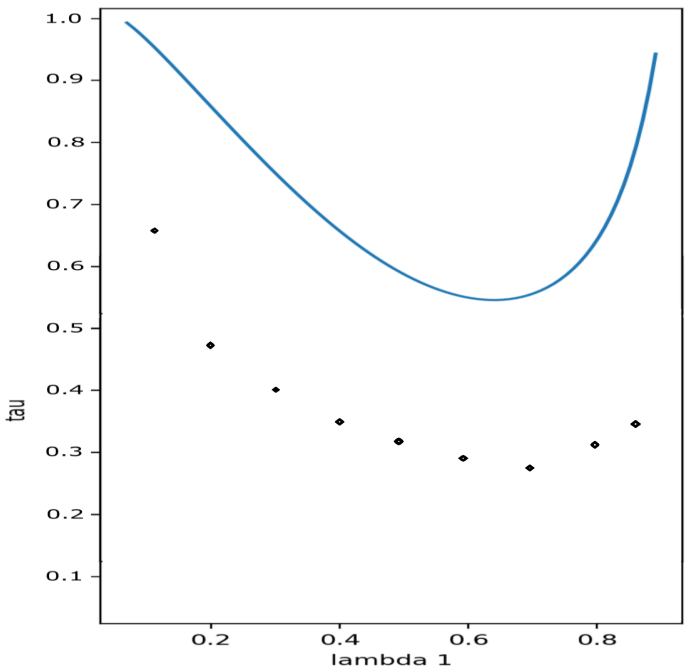
$$\begin{aligned} & y_1(1-\varepsilon)\frac{\varepsilon}{\lambda_1(1-\varepsilon)+(1-\lambda_1)\{(1-\varepsilon)-\tau(1-2\varepsilon)\}+\varepsilon} \\ & +y_1\varepsilon\frac{1-\varepsilon}{\lambda_1\varepsilon+(1-\lambda_1)\{\varepsilon+\tau(1-2\varepsilon)\}+1-\varepsilon} \\ & +(1-\varepsilon)v_B(\Lambda(\lambda_1,0,0))+\varepsilon v_B(\Lambda(\lambda_1,1,0)), \end{aligned} \tag{4}$$

and her total utility from lying ($m_1^S=1$) is

$$\begin{aligned} & y_1(1-\varepsilon)\frac{1-\varepsilon}{\lambda_1\varepsilon+(1-\lambda_1)\{\varepsilon+\tau(1-2\varepsilon)\}+1-\varepsilon} \\ & +y_1\varepsilon\frac{\varepsilon}{\lambda_1(1-\varepsilon)+(1-\lambda_1)\{(1-\varepsilon)-\tau(1-2\varepsilon)\}+\varepsilon} \\ & +(1-\varepsilon)v_B(\Lambda(\lambda_1,1,0))+\varepsilon v_B(\Lambda(\lambda_1,0,0)) \end{aligned} \tag{5}$$

In equilibrium, either the bad advisor always lies ($\tau=1$) and (5) exceeds (4), or $0<\tau<1$ and (4) equals (5). There is always a unique such τ since expression (4) is strictly increasing in τ and expression (5) is strictly decreasing in τ. For example, if $\varepsilon=\frac{1}{4}$, $y_1=\frac{1}{10}$, and $y_2=1$, so that the bad advisor cares much more about the second-period outcome than the first-period outcome, then the unique value of τ is plotted as a function of λ_1 in Figure 1. The solid line in Figure 1 represents the value of τ and the dots represent the corresponding values (denoted by τ_{Morris}) of Morris (2001) with $1-\gamma=\frac{1}{4}$. Observe that when her initial reputation is very low or very high, she knows that the improvement of her reputation by her report is limited, so she lies most of the time.

Figure 1: τ as a function of λ_1 when $\varepsilon = \frac{1}{4}$, $y_1 = \frac{1}{10}$, and $y_2 = 1$



Note: The solid line represents the value of τ and the dots represent the corresponding values (τ_{Morris}) of Morris (2001) with $1-\gamma = \frac{1}{4}$.

If $\varepsilon = \frac{1}{4}$, $y_1 = 1$, and $y_2 = 1$, so that the decision problems are equally important to the bad advisor, then reputational concerns are too small to persuade the bad advisor to tell the truth, and the bad advisor always announces 1. That is, as the second period becomes more important, the bad advisor is more likely to tell the truth in the first period in this numerical example.

So far, it has been assumed that a good advisor tells the truth. Finally, we need to check if truth-telling is really optimal for the good advisor. If the good advisor observes state 0, she has an unambiguous incentive to tell the truth for current and future utility, since this will lead the decision maker to choose a lower action in this period and it will also enhance her reputation. If she observes state 1, she will gain in terms of the current outcome if she tells the truth. Furthermore,

there is no loss in her reputation since $\Lambda(\lambda_1,1,1) = \Lambda(\lambda_1,0,1) = \lambda_1$. Hence, truth-telling is optimal for the good advisor. This means that each type's strategy described in Table 2 and the associated action of the decision maker constitute an equilibrium of the game.

These results, albeit with different concepts of noise and the bad type, reconfirm findings from Sobel (1985) and Benabou and Laroque (1992) that, given an *assumption* that the good advisor tells the truth, the bad advisor sometimes has an incentive to tell the truth for reputation building and sometimes to lie for exploiting that reputation. Instead, we endogenize the behavior of the advisor as an equilibrium behavior, rather than an assumption.

The results of the current paper are also different from the implication of communication error illustrated in the literature on static cheap talk. If $x_1 = y_1 = 0$ (or $x_2 = y_2 = 0$), then our model resembles the existing static cheap talk model with communication error like Blume *et al.* (2007) and Myerson (1991). However, there are some differences in the structures of noise and the existence of "type" of a player, and, in particular, there is no reputational concern in those static cheap talk models. Therefore, direct comparison is limited. Nevertheless, in our model of repeated cheap talk with reputational concern, communication error helps the advisor lie for her current utility maximization. This is a stark difference.

4.1.3 DISCUSSION ON THE OVERALL WELFARE

Given that the good advisor tells the truth, as shown in Table 2, the bad advisor randomizes between 0 with probability $1-\tau$ and 1 with probability τ when she observes the state 0. And Figure 1 shows that τ is greater than τ_{Morris} for all λ_1. That is, the bad advisor announces 1 more often in this model than she does in Morris (2001). This phenomenon is well explained by the notion of "plausible deniability" by Blume *et al.* (2019). The bad advisor would feel more free to announce 1 in the current paper, because even if the state 0 is revealed at the end of the period, the bad advisor can simply argue that she sent 0 but the message is delivered wrong as 1 due to the communication error. By this deniability, the bad advisor increases her first period utility without reputational loss. If this is true, the good advisor would lie much less. Hence, the overall effect on information transmission is mixed.

As described in Proposition 1, in any informative equilibrium, the good advisor always reveals her signal truthfully when she observes state 0 and reputational incentives reduce the good advisor's incentive to lie when she observes state 1, and therefore the good type's information transmission improves, com-

pared to Morris (2001). However, unlike the result of this example, the reputational effects on the bad advisor are not obvious, in general.

The similar argument can be applied to welfare. Morris (2001) explains three welfare effects: (i) To enhance her reputation, the bad advisor may sometimes announce 0 when observing the state 0 (i.e., the discipline effect). (ii) The decision maker learns and updates about the advisor's type from first-period play (i.e., the sorting effect), and (iii) the decision maker's concern about the type of the advisor may provide incentives to the good advisor to lie in the first period (i.e., the political correctness effect). "Plausible deniability" from communication error affects both the discipline effect and the political correctness effect in the current paper. That is, the discipline effect is strengthened and the political correctness effect is weakened. The overall welfare effect is ambiguous as in Morris (2001).

5. CONCLUSION

We revisited twice-repeated cheap talk game of Morris (2001) but with communication error and investigated how communication error affects the advisor's reputational incentives and her information transmission.

The advisor still cares about her reputation instrumentally in the repeated interaction but the relationship of reputational incentives varies depending on the structure of noise and equilibrium strategies.

The "plausible deniability" by Blume *et al.* (2019) can provide an intuitive explanation on the effects of communication error. The bad advisor would feel more free to lie in the current paper, because even if the other state is revealed at the end of the period, the bad advisor can deny her lying simply arguing that she sent the signal truthfully but the message was delivered wrong due to the communication error. By this deniability, the bad advisor increases her first period utility without reputational loss. If this is true, the good advisor would lie much less. Hence, the overall effect on information transmission is mixed. The similar argument can be applied to welfare.

For future research, analyzing a model admitting both forms of noise seems promising. Then, we can see both the individual and the combined effects on the advisor's reputational incentives, considering model of Morris (2001), the current model and a model having both forms of noise as subcases. In the proposed model, we may further ask whether results in Morris (2001) with noisy signals are robust to the addition of different form of noise. An experimental study comparing the effects of different types of noise on information transmission in a

repeated relationship is worth conducting.

REFERENCES

Benabou, R. and G. Laroque (1992). "Using Privileged Information To Manipulate Markets: Insiders, Gurus, and Credibility," *Quarterly Journal of Economics* 107(3), 921-958.

Blume, A., Board, O. and K. Kawamura (2007). "Noisy talk," *Theoretical Economics* 2, 395-440.

Blume, A., Lai, E. K. and W. Lim (2019). "Eliciting private information with noise: The case of randomized response," *Games and Economic Behavior* 113, 356-380.

Crawford, V. and J. Sobel (1982). "Strategic Information Transmission," *Econometrica* 50, 1431-1451.

Foerster, M. (2019). "Dynamics of strategic information transmission in social networks," *Theoretical Economics* 14, 253-295.

Blume, A., Lai, E. K. and W. Lim (2014). "Dynamic strategic information transmission," *Journal of Economic Theory* 151, 304-341.

Guembel, A. and S. Rosseto (2009). "Reputational Cheap Talk with Misunderstanding," *Games and Economic Behavior* 67, 736-744.

Morris, S. (2001). "Political Correctness," *Journal of Political Economy* 109(2), 231-265.

Myerson, R. (1991). *Game Theory: Analysis of Conflict*, Harvard University Press, Cambridge, Massachusetts.

Ottaviani, M. and P. Sørensen (2006). "Reputational cheap talk," *Rand Journal of Economics* 37(1) 155-175.

Renault, J., Solan, E. and N. Vieille (2013). "Dynamic sender–receiver games," *Journal of Economic Theory* 148(2), 502-534.

Sobel, J. (1985). "A Theory of Credibility," *Review of Economic Studies* 52(4), 557-573.

APPENDIX

Proof of Proposition 1. As in Morris (2001), we provide a formal proof for the properties of the informative equilibria by analyzing a static version of the advice game, in which advisors have exogenous reputational payoff functions (1) and (2).

Let $\sigma_I(\theta)$, $I = G$ or B, be the type I advisor's probability of announcing message $m^S = 1$ when the state is θ and $\chi(m^R)$ be the decision maker's action if m^R is the received message from the advisor.

The posterior probability that the advisor is good if he receives message m^R and state θ is realized is

$$\Lambda\left(m^R,\theta\right)=\frac{\lambda\phi_G(m^R|\theta)}{\lambda\phi_G\left(m^R|\theta\right)+(1-\lambda)\phi_B(m^R|\theta)}\,, \tag{A1}$$

where $\phi_I(m^R|\theta)$ is the probability that the decision maker receives message m^R by the type I advisor given state θ; that is, $\phi_I\left(1|\theta\right)=(1-\varepsilon)\,\sigma_I\left(\theta\right)+\varepsilon\left(1-\sigma_I\left(\theta\right)\right)=\varepsilon+(1-2\varepsilon)\,\sigma_I\left(\theta\right)$, $\phi_I\left(0|\theta\right)=(1-\varepsilon)\left(1-\sigma_I\left(\theta\right)\right)+\varepsilon\sigma_I\left(\theta\right)=(1-\varepsilon)-(1-2\varepsilon)\,\sigma_I\left(\theta\right)$, and $\phi_I\left(1|\theta\right)+\phi_I\left(0|\theta\right)=1$.

The decision maker's posterior belief that the true state is 1 if message m^R arrives is

$$\Gamma\left(m^R\right)=\frac{\lambda\phi_G\left(m^R\middle|1\right)+(1-\lambda)\phi_B(m^R|1)}{\lambda\phi_G\left(m^R|1\right)+(1-\lambda)\,\phi_B\left(m^R|1\right)+\lambda\phi_G\left(m^R|0\right)+(1-\lambda)\phi_B(m^R|0)} \tag{A2}$$

Write $\widehat{u}_G\left(q,\theta\right)$ and $\widehat{u}_B\left(q\right)$ for the expected values of $u_G(a,\theta)=-(a-\theta)^2$ and $u_B(a)=a$ if the advisor has observed state θ and the decision maker believes that the true state is 1 with probability q.

Let $\Pi_I^C(\theta)$ denote the net current expected gain to the type I advisor choosing message $m^S=1$, rather than $m^S=0$, when she observes state θ, if the decision maker follows his optimal strategy, that is,

$$\begin{aligned}\Pi_G^C\left(\theta\right)=x\,[&(1-\varepsilon)\,\widehat{u}_G\left(\Gamma\left(1\right),\theta\right)\\&+\varepsilon\widehat{u}_G\left(\Gamma\left(0\right),\theta\right)-(1-\varepsilon)\,\widehat{u}_G\left(\Gamma\left(0\right),\theta\right)-\varepsilon\widehat{u}_G\left(\Gamma\left(1\right),\theta\right)]\\=x\,(1-2\varepsilon)\,&[\widehat{u}_G\left(\Gamma\left(1\right),\theta\right)-\widehat{u}_G\left(\Gamma\left(0\right),\theta\right)]\end{aligned}$$

$$\Pi_B^C(\theta) = \Pi_B^C(0) = \Pi_B^C(1) = y(1-2\varepsilon)\left[\widehat{u}_B(\Gamma(1)) - \widehat{u}_G(\Gamma(0))\right] \tag{A3}$$

where $x > 0,\ y > 0$ are weights on the current utility.

Let $\Pi_I^R(\theta)$ denote the net expected reputational gain to the type I advisor of choosing $m^S = 0$ rather than $m^S = 1$ when she observes state θ, that is,

$$\begin{aligned}\Pi_I^R(1) = &\left[(1-\varepsilon)v_I(\Lambda(0,1)) + \varepsilon v_I(\Lambda(1,1))\right.\\ &\left. -(1-\varepsilon)v_I(\Lambda(1,1)) - \varepsilon v_I(\Lambda(0,1))\right]\\ = &(1-2\varepsilon)\left[v_I(\Lambda(0,1)) - v_I(\Lambda(1,1))\right]\end{aligned}$$

$$\Pi_I^R(0) = (1-2\varepsilon)\left[v_I(\Lambda(0,0)) - v_I(\Lambda(1,0))\right] \tag{A4}$$

Note that $v_I\left(\Lambda\left(m^R,\theta\right)\right)$ should be exactly the same functions as defined in (1) and (2) in Section 3.1 since there exists a unique informative equilibrium in the second period of the game.

Now, we prove the properties in several steps by applying the similar logic of Morris (2001). It is assumed without loss of generality that $\Gamma(1) \geq \Gamma(0)$ and thus $\chi(1) \geq \chi(0)$.

Step 1 $\Lambda(0,0) \geq \Lambda(1,0)$ and $\Lambda(0,1) \geq \Lambda(1,1)$.

This asserts that there must be a weak reputational incentive to announce 0. The proof shows by contradiction that no equilibrium exists if one of these conditions is violated.

(1) Suppose that $\Lambda(1,0) > \Lambda(0,0)$ and $\Lambda(1,1) > \Lambda(0,1)$. Then $\Pi_B^R(\theta) < 0$ and $\Pi_B^C(\theta) \geq 0$ for each $\theta = 0,1$ and thus we must have $\sigma_B(0) = \sigma_B(1) = 1$. But if $\sigma_G(0) = \sigma_G(1) = 1$, $\Lambda(0,0) = \Lambda(1,0) = \Lambda(0,1) = \Lambda(1,1)$, a contradiction. But if $\sigma_G(0) \neq 1$ or $\sigma_G(1) \neq 1$, then $\Lambda(0,0) > \Lambda(1,0)$ or $\Lambda(0,1) > \Lambda(1,1)$, another contradiction.

(2) Suppose that $\Lambda(0,0) \geq \Lambda(1,0)$ and $\Lambda(0,1) < \Lambda(1,1)$. By the definition of $\Lambda\left(m^R,\theta\right)$ in (A1), we have $\phi_G(1|0) \leq \phi_B(1|0)$ and $\phi_G(1|1) > \phi_B(1|1)$. Then $\sigma_G(0) \leq \sigma_B(0)$ and $\sigma_G(1) > \sigma_B(1)$. Observe also that $\Pi_I^R(0) > \Pi_I^R(1)$ and $\Pi_I^C(0) \leq \Pi_I^C(1)$. Thus for both I, $\sigma_I(0) = 0$ or $\sigma_I(1) = 1$, a contradiction.

(3) Suppose that $\Lambda(0,0) < \Lambda(1,0)$ and $\Lambda(0,1) \geq \Lambda(1,1)$. By the definition of $\Lambda\left(m^R,\theta\right)$, we have $\phi_G(1|0) > \phi_B(1|0)$ and $\phi_G(1|1) \leq \phi_B(1|1)$. Then we have $\sigma_G(0) > \sigma_B(0)$ and $\sigma_G(1) \leq \sigma_B(1)$. And observe that $\Pi_B^R(1) > \Pi_B^R(0)$ and $\Pi_B^C(1) = \Pi_B^C(0)$, so either $\sigma_B(1) = 0$ or $\sigma_B(0) = 1$. Thus $\phi_B(1|1) \leq \phi_B(1|0)$.

Since $\phi_G(1|0) > \phi_B(1|0)$ and $\phi_G(1|1) \leq \phi_B(1|1)$, this implies $\phi_G(1|1) < \phi_G(1|0)$. But now $\Gamma(1) < \Gamma(0)$, a contradiction.

Step 2 $\Lambda(0,0) \geq \Lambda(1,0)$ and $\Lambda(0,1) \geq \Lambda(1,1)$, and at least one of these inequalities is strict.

This asserts that there must be a strict reputational incentive to announce 0. Suppose that both held with equality. Recall that $\chi(1) \geq \chi(0)$ by assumption. If $\chi(1) > \chi(0)$, the bad advisor would have strict incentive to choose 1 (whatever her state), leading to a contradiction. But if $\chi(1) = \chi(0)$, we have a babbling equilibrium.

Step 3 $\sigma_B(0) \geq \sigma_G(0)$ and $\sigma_B(1) \geq \sigma_G(1)$, and at least one of these inequalities is strict.

This property directly holds by Step 2 and by the definition of $\Lambda\left(m^R, \theta\right)$.

Step 4 $\chi(1) > \chi(0)$.

If $\chi(1) = \chi(0)$, then (by Step 2) the bad advisor has a strict incentive to choose 0 (whatever her state), leading to a contradiction.

Step 5 $\sigma_G(0) = 0$.

By Step 2, $\Pi_G^R(0) \geq 0$, and by Step 4, $\Pi_G^C(0) < 0$. So $\sigma_G(0) = 0$.

Step 6 $\sigma_G(1) > 0$.

Suppose $\sigma_G(1) = 0$. To have $\Gamma(1) > \Gamma(0)$, we must have $\sigma_B(1) > \sigma_B(0)$. These imply $\Lambda(0,1) > \Lambda(0,0) > \lambda$ and $\Lambda(1,1) > \Lambda(1,0) = 0$. Then $\Pi_B^R(1) > \Pi_B^R(0)$, and thus $\sigma_B(1) \leq \sigma_B(0)$, a contradiction.

Step 7

$$\begin{cases} \Lambda(0,0) > \lambda > \Lambda(1,0) \text{ and } \Lambda(0,1) \geq \lambda \geq \Lambda(1,1) \text{ if } \sigma_B(0) > 0 \\ \Lambda(0,0) = \lambda = \Lambda(1,0) \text{ and } \Lambda(0,1) > \lambda > \Lambda(1,1) \text{ if } \sigma_B(0) = 0 \end{cases}.$$

Properties of Step 7 follow from Step 3, Step 4, and Step 5, and from the definition of $\Lambda\left(m^R, \theta\right)$. $\Lambda(0,0) = \lambda = \Lambda(1,0)$ is satisfied when $\sigma_G(0) = \sigma_B(0)$, and $\Lambda(0,1) = \lambda = \Lambda(1,1)$ is satisfied when $\sigma_G(1) = \sigma_B(1)$.

Finally, part (i) is proved by Step 5 and Step 6, part (ii) is proved by Step 3, and part (iii) is proved by Step 7.