Cheap Talk with Prior-biased Inferences*

Yong-Ju Lee† Wooyoung Lim‡ Chen Zhao§

September 26, 2022

Abstract

We investigates how prior-biased inferences change players' strategic incentives and result in novel welfare implications in the canonical framework of strategic information transmission. The ex ante social welfare achieved in our model exceeds the upper bound characterized in the standard environment without prior bias. The welfare gain stems from the fact that the receiver's prior bias weakens the link between the sender's message and the receiver's response without contaminating the actual content of the messages. We further show that direct communication is optimal among all possible communication protocols in the presence of a sufficient degree of prior bias.

Keywords: Communication, Information Transmission, Cheap Talk, Prior Bias, Non-Bayesian Updating

***JEL* classification numbers**: C72, D82, D83, D91

*The paper has greatly benefited from input by the editor, advisory editor, and two referees. We are grateful to Pak Hung Au, Andreas Blume, Vince Crawford, Chishio Furukawa, Sidartha Gordon, Navin Kartik, Bryant Kim, Yong-Gwan Kim, Run Li, Shintaro Miura, Jawwad Noor, Joel Sobel, and Jianrong Tian for their valuable comments and suggestions. For helpful comments and discussions, we thank the seminar participants at HKUST, Lingnan college at SYSU, and Sungkyunkwan University, and conference and workshop participants at the conference on Strategic Information Transmission at Peking University, the 13th East Asian Contract Theory Conference, and the 2020 (virtual) Korea Economic Review International Conference. A previous version of this paper has been circulated under the title of "Cheap Talk with Non-Bayesian Updating."

†School of Economics and Finance, Yeungnam University. *Email:* `yongjulee@ynu.ac.kr`

‡Department of Economics, The Hong Kong University of Science and Technology. *Email:* `wooyoung@ust.hk`

§Faculty of Business and Economics, University of Hong Kong. *Email:* `czhao@hku.hk`

1 Introduction

Bayes' rule provides a normative way of combining prior information with additional information that arises from a new observation, and is regarded as the canonical model of information processing in economics and other social sciences. In contrast, laboratory evidence suggests that how people perceive probability and process information deviates systematically from what Bayes' rule predicts (e.g., Kahneman and Tversky, 1972, 1973; Grether, 1978, 1992).[1] In light of this evidence, many alternative modeling approaches have been offered in behavioral economics (e.g., Epstein, 2006; Ortoleva, 2012; Zhao, 2016).[2]

In this paper, we investigate how non-Bayesian updating changes players' strategic incentives, behavioral predictions and, most importantly, welfare implications in the canonical environment of strategic information transmission. Precisely, we explore a model in which a fully informed expert sends a cheap-talk message to a decision maker who does not update her belief according to Bayes' rule. We assume that the decision maker is prone to a particular type of non-Bayesian updating bias, *prior bias*.

Prior bias, also known as conservatism or inertia, captures inferences drawn in favor of the current belief, and it belongs under the umbrella of confirmation bias. Evidence of prior bias and its related behaviors is prevalent not only in laboratory settings (e.g., Pitz, Downing, and Reinhold, 1967; Geller and Pitz, 1968; Pitz, 1969), but also in real life situations that involve strategic communication (e.g., Weible, Heikkila, deLeon, and Sabatier, 2012; Cairney, 2016; King, 2016).

To model prior bias, we adopt the following updating rule, as axiomatized in Epstein (2006). For each signal m,

$$q(\cdot|m) = (1-a)p(\cdot|m) + ap_0(\cdot) \tag{1.1}$$

where $p_0(\cdot)$ is the prior and $p(\cdot|m)$ is the Bayesian update of p_0.[3] Hence, $q(\cdot|m)$ captures prior-biased inferences where $a \in (0,1)$ is the *degree of prior bias*. The degree of prior bias reflects the *mistrust* a decision maker has for the information source. In particular, a decision maker that follows (1.1) behaves as if she thinks that with probability a, the source is not credible at all.

We incorporate the prior-biased belief updating presented in equation (1.1) into the model of Crawford and Sobel (1982) and investigate how the presence of prior bias changes players' strategic incentives in a communication environment. Introducing prior bias creates the following key tradeoff: On one hand, information transmission is undermined because the actions taken by the decision maker are distorted by her prior bias (distortionary effect). On the other hand, knowing that the decision maker has prior bias, the expert has an incentive to send more informative signals since the link between his message and the decision maker's response is weakened (strategic effect).

Consistent with the standard findings in the literature, all our equilibria are interval partitional. We show that under a variant of the standard monotonicity condition, there exists a unique, most informative equilibrium that ex ante Pareto dominates all other equilibria of the game. Specifically, when the conflict

[1] See Camerer (1995) and Rabin (1998) for surveys of the relevant papers in psychology.

[2] See Benjamin (2019) for a survey of non-Bayesian updating models.

[3] In Epstein (2006), the degree of prior bias may in principle depend on the message received. However, a constant a allows us to obtain sharper comparative statics and welfare implications.

of interest is small relative to the degree of prior bias, the most informative equilibrium induces infinitely many actions. Such an equilibrium can be constructed with an appropriately chosen degree of prior bias even in situations where no communication can be sustained in standard cheap talk.

We further show that if the degree of prior bias is within a certain range, the most informative equilibrium of our game achieves social welfare that is strictly higher than the upper bound obtained with the standard information-garbling devices of Blume, Board, and Kawamura (2007) and Goltsman, Hörner, Pavlov, and Squintani (2009). Effectively, prior bias garbles the information *in the head* of the decision maker, which relieves the conflict of interest between parties without actually contaminating the content of the messages. In other words, prior bias is less costly than standard information-garbling devices.

To understand our welfare result more systematically, we decompose the welfare gain into the strategic effect and the distortionary effect. Compared with the optimal noise equilibrium in Blume et al. (2007), the most informative equilibrium in our model, with a properly chosen degree of prior bias, offers a significantly higher strategic effect and a similar distortionary effect. We find that to achieve the same level of strategic effect in a noise equilibrium, a much larger distortionary effect will have to be incurred because, in the event that the message transmitted is a noise, the action taken by the decision maker is independent of her optimal action given the equilibrium partition. This type of distortion does not exist in our model since the message from the expert is always delivered verbatim. Thus, although the information will be discounted in the same way, our decision maker always receives the right information to begin with.

We then incorporate prior bias into the optimal mediation problem à la Goltsman et al. (2009). We characterize the maximum social welfare given the conflict of interest and the degree of prior bias. When the most informative equilibrium in the cheap-talk game induces infinitely many actions, i.e., when the conflict of interest between parties is small relative to the degree of prior bias, the maximum social welfare with mediation turns out to be exactly the same as the social welfare achieved by this most informative equilibrium. This result implies that, with a sufficient degree of prior bias, direct communication between the expert and the decision maker is optimal among all possible communication protocols.

As we mentioned earlier, prior bias naturally captures *mistrust*. The decision maker mistrusts the expert, and *incorrectly* thinks the expert's advice is noisy. In particular, the decision maker believes that, with a positive probability, the message she receives from the expert is completely uninformative, in which case, she does not update her beliefs. As a result, welfare evaluated under the true, noise-free model could exceed the standard upper bound given by Goltsman et al. (2009).[4] The endogeneity of the noise distribution and the fact that the expert's message is actually free from noise (but the decision maker wrongly believes it to be noisy) are the main differences between our model and that of Blume et al. (2007).

One of the most prevalent findings in the experimental literature on strategic information transmission is *over-communication* (see, e.g., Dickhaut, McCabe, and Mukherji, 1995; Blume, DeJong, Kim, and Sprinkle, 2001), the phenomenon that more information is transmitted from the sender to the receiver than the most

[4]This method of evaluating welfare under the true model is standard in the behavioral economics literature in the presence of time inconsistency. See, for example, O'Donoghue and Rabin (1999) and O'Donoghue and Rabin (2003). For additional discussion on the welfare criterion, see Section 5.

informative equilibrium of the model.[5] Two main explanations provided by the literature are truth-telling preference (Gneezy, 2005; Gneezy, Kajackaite, and Sobel, 2018; Abeler, Nosenzo, and Raymond, 2019) and individual heterogeneity of strategic sophistication (Cai and Wang, 2006; Wang, Spezio, and Camerer, 2010; Lafky, Lai, and Lim, 2022). Our model provides a novel way of rationalizing the over-communication phenomenon: when people bring some home-grown mistrust to the lab, our theory suggests that more information transmission than what the standard theory predicts may result.

The remainder of the paper is organized as follows. The rest of this section reviews the related literature. Section 2 presents our main model, in which players have a quadratic loss utility, and characterizes the set of equilibria. General welfare implications under a broad class of type distributions are also examined. In Section 3, we focus on the uniform prior to present equilibrium properties and welfare implications. Section 4 discusses the optimal mediation problem in the uniform-quadratic environment. We discuss the welfare criterion in Section 5. Section 6 concludes.

1.1 Literature Review

In their seminal paper, Crawford and Sobel (1982) (hereafter CS) consider a model of strategic information transmission in which a fully informed sender sends a cheap-talk message to a receiver, who then takes action that affects the payoffs of both. The main insight obtained by CS is that more informative communication is possible when players' preferences are more aligned. Blume et al. (2007) extend the CS model to the situation in which the communication channel is noisy. When the sender sends a message, there is a chance that any message in the message space is randomly transmitted to the receiver. Blume et al. (2007) show that with an appropriately chosen degree of noise, there exists an equilibrium that is more informative than the most informative equilibrium of CS.[6] Goltsman et al. (2009) consider an optimal mediation problem in the uniform-quadratic environment of CS and characterize the upper bound of the social welfare. We consider the problem in the presence of prior bias and find that direct communication is optimal among all possible communication channels, and the maximum social welfare strictly exceeds the upper bound characterized by Goltsman et al. (2009). We identify that the main source of the welfare gain relative to Goltsman et al. (2009) is the fact that prior bias induces information garbling only in the head of the receiver and thus is less costly to society than directly introducing noise.

Many papers in the literature consider a type-dependent conflict of interest between the sender and the receiver. Melumad and Shibano (1991) consider the standard communication game with quadratic utilities and notice the existence of an equilibrium with infinitely many intervals. Gordon (2010) studies a more general class of preferences and characterizes the conditions for the existence of such type of equilibria. However, no welfare analysis of this type of equilibria has been conducted by either paper. In this strand of research, our paper is most related to Kawamura (2015), who considers a uniform-quadratic setting in which the sender and the receiver have asymmetric beliefs about the quality of the sender's observed signal. The uniform-quadratic specification of our model corresponds to the case in which the sender is

[5] See Blume, Lai, and Lim (2020) for the review of the most recent development of the literature.

[6] Earlier papers that illustrate the potential for mediation and a noisy communication channel to improve information transmission include Forges (1985) and Myerson (2013).

overconfident about the quality of his signal relative to the receiver. Kawamura (2015) shows that in this case, finite-interval equilibria with more steps Pareto-dominate those with fewer steps, and a small level of such overconfidence may improve information transmission beyond the level in CS. We extend the welfare analysis to general type distributions, and show that prior bias can improve social welfare even beyond the bound obtained by Blume et al. (2007), Goltsman et al. (2009), and Ivanov (2010).

In summary, our contribution to the literature is threefold: First, we provide a welfare analysis of the infinite-interval equilibrium under general distributions. In particular, we show that under a variant of the standard monotonicity condition, the infinite-interval equilibrium generates a higher welfare than any finite-interval equilibrium. Second, we present a novel setting in which the conflict of interest between the sender and the receiver is effectively type-dependent, but the ex ante preferences are aligned. This feature allows us to compare the ex ante welfare obtained in our model with standard welfare bounds in Crawford and Sobel (1982), Blume et al. (2007), and Goltsman et al. (2009). We find that the main ingredient of our model, prior bias, may lead to an ex ante welfare that exceeds these welfare bounds, and that the infinite-interval equilibrium plays an important role in the welfare gain. Third, we extend the optimal mediation results in Goltsman et al. (2009) to the prior bias setting. We fully characterize the optimal communication protocol, and show that mediation is unnecessary if the infinite-interval equilibrium is played in the direct talk game.

Various behavioral models have been explored in the literature on strategic information transmission. Kartik, Ottaviani, and Squintani (2007) and Kartik (2009) introduce naïveté of senders by assuming that they have preferences for telling the truth. Ottaviani and Squintani (2006), Kartik et al. (2007), and Chen (2011) relax the full rationality of receivers by assuming they are credulous. Jehiel and Koessler (2008) applies the notion of analogy-based expectations equilibrium to the CS model and show that the analogy grouping of the receiver may improve information transmission.

2 Setup

Consider a standard cheap-talk game. There are two players, an expert (henceforth the sender, or S), and a decision maker (henceforth the receiver, or R). The sender is privately informed about the state of nature $\theta \in [0,1]$, which is drawn from a common knowledge probability measure p_0 with mean μ. We assume that p_0 has positive and continuous density everywhere in the support. Knowing θ, the sender sends a message $m \in M = [0,1]$ to the receiver, who then takes an action $y \in Y = \mathbb{R}$. Both players' payoffs depend on the state of nature θ and the action y taken. In particular, the sender's payoff is $U_S(y,\theta) = -(y-\theta-b)^2$, and the receiver's payoff is $U_R(y,\theta) = -(y-\theta)^2$, in which $0 < b < \mu$. The parameter b measures the conflict of interest between the two players and is assumed to be constant.

2.1 Prior-biased Receiver

We assume the receiver is prior-biased, as in Epstein (2006). Let p be the joint distribution of messages and types, given the strategies of the players. On receiving each message m, a prior-biased receiver fails to fully update her beliefs to the Bayesian conditional distribution $p(\cdot|m)$. Instead, she adopts $q(\cdot|m)$ as her posterior, given by

$$q(\cdot|m) = (1-a)p(\cdot|m) + ap_0(\cdot), \tag{2.1}$$

where $a \in (0,1)$ measures the degree of prior bias. Thus, the receiver partially neglects the information content of the signal in favor of her prior. We assume the sender knows a and takes the prior bias of the receiver into account in his signaling strategy.

An alternative way to interpret our model is to consider the case in which the receiver mistrusts the sender. The receiver incorrectly believes that, with probability a, the message she receives from the sender is a pure noise, in which case, she does not update her beliefs.[7] Under this interpretation, a captures the degree of mistrust.

Let $y_R(\theta)$ be the best response of the prior-biased receiver if the sender of type θ reveals himself. With quadratic utility, $y_R(\theta) = (1-a)\theta + a\mu$. For the sender, the same as in the standard setting, let $y_S(\theta) = \arg\max_y U_S(y,\theta) = \theta + b$. Let θ^* solve the equation $y_R(\theta^*) = y_S(\theta^*)$, i.e., $\theta^* = \mu - \frac{b}{a}$. We call θ^* the *agreement type* when it is nonnegative.

Note that the quadratic loss function we assume is not entirely harmless. Apart from generating the interval-partitional equilibrium structure, quadratic utility plays an indispensable role in the analysis of the optimal mediation problem as it does in Goltsman et al. (2009). With quadratic utility, the sufficient statistic for welfare is the covariance between equilibrium actions and the state of nature, which is key in the welfare calculations under the uniform-prior assumption. However, our results for general distributions can be extended to preferences that have strict single-crossing expectational differences (Kartik, Lee, and Rappoport, 2019) and admit at most one agreement type within the state space.

2.2 Equilibrium

Let $\sigma : \Theta \to \Delta(M)$ be the behavioral strategy of the sender that specifies a distribution of messages for each type θ. For the receiver, by the strict concavity of U_R in y, it is without loss of generality to restrict attention to pure strategies $\rho : M \to Y$. The joint distribution of messages and types, p, is then given by $p(m,\theta) = \sigma(m|\theta)p_0(\theta)$ for any m and θ. Let $\mathbb{E}_\nu$ be the expected value operator with respect to probability measure ν.

Our equilibrium notion differs from the standard perfect Bayesian equilibrium in that the receiver updates with prior bias a. In particular, given message $m \in M$, the receiver chooses the action that maximizes her expected payoff given beliefs $q(\cdot|m)$, rather than the Bayesian conditional $p(\cdot|m)$.

Definition. *Given a, $\{\sigma,\rho\}$ is a prior-biased equilibrium (henceforth equilibrium) if*

1. *for all $\theta \in \Theta$, supp$(\sigma(\cdot|\theta)) \subseteq \arg\max_{m' \in M} U_S(\rho(m'),\theta)$, and*

[7] This happens if the message is randomly drawn from the ex ante distribution of equilibrium messages, independent to the state of nature.

2. *for all $m \in M$, $\rho(m) = \arg\max_{y \in Y} \int_\Theta U_R(y,\theta) q(d\theta|m)$ where*

$$q(\theta|m) = (1-a)\frac{\sigma(m|\theta)p_0(\theta)}{\int_\Theta \sigma(m|\theta)p_0(d\theta)} + ap_0(\theta)$$

for all $\theta \in \Theta$ and any $m \in M$ such that $\int_\Theta \sigma(m|\theta)p_0(d\theta) > 0$.

In fact, any prior-biased equilibrium can be interpreted as a perfect Bayesian equilibrium under the assumption that the receiver holds incorrect beliefs about the information transmitted. In particular, consider a receiver who mistakenly believes that, with probability a, the message she receives is just a random draw from the ex ante distribution of equilibrium messages, $\int_\Theta \sigma(\cdot|\theta)p_0(d\theta)$. In this case, the "noise" is independent to equilibrium messages and thus is completely uninformative about the state. Thus, on receiving any equilibrium message m, the Bayesian update of such a receiver will exactly be $q(\cdot|m)$. Hence, the main differences between this paper and Blume et al. (2007) are the endogeneity of the noise distribution and the welfare considerations that result from the receiver's incorrect beliefs.

Our equilibrium notion is also tightly connected to that of cursed equilibrium (Eyster and Rabin, 2005), in which each player partially neglects the information content of other players' strategies. In fact, any prior-biased equilibrium in the strategic communication game can be expressed as a cursed equilibrium in which the receiver is cursed.

Observe that a babbling equilibrium in which no information is transmitted always exists. In order to characterize the set of informative equilibria, it is useful to define the notion of outcome equivalence. Let ν and ν' be the joint distribution of types and actions induced by $\{\sigma, \rho\}$ and $\{\sigma', \rho'\}$, respectively. We say that $\{\sigma, \rho\}$ and $\{\sigma', \rho'\}$ are *outcome equivalent* if $\nu(\cdot|\theta) = \nu'(\cdot|\theta)$ for all but a set of types that is at most countable. An equilibrium is *interval-partitional* if (i) each type θ induces a single action and (ii) for each $y \in Y$, the set of all types that induces y is a (possibly degenerate) interval.

As usual, any equilibrium in our model is outcome equivalent to an interval-partitional equilibria.[8] In an interval-partitional equilibrium, a closed interval $I \subseteq [0,1]$ is a *pooling interval* if all types in the interior of I induce the same action and any type outside I induces a different action. Any boundary point of a pooling interval is called *boundary type*. We say that a sender of type θ *separates* if he induces an action distinct from all other types. Given any closed interval $[\tau, \tau'] \subseteq [0,1]$, let $\gamma(\tau,\tau')$ denote the conditional mean of θ with respect to p_0 given that $\theta \in [\tau, \tau']$. Note that since p_0 has positive and continuous density everywhere in the support, γ is strictly increasing in both arguments and $\gamma(\tau,\tau) = \tau$ for all $\tau \in [0,1]$.

For any $1 \le n \le \infty$, an interval-partitional equilibrium is said to be an *n-step equilibrium* if n non-degenerate pooling intervals are induced. In terms of descriptive properties, our model belongs to the class of communication games in which the conflict of interest between the sender and the receiver depends on the underlying type, which is first considered by Melumad and Shibano (1991), and further explored by Gordon (2010) and Kawamura (2015). In particular, when $\theta^* \ge 0$, the sender exhibits outward bias; the locus of the sender's most preferred action, $[y_S(0), y_S(1)]$, contains the locus of the receiver's best

[8]See Blume et al. (2007, Proposition 1) for a proof. The proposition applies directly because it exploits only the optimality of the sender whose preferences satisfy the standard assumptions imposed by Blume et al. (2007).

response, $[y_R(0), y_R(1)]$. By Gordon (2010, Theorem 4), in the case of outward bias, there is at least one ∞-step equilibrium. Let $N(a,b) \in \mathbb{Z}^+ \cup \{\infty\}$ be the maximum number of non-degenerate pooling intervals supported by a and b.

Consider an arbitrary interval-partitional equilibrium. Let $[\theta_{i-1}, \theta_i]$ and $[\theta_i, \theta_{i+1}]$ be two neighboring pooling intervals that induce actions y_i and y_{i+1}, respectively. The incentive compatibility of the receiver and the sender requires

$$\begin{aligned} y_i &= (1-a)\gamma(\theta_{i-1}, \theta_i) + a\mu, \\ y_{i+1} &= (1-a)\gamma(\theta_i, \theta_{i+1}) + a\mu, \\ \theta_i &= \frac{y_i + y_{i+1}}{2} - b. \end{aligned}$$

Combining the equations above yields the following second-order difference equation:

$$V(\theta_{i-1}, \theta_i, \theta_{i+1} | a, b) \equiv (1-a)\frac{\gamma(\theta_{i-1}, \theta_i) + \gamma(\theta_i, \theta_{i+1})}{2} + a\mu - b - \theta_i = 0. \tag{2.2}$$

For any $N \in \mathbb{Z}^+ \cup \{\infty\}$, an increasing (decreasing) sequence $\{\tau_i\}_{i=0}^N$ is a *forward* (*backward*) *solution* to (2.2) if $\tau_0 < \tau_1$ ($\tau_0 > \tau_1$) and $V(\tau_{i-1}, \tau_i, \tau_{i+1} | a, b) = 0$ for $1 \le i < N$.

The following proposition summarizes the characterization results that are familiar to the literature.

Proposition 1. *$N(a,b) = \infty$ if and only if $\theta^* \equiv \mu - \frac{b}{a} \ge 0$. Furthermore, in any ∞−step equilibrium, θ^* is the unique type that separates, and the set of boundary types is given by $\{\theta_i\}_{i=0}^\infty \cup \{\theta^*\} \cup \{\theta'_i\}_{i=0}^\infty$ in which*

(i) *if $\theta^* > 0$, $\{\theta_i\}_{i=0}^\infty$ is a forward solution to (2.2) with $\theta_0 = 0$; if $\theta^* = 0$, then $\theta_i = 0$ for all i,*

(ii) *$\{\theta'_i\}_{i=0}^\infty$ is a backward solution to (2.2) with $\theta'_0 = 1$,*

(iii) $\lim_{i\to\infty} \theta_i = \lim_{i\to\infty} \theta'_i = \theta^*$.

Proof. See Appendix A. □

Hence, in any ∞-step equilibrium, the agreement type is the unique type that separates, and the pooling intervals become arbitrarily small as they approach the agreement type from either side. The existence of an ∞-step equilibrium with similar features in the presence of an agreement type is also noted by Kawamura (2015) under the uniform type distribution. Next, we turn to the focus of this paper, the welfare results.

2.3 Welfare

Non-Bayesian agents are potentially time-inconsistent. Therefore, we will need to be more careful when selecting the welfare criterion. Given a strategy profile $\{\sigma, \rho\}$, for $j = S, R$, the ex ante expected payoff of j is given by $\mathrm{EU}_j \equiv \mathbb{E}_p[U_j(\rho(m), \theta)]$. For a sender of type θ, his interim expected payoff is $\mathrm{IU}_S(\theta) \equiv \mathbb{E}_p[U_S(\rho(m), \theta)|\theta]$; upon receiving message m, the interim expected payoff of the receiver is $\mathrm{IU}_R(m) \equiv \mathbb{E}_q[U_R(\rho(m), \theta)|m]$, the ex ante expected value of which is $\mathrm{EIU}_R \equiv \mathbb{E}_p[\mathbb{E}_q[U_R(\rho(m), \theta)|m]]$.

The beliefs of the receiver, albeit prior-biased, are a martingale. In other words, the expected interim beliefs of the receiver coincide with the prior, which directly implies that $\mathrm{EIU}_R = \mathrm{EU}_R$. Moreover, under quadratic utility, this martingale property ensures that in equilibrium, the ex ante expected action of the receiver is still the mean of θ as in standard cheap-talk games. Thus, the following proposition holds, which shows that the ex ante welfare is aligned among the sender, the ex ante self of the receiver, and the interim self of the receiver. This result allows us to welfare-rank equilibria according to EU_R without worrying about the welfare consequences of time inconsistency. We further discuss this welfare criterion in Section 5.

Proposition 2. *If $\{\sigma, \rho\}$ is a prior-biased equilibrium, then*

$$EU_R = EIU_R = EU_S + b^2.$$

Proposition 2 marks the main difference between our model and other cheap-talk models with a type-dependent bias (e.g., Melumad and Shibano, 1991): in our model, ex ante welfare is aligned between the sender and the receiver irrespective of the degree of prior bias. This observation enables us to Pareto-rank equilibria according to EU_R across different degrees of prior bias and thus allows us to compare the level of welfare achieved in our model with that in standard strategic communication games.

To obtain comparative statics results, Crawford and Sobel (1982) introduce a monotonicity condition called Condition (M). We now introduce the same condition in our context.

Condition (M). *Given a and b, if $\bar{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\tau}}$ are two forward (or two backward) solutions to (2.2) with $\bar{\tau}_0 = \hat{\tau}_0$ and $\bar{\tau}_1 > \hat{\tau}_1$, then $\bar{\tau}_i > \hat{\tau}_i$ for all $i \geq 2$.*

In our model, Condition (M) is not sufficient for uniqueness of n-step equilibrium for every n. In regard to ∞−step equilibria, even if $\bar{\tau}_i > \hat{\tau}_i$ for all i, the sequences can still have the same limit, and thus, the uniqueness of equilibrium partitions is not guaranteed. Therefore, we state a stronger form of Condition (M). This stronger form, called Condition (M*), also holds, given the sufficient condition of Condition (M) provided by Crawford and Sobel (1982, Theorem 2).

Condition (M*). *Given a and b, if $\bar{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\tau}}$ are two forward (or two backward) solutions to (2.2) with $\bar{\tau}_0 = \hat{\tau}_0$ and $\bar{\tau}_1 - \hat{\tau}_1 > 0$, then there is $\epsilon > 0$ such that $\bar{\tau}_i - \hat{\tau}_i > \epsilon$ for all $i \geq 1$.*

Besides the uniqueness of n-step equilibrium, Condition (M*) also ensures that both the sender and the receiver prefer the equilibrium with the largest number of steps among all equilibria.

Theorem 1. *Given a and b, if Condition (M*) holds, then*

(i) *for any $1 \leq n \leq N(a, b)$, all n-step equilibria are outcome equivalent;*

(ii) *ex ante, any $N(a, b)$-step equilibrium Pareto-dominates equilibria with fewer steps.*

Proof. See Appendix A. □

In Kawamura (2015), it is shown that under the uniform type distribution, finite-step equilibria with more steps Pareto-dominate those with fewer steps. We consider general type distributions and include the ∞-step equilibrium into the welfare anaylsis. In particular, Theorem 1 shows that under Condition (M*), equilibria with the largest number (∞ included) of steps always Pareto-dominate those with fewer steps.

Note that Theorem 1 is silent on how equilibria with fewer than $N(a,b)$ steps are ranked. When the agreement type does not exist ($\theta^* < 0$), standard welfare arguments apply, which ensures that any equilibrium with more steps dominates equilibria with fewer steps. However, when the agreement type exists ($\theta^* > 0$), sender types below θ^* demand a lower action than the receiver, and thus, ex ante welfare may not be increasing in the second largest boundary type. In this case, we show that the ∞-step equilibrium still Pareto-dominates all other equilibria.

Now, we briefly sketch the proof for part (ii) when $\theta^* > 0$. First, for every finite-step equilibrium, we perturb it by artificially inserting θ^* as a boundary type. In this way, we are making every finite-step equilibrium strictly more informative. Focus on $[0, \theta^*]$ (as the argument for $[\theta^*, 1]$ is symmetric). Within $[0, \theta^*]$, by $y_S(\theta) \leq y_R(\theta)$, the sender always demands a weakly lower action than the receiver. Then, standard welfare arguments imply that a decrease in the first cutoff θ_1 results in an increase in the number of partitions within $[0, \theta^*]$ and increases the ex ante welfare within $[0, \theta^*]$. In addition, by Proposition 1, in any ∞-step equilibrium, θ^* must be a boundary type. Hence, as the number of equilibrium partitions goes to infinity, the welfare of the perturbed equilibrium converges to the welfare of the ∞-step equilibrium, which implies that the ∞-step equilibrium Pareto dominates all finite-step equilibria.

The following proposition shows that log-concavity of the type distribution is sufficient for Condition (M*) to hold.[9]

Proposition 3. *If p_0 has continuously differentiable and log-concave density, then Condition (M*) holds for any $a \in (0,1)$ and $b \in (0, \mu)$.*

Proof. See Appendix A. □

To see the main idea of the proof, let $\{\theta_0, \theta_1, \theta_2\}$ be a forward solution to (2.2) and consider increasing both θ_1 and θ_2 by Δ. Due to log-concavity, it can be shown that $\frac{\partial \gamma(\underline{\theta}+\Delta, \overline{\theta}+\Delta)}{\partial \Delta}|_{\Delta=0} = \gamma_1(\underline{\theta}, \overline{\theta}) + \gamma_2(\underline{\theta}, \overline{\theta}) \leq 1$ for any $\underline{\theta} < \overline{\theta}$.[10] Consider increasing θ_0 by Δ and θ_1 by Δ' for some small Δ and Δ' such that $\Delta' > \Delta$. Since the conditional expectation $\gamma(\theta_0, \theta_1)$ will increase by less than Δ', the action induced by $[\theta_0, \theta_1]$ will increase by less than $(1-a)\Delta'$. As θ_1 is increased by Δ', its optimal action is also increased by Δ'. Thus, for θ_1 to be indifferent, the action induced by $[\theta_1, \theta_2]$ will need to be increased by more than Δ', which requires that θ_2 be increased by more than Δ'. Inductively, it is easy to see that if $\bar{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\tau}}$ are two forward solutions to (2.2) with $\bar{\tau}_0 = \hat{\tau}_0$ and $\bar{\tau}_1 - \hat{\tau}_1 > 0$, then $\bar{\tau}_i - \hat{\tau}_i \geq \bar{\tau}_{i-1} - \hat{\tau}_{i-1} \geq \cdots \geq \bar{\tau}_1 - \hat{\tau}_1$.

For the purpose of equilibrium refinement, we adopt the *no-incentive-to-separate* criterion proposed by Chen, Kartik, and Sobel (2008), which requires that the worst type does not want to deviate by separating

[9]By contrast, Szalay (2012) shows that if $\frac{\partial U_R(y,\theta)}{\partial y} + \frac{\partial U_R(y,\theta)}{\partial \theta}$ is non-increasing in y, $\frac{\partial U_S(y,\theta)}{\partial y} + \frac{\partial U_S(y,\theta)}{\partial \theta}$ is non-decreasing in y, and p_0 has a log-concave density, then for any *finite* n, all n-step equilibria, if they exist, must be outcome equivalent.

[10]See Lemma 2 in Appendix A for a detailed proof.

himself. In Appendix B, we show that with log-concavity, generically only the most informative equilibrium survives the refinement. Hereafter, we will focus on the $N(a,b)$-step equilibrium when presenting the remaining welfare results.

Next, we compare the welfare achieved in our model with the welfare achieved in the standard cheap talk. With a harmless abuse of notation, let $\mathrm{EU}_R(a,b)$ denote the ex ante welfare of the receiver in the $N(a,b)$-step equilibrium, and let $\mathrm{EU}_R(0,b)$ denote the ex ante welfare of the receiver in the most informative cheap-talk equilibrium. The following proposition shows that under Condition (M*), if the conflict of interest is sufficiently large, then setting the degree of prior bias to be $\frac{b}{\mu}$ will improve welfare relative to the case without prior bias.

Proposition 4. *If p_0 has continuously differentiable and log-concave density, then there exists $b^* \in (0, \frac{\mu}{2})$ such that $EU_R(\frac{b}{\mu}, b) > EU_R(0,b)$ for any $b \in (b^*, \mu)$.*

Proof. See Appendix A. □

Note that if $a = \frac{b}{\mu}$, the agreement type θ^* is 0. Thus, $\mathrm{EU}_R(\frac{b}{\mu}, b)$ is in fact the welfare of an ∞-step equilibrium in which type 0 separates. We can always construct such an ∞-step equilibrium for any $b \in (0,\mu)$. In contrast, for sufficiently large values of b, information transmission in the standard cheap-talk equilibrium is low. In this case, the strategic advantage generated by a sufficient level of prior bias outweighs the distortion towards the actions. Next, we turn to the uniform distribution for more welfare results.

3 The Uniform Case

In this section, we further assume that p_0 is the uniform distribution on $[0,1]$. The uniform prior enables us to compare our welfare results with several important recent contributions in the literature, including Blume et al. (2007) and Goltsman et al. (2009). Note that the uniform distribution is log-concave, which implies that Condition (M*) is satisfied. The remainder of the paper focuses on this uniform-quadratic case unless stated otherwise.

Figure 1 illustrates how the maximum number of intervals $N(a,b)$ varies as a function of a and b. Consistent with the standard finding in the literature, given the degree of prior bias, the maximum number of equilibrium partitions weakly decreases in the degree of conflict of interest between the parties. Moreover, given the degree of conflict of interest, the maximum number of equilibrium partitions weakly *increases* in the degree of prior bias. This observation confirms our intuition that prior bias encourages the sender to be more informative in his messages. Specifically, when the conflict of interest is small relative to the degree of prior bias, there exists an ∞-step equilibrium. In Appendix C, we present the full characterization of all equilibria in the model.

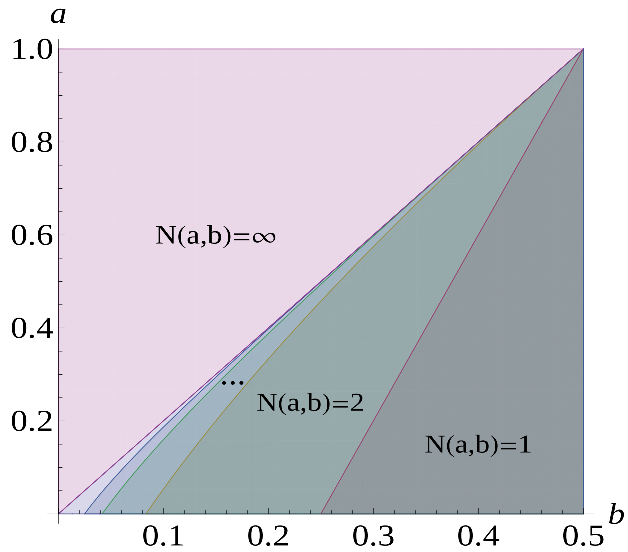


Figure 1: Maximum Number of Equilibrium Intervals

3.1 Welfare Results of the Uniform Case

In this section, we compare the welfare obtained in our model with that in Goltsman et al. (2009). Goltsman et al. (2009) show that in the uniform-quadratic setting, the maximum welfare for a Bayesian receiver across all communication protocols, denoted as $\mathrm{EU}_R^{\mathrm{B}}(b)$, is $-\frac{1}{3}b(1-b)$. This upper bound can also be achieved in Blume et al. (2007) via a noisy communication channel and in Krishna and Morgan (2004) with multi-stage communication.

The following theorem states that with a properly chosen degree of prior bias, $\mathrm{EU}_R(a,b)$ exceeds the upper bound $\mathrm{EU}_R^{\mathrm{B}}(b)$.

Theorem 2. *Given any $b \in (0, \frac{1}{2})$, there exists $a^* \in (0, 2b)$ and $a^{**} \in (2b, 1)$ such that $EU_R(a,b) > EU_R^B(b)$ for any $a \in (a^*, a^{**})$.*

Proof. See Appendix A. □

The main source of the welfare gain relative to Goltsman et al. (2009) is the fact that prior bias garbles the information only in the head of the interim receiver, which relieves the conflict of interest between parties without actually contaminating the content of the messages. On one hand, because of this information garbling, the sender is encouraged to send more informative messages since the link between his message and the receiver's response is weakened. On the other hand, since the information is not garbled in reality, the receiver always receives the right message to begin with, and thus the mismatch between messages and actions that comes with standard garbling devices is avoided. In this sense, prior bias can be regarded as a nonstandard, cheaper information-garbling device.

Note that the existence of ∞-step equilibria plays an important role in Theorem 2. Recall that $a = 2b$ is the boundary condition under which the most informative equilibrium induces infinitely many steps. By Theorem 2, this boundary condition always ensures that the resulting ∞-step equilibrium strictly Pareto-dominates the upper bound $\mathrm{EU}_R^{\mathrm{B}}(b)$. In Blume et al. (2007), ∞-step equilibria exist when there

	$\mathrm{EU}_R^{\mathrm{CS}}$	$\xrightarrow[\text{effect}]{\text{strategic}}$	EU_R^{0}	$\xrightarrow[\text{effect}]{\text{distortionary}}$	EU_R
Prior Bias $a = 0.2$	-0.0308	(+0.0100)	-0.0208	**(-0.0025)**	-0.0233
Blume et al. $\epsilon = \frac{1}{126}$	-0.0308	(+0.0028)	-0.0280	**(-0.0020)**	-0.0300

Figure 2: Welfare Comparison versus Blume et al. (2007)

is a sufficient level of noise. In their ∞-step equilibria, although the sender is very informative with his messages, a large amount of information is lost in the transmission before the receiver further discounts the content. In our model, by contrast, the fact that the messages are delivered verbatim allows both parties to gain from the fine information partition induced in the ∞-step equilibria, making the equilibria much more attractive.

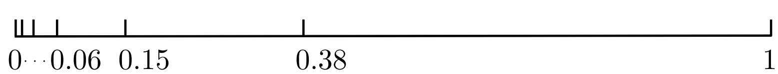
To understand the source of the welfare gain more systematically, we conduct a simple decomposition exercise on the effects of prior bias as an information-garbling device. On the positive side, we define any changes in welfare resulting from changes in the equilibrium partition as the *strategic effect*. On the negative side, information garbling, broadly defined, distorts the actions of the receiver, which we call the *distortionary effect*.

In particular, given any equilibrium partition, let y_0 be the best response of the receiver if no noise or bias is introduced, which represents the information content of the sender's messages. Then, in any interval-partitional equilibrium with information garbling, quadratic utility implies that

$$\mathrm{EU}_{\mathrm{R}} \quad = \quad -\mathbb{E}[(y_0 - \theta)^2] - \mathbb{E}[(y - y_0)^2].$$

The first term, $-\mathbb{E}[(y_0 - \theta)^2]$, is the receiver's payoff given the equilibrium partition if no noise or bias were introduced. The second term, $-\mathbb{E}[(y - y_0)^2]$, measures the welfare loss due to distortion of actions induced by information garbling.

Example 1. Let $a = 0.2$ and $b = 0.1$. We have that $\theta^* = 0$ and $\alpha = \frac{3+\sqrt{5}}{2}$. The information partition induced by the ∞-step equilibrium is illustrated as follows:



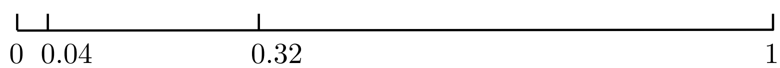
Then, $\mathrm{EU}_R \approx -0.0233 > -0.03 = -\frac{1}{3}b(1-b)$. In the CS model, the Pareto-optimal equilibrium (with two steps) achieves $\mathrm{EU}_R^{\mathrm{CS}} \approx -0.0308$. Given the partition above, if the receiver were to select the Bayesian response, her payoff would be $\mathrm{EU}_{\mathrm{R}}^0 \approx -0.0208$. The difference, $\mathrm{EU}_R^0 - \mathrm{EU}_R^{\mathrm{CS}} \approx 0.0100$, captures the strategic effect. The remaining change, $\mathrm{EU}_R - \mathrm{EU}_R^0 \approx -0.0025$, represents the distortionary effect.

Figure 2 compares the welfare decomposition between Example 1 in our model and Example 1 in

Prior Bias $a = 0.2$	-0.0308	(+0.0100)	-0.0208	**(-0.0025)**	-0.0233
	EU_R^{CS}	$\xrightarrow[\text{effect}]{\text{strategic}}$	EU_R^{0}	$\xrightarrow[\text{effect}]{\text{distortionary}}$	EU_R
Endogenous Noise $a = 0.2$	-0.0308	(+0.0100)	-0.0208	**(-0.0225)**	-0.0433

Figure 3: Welfare Comparison versus the Endogenous Noise Model

Blume et al. (2007).[11] In Blume et al. (2007), the receiver is Bayesian but the message from the sender is replaced with probability ϵ by a random draw from the uniform distribution over the message space. In their Example 1, Blume et al. (2007) consider a very low level of noise $\epsilon = \frac{1}{126}$ and show that the three-step equilibrium with the following partition achieves the welfare bound for $b = 0.1$:



Compared with this example, on one hand, our Example 1 generates a substantially larger strategic effect (+0.0100 versus +0.0028) due to the fact that a sufficient level of information garbling ($a = 0.2$) leads to a much finer equilibrium partition. On the other hand, since information is not garbled in reality, the distortionary effect of a relatively large a in our model is similar to that of a very small ϵ in Blume et al. (2007) (-0.0025 versus -0.0020).

To better compare prior bias with noise, we introduce the *endogenous noise model* as a second benchmark: Suppose the receiver is Bayesian but the message from the sender reaches the receiver only with probability $1-a$. With probability a, the message delivered to the receiver is drawn randomly from the ex ante marginal distribution of equilibrium messages.[12] This model implements the same updating behavior as equation (1.1) by injecting real noise into the communication channel. In fact, the endogenous noise model generates exactly the same equilibrium partitions and ex ante distribution of actions as the prior bias model does. As a result, comparing the welfare achieved in the endogenous noise model and that in the prior bias model highlights the difference between standard information garbling devices and prior bias.

Figure 3 compares the welfare decomposition between the ∞-step equilibrium in our model and that in the endogenous noise model for $a = 0.2$ and $b = 0.1$. First, observe that both models achieve the same level of strategic effect (+0.0100), simply because the same equilibrium partition is generated. Second, the endogenous noise model incurs a much larger distortionary effect than the prior bias model (-0.0225 versus -0.0025). This difference in the distortionary effect is significant enough to bring the welfare in the endogenous noise model (-0.0433), which is far from the upper bound (-0.0300), up to the level that beats

[11] We define the distortionary effect differently from Blume et al. (2007). Our distortionary effect is the sum of the direct effect and the distortion effect in Blume et al. (2007). We combine the effects since (i) the direct effect also distorts the receiver's action away from her optimal response given the equilibrium partition; (ii) no direct effect is induced in the prior bias model.

[12] In other words, the "incorrect" mental model of the receiver under our mistrust interpretation now matches the reality.

it (-0.0233).

Note that since the two models induce the same partition and ex ante distribution of actions, the difference in the distortionary effect boils down to the difference in $cov(y, y_0)$—how well the receiver's actions match the sender's messages. In the endogenous noise model, conditioning on the event that the true message is delivered, the covariance between y and y_0 exactly coincides with the unconditional covariance in the prior bias model. However, conditioning on the noise event, the action taken by the receiver is independent of the true state of the world, and thus, the covariance between y and y_0 is zero. Thus, the endogenous noise model always generates a worse match between actions and messages compared with the prior bias model.

To summarize, the introduction of prior bias may create an agreement type between the sender and the receiver, which leads to the existence of an ∞-step equilibrium, and thus more informative messages sent from the sender. In addition, since prior bias only affects the mental model of the receiver, it avoids the mismatch between messages and actions that comes with standard information garbling devices. Note that the existence of the ∞-step equilibrium with a sufficient level of prior bias, and the fact that the prior bias model Pareto-dominates the endogenous noise model, hold irrespective of the underlying distribution of types.

Nevertheless, the optimal information garbling in Goltsman et al. (2009) does not exactly match the endogenous noise model. While our general intuition that prior bias avoids the mismatch between messages and actions still hold, whether prior bias dominates optimal information garbling depends on the underlying distribution. With uniform prior, the sender's incentive compatibility constraints directly impose an upper bound on the covariance between equilibrium actions and the state of nature, which is the sufficient statistic for ex ante welfare. As a result, the interim payoff of sender type 0 plays a vital role in determining welfare. In fact, with uniform prior, it can be shown that in any equilibrium, with or without information garbling,

$$-\mathbb{E}[(y_0 - \theta)^2] = g_1(a) \cdot h(\mathrm{IU}_S(0)) - var(\theta) \qquad \text{(Strategic effect)}$$
$$-\mathbb{E}[(y - y_0)^2] = -g_2(a) \cdot h(\mathrm{IU}_S(0)) \qquad \text{(Distortionary effect)}$$

where $\mathrm{IU}_S(0)$ is the interim expected payoff for the type 0 sender, and g_1, g_2, h, and $g_1 - g_2$ are nonnegative, strictly increasing functions within the relevant domain. At $a = \frac{b}{\mu}$, the interim expected payoff $\mathrm{IU}_S(0)$ achieves the maximum possible level, 0, in the ∞-step prior-biased equilibrium, and so it does under optimal information garbling in standard models with $a = 0$.[13] Thus, the welfare gain is mainly due to the fact that $g_1(a) - g_2(a)$ is larger than $g_1(0) - g_2(0)$ for any $a \in (0, 1)$. As a result, for the same level of strategic effect, prior bias always induces a smaller distortionary effect than standard information garbling devices.

In Appendix D, we show that essentially the same arguments apply to distributions such that the reciprocal of the hazard function is linear, and also to distributions close to this class. For these distributions, when $a \in (0, \frac{b}{\mu})$, an increase in a tends to increase welfare since the effective conflict of interest for type 0 is

[13] As is shown in Goltsman et al. (2009), under the uniform-quadratic setting, a mediation policy achieves the upper bound if and only if $\mathrm{IU}_S(0) = 0$.

mitigated. When $a \in (\frac{b}{\mu}, 1)$, an increase in a increases the conflict of interest for type 0, and thus tends to decrease the interim payoff of type 0. However, the effect on welfare may be ambiguous within this range, since $g_1 - g_2$ is increasing in a. As a approaches 1, ex ante welfare will inevitably decrease and approach the level in the babbling equilibrium, simply because all actions that the receiver takes will approach the mean.

4 Optimal Communication Protocol

In this section, we explore the optimal mediation problem for $b \in (0, \frac{1}{2})$.[14] We characterize the maximum welfare achieved by any communication protocol in the presence of prior bias and show that direct communication between the sender and the receiver is optimal when the conflict of interest between parties is small relative to the degree of prior bias. When direct communication is suboptimal, we show that the maximum welfare in this case always exceeds the upper bound obtained by Goltsman et al. (2009).

Suppose that there exists a mediator who can design the joint distribution of actions and types. Then optimal mediation is a probability measure on $Y \times \Theta$ that solves the following optimization problem:

$$\max_{\nu \in \Delta(Y\times\Theta)} \mathrm{EU}_R = -\int_{Y\times\Theta} (y-\theta)^2 \nu(dy, d\theta) \tag{4.1}$$

$$\text{subject to} \quad 1 = \int_Y \nu(dy, \theta), \quad \forall \theta \in \Theta; \tag{Prob}$$

$$\theta = \arg\max_{\hat\theta \in \Theta} \left[-\int_Y (y-\theta-b)^2 \nu(dy, \hat\theta) \right], \quad \forall \theta \in \Theta; \tag{IC$_S$}$$

$$y = (1-a) \cdot \frac{\int_\Theta \theta \nu(y, d\theta)}{\int_\Theta \nu(y, d\theta)} + a \cdot \frac{1}{2}, \quad \forall y \in Y \text{ s.t. } \int_\Theta \nu(y, d\theta) > 0, \tag{IC$_R$}$$

where (Prob) is the feasibility constraint that requires the marginal distribution of types be uniform, (IC$_S$) is the incentive compatibility constraint for the sender to truthfully report his type, and (IC$_R$) ensures that the prior-biased receiver follows the recommendation of the mediator. Since the ex ante welfare is aligned between the sender and the receiver, (4.1) can be recast into the problem of maximizing the sender's ex ante welfare in the presence of a type-dependent conflict of interest on the receiver's side.[15] We say that $\nu \in \Delta(Y \times \Theta)$ is a *mediation rule* if ν satisfies (Prob), (IC$_S$) and (IC$_R$). Given a mediation rule ν, again let $\mathrm{IU}_S(\theta)$ be the interim expected payoff for a sender of type θ, i.e., $\mathrm{IU}_S(\theta) = \mathbb{E}_\nu[U_S(y,\theta)|\theta]$.

Goltsman et al. (2009) consider optimization problem (4.1) with $a = 0$. They first express social welfare as a function of $\mathrm{IU}_S(0)$ using an envelope condition and then show that optimal mediation must assign the maximum possible interim payoff to type 0 and thus set $\mathrm{IU}_S(0) = 0$. In their analysis, type 0 is special because it is the worst type that no other type wants to mimic. This is not always the case when $a > 0$. In

[14] Goltsman et al. (2009) consider the same problem without the incentive constraints for the receiver. They show that if $b \geq \frac{1}{2}$, the optimal policy will be to recommend the prior mean no matter what the sender reports, which clearly satisfies the incentive constraints in our case. Thus, the same policy is also the optimal mediation rule in our case if $b \geq \frac{1}{2}$.

[15] In contrast, Alonso and Rantakari (2013) consider the optimal mediation problem in which the receiver is Bayesian but the sender has a type-dependent conflict of interest. They show that when both parties agree at some *extreme* type, the optimal cheap-talk equilibrium may assign the maximum possible interim payoff to that type, and thus maximize the receiver's welfare.

particular, when $b \in (0, \frac{a}{2})$, a sender of type $\frac{a}{2} - b$, if he manages to pretend to be of type 0, will induce his ideal action, $\frac{a}{2}$.

The following lemma plays a vital role in our analysis of the optimal mediation problem (4.1).

Lemma 1. *Let X and Z be two L^1 random variables such that $\mathbb{E}[X|Z] = Z$. Then, for any $t \in \mathbb{R}$,*

$$\mathbb{E}[Z\mathbf{1}_{X\le t}] \ge \mathbb{E}[X\mathbf{1}_{X\le t}]$$

where $\mathbf{1}_{X\le t} = 1$ if $X \le t$ and 0 otherwise.

Proof. See Appendix A. □

To see why Lemma 1 is relevant in the context of strategic communication, note that under quadratic preferences, in any strategic communication game with a Bayesian receiver, the prior distribution of the state of nature is a mean-preserving spread of the ex ante distribution of equilibrium actions, i.e., $\mathbb{E}[\theta|y] = y$. Then, Lemma 1 implies that given any threshold t, a sender of types below t, on average, induces actions above their types. Specifically, under the uniform-quadratic setting, Lemma 1 provides a lower bound for the additional information rent of type t over type 0, since additional rent is increasing in the average action induced by types below t.[16] In short, this lemma relates an equilibrium object (average Bayesian response) to the primitives (average type) regardless of the communication protocol.

With Lemma 1, we arrive at the following theorem, which characterizes when a mediation rule—i.e., a feasible probability measure—solves the optimization problem (4.1).

Theorem 3. *A mediation rule is optimal if and only if the following two conditions are met:*

(i) $IU_S(\max\{0, \theta^*\}) = 0$;

(ii) $\mathbf{1}_{\theta\le\theta^*} = \mathbf{1}_{y\le\theta^*+b}$ *almost surely.*

Proof. See Appendix A. □

When $\theta^* \le 0$, (ii) is satisfied automatically. In this case, a mediation rule is optimal if and only if it assigns the maximum possible interim payoff to type 0. This is the direct counterpart of Lemma 1 in Goltsman et al. (2009) under the prior bias setting. As is the case in Goltsman et al. (2009), the interim expected payoff of type 0 sender uniquely pins down the ex ante welfare. Hence, any mediation rule that assigns an interim payoff of 0 to type 0 must maximize welfare.

When $\theta^* > 0$, (i) requires that the maximum possible interim payoff be assigned to type θ^*, and (ii) requires that the mediator only recommend actions below $\theta^* + b$ if the sender reports a type below θ^*, and vice versa, essentially segregating the state space into $[0, \theta^*]$ and $[\theta^*, 1]$. These conditions, while unfamiliar to the literature, are consequences of the interplay between Lemma 1 and the sender's incentive

[16] Under the uniform-quadratic setting, incentive compatibility of the sender implies that

$$\mathrm{IU}_S(t) - \mathrm{IU}_S(0) = \int_0^t 2(\mathbb{E}[y|\theta] - (\theta + b))d\theta = 2\mathbb{E}[y\mathbf{1}_{\theta\le t}] - t^2 - 2bt.$$

compatibility constraints. When $\theta^* > 0$, by Lemma 1, the additional information rent of type θ^* over type 0 achieves its lower bound when the average action induced by types below θ^* matches the average type below θ^*. This only happens when (ii) holds. Hence, when (i) and (ii) both hold, the interim payoff of type 0, and thus the ex ante welfare, is maximized.

Let $V(a,b)$ denote the maximum welfare attained in (4.1), given a and b. Recall that the ∞-step equilibrium, if it exists, satisfies both conditions of Theorem 3, since the equilibrium is interval partitional and type θ^* induces his ideal action $\theta^* + b$. Hence, if $b \in (0, \frac{a}{2}]$, the ∞-step equilibrium in the direct-talk game must have already achieved $V(a,b)$.

Theorem 4. (i) *If $b \in (0, \frac{a}{2}]$, then $V(a,b) = \frac{1-a^2}{3+a}\left(\frac{1}{4} - \frac{b^2}{a}\right) - \frac{1}{12}$ and direct communication is optimal.*

(ii) *If $b \in (\frac{a}{2}, \frac{1}{2})$, then $V(a,b) = \frac{1+a}{3+a}\left(b - \frac{1}{2}\right)^2 - \frac{1}{12}$ and direct communication is suboptimal.*

Proof. See Appendix A. □

Thus, when an ∞-step equilibrium is played in the direct talk game, no other communication protocol, including the noisy communication channel (Blume et al., 2007), multi-stage direct talk (Krishna and Morgan, 2004), communication via a trustworthy mediator (Goltsman et al., 2009), and communication via a strategic mediator (Ivanov, 2010), can further improve efficiency. This is against the conventional wisdom that a neutral mediator enhances communication when the receiver is conservative or when she mistrusts the sender. When the level of conservatism/mistrust is high, it may create a state at which the sender and the receiver agree on the optimal action, around which communication can be very effective. In this case, a one-shot direct communication between the parties can already be optimal.

On the other hand, if no ∞-step equilibrium exists, additional mediation may improve information transmission. In this case, we construct an optimal mediation rule in Appendix E. With the help of mediation, the maximum welfare attained in this case is always higher than the upper bound $-\frac{1}{3}b(1-b)$. The results suggest that under uniform-like distributions, prior bias is a cheaper substitute for mediation in generating the ideal action for type 0—prior bias distorts the receiver's response toward the mean when the sender type is 0 but avoids the mismatch between actions and messages. When prior bias is sufficiently large to induce the ideal action for type 0, no mediation is necessary; when prior bias is not sufficient, additional mediation is needed, but the level needed is lower than the case without prior bias.

5 A Discussion of the Welfare Criterion

In this section, we discuss the welfare criterion that we used in our analysis. For any prior p_0 and joint distribution of messages and types p, define

$$q(m,\theta) \equiv (1-a)p(m,\theta) + ap_0(\theta)\int_\Theta p(m,d\theta)$$

for any m and θ; that is, q is the ex ante joint distribution of messages and types according to the endogenous noise model, which is also the incorrect mental model of the interim receiver under our mistrust interpretation. Let $\mathrm{EU}_R^* \equiv \mathbb{E}_q[U_R(\rho(m),\theta)]$ be the ex ante expected payoff of the receiver under the incorrect beliefs q.

Proposition 5. *If $\{\sigma,\rho\}$ is a prior-biased equilibrium, then*

$$EU_R = \frac{1+a}{1-a}EU_R^* + \frac{2a}{1-a}var(\theta).$$

Proof. See Appendix A. □

Proposition 5 suggests that holding a constant, EU_R^* and EU_R are aligned. Hence, the Pareto optimality of the most informative equilibrium in Theorem 1, the characterization of the optimal mediation rule in Theorem 3, and the (sub)optimality of direct communication in Theorem 4, all hold irrespective of the welfare criterion. Although the welfare comparison with cheap talk in Proposition 4 involves different values of a, the result only depends on the existence of the ∞-step equilibrium when $a = \frac{b}{\mu}$. Thus, the counterpart of Proposition 4 holds even if EU_R^* is adopted.

Nevertheless, the interpretation of Theorem 2 relies crucially on using EU_R as the welfare criterion. As the endogenous noise model can be implemented by a mediator, we always have $\mathrm{EU}_R^* \leq \mathrm{EU}_R^B(b)$. This means that the counterpart of Theorem 2 for EU_R^* will not hold. On a brighter note, since the sender is fully time-consistent, EU_S unequivocally measures the ex ante welfare for the sender. Then, by Proposition 2, we can rank the ex ante welfare of the sender in the same order as in Theorem 2, and conclude safely that the receiver's prior bias may increase the ex ante welfare of the *sender* beyond the corresponding bound achieved with optimal mediation.

Our rationale for selecting EU_R as the welfare criterion is as follows. First, EU_R evaluates welfare from a rational outside observer's perspective and thus is the natural and normative criterion. Furthermore, while EU_R^* may better represent the preference of a naive receiver who simply considers q to be the correct model, EIU_R represents the preference of a sophisticated receiver who matches the description in Epstein (2006)—the agent is "self-aware and anticipates her updating behaviour when formulating plans." As EU_R and EIU_R are aligned, it is least controversial to use EU_R as the welfare criterion. In fact, our method of evaluating welfare under the correct model is standard in the behavioral economics literature in the presence of time inconsistency. See, for example, O'Donoghue and Rabin (1999).

6 Conclusion

In this paper, we incorporate prior-biased belief updating into the communication environment of CS. To the best of our knowledge, this paper is the first to consider an updating bias in the strategic communication setting. We find that prior bias can be treated as a nonstandard garbling device that weakens the link between the sender's message and the receiver's response and, more importantly, does so without contaminating the actual content of the messages. As a result, society benefits from the strategic advantage of information garbling without paying the full cost. Thus, the efficiency bound characterized in our model exceeds that of Goltsman et al. (2009). Moreover, when the degree of prior bias is sufficient, no additional garbling device is useful to further improve the efficiency of communication—direct communication is optimal in the presence of a sufficient degree of prior bias.

A shared concern in the literature on strategic communication is the dependence on the uniform-quadratic environment. The welfare results in both Blume et al. (2007) and Goltsman et al. (2009) largely rely on the uniform-quadratic assumption, as do most of the welfare results in this paper. Nevertheless, we believe that our intuition still holds in more general settings. First, the introduction of an updating bias may create an agreement type between the sender and the receiver, which leads to the existence of an ∞-step equilibrium or, in other words, more communication. Second, since the updating bias affects only the mental model of the receiver, it may create less distortion of the receiver's action compared with standard implementations of the same equilibrium partition.

References

Abeler, J., D. Nosenzo, and C. Raymond (2019). Preferences for truth-telling. *Econometrica 87*(4), 1115–1153.

Alonso, R. and H. Rantakari (2013). The art of brevity. *Working Paper*.

Benjamin, D. J. (2019). Chapter 2 - errors in probabilistic reasoning and judgment biases. Volume 2 of *Handbook of Behavioral Economics: Applications and Foundations 1*, pp. 69 – 186. North-Holland.

Blume, A., O. J. Board, and K. Kawamura (2007). Noisy talk. *Theoretical Economics 2*(4), 395–440.

Blume, A., D. V. DeJong, Y.-G. Kim, and G. B. Sprinkle (2001). Evolution of communication with partial common interest. *Games and Economic Behavior 37*(1), 79–120.

Blume, A., E. K. Lai, and W. Lim (2020). Strategic information transmission: A survey of experiments and theoretical foundations. In *Handbook of Experimental Game Theory*. Edward Elgar Publishing.

Cai, H. and J. T.-Y. Wang (2006). Overcommunication in strategic information transmission games. *Games and Economic Behavior 56*(1), 7–36.

Cairney, P. (2016). *The Politics of Evidence-Based Policymaking*. London.

Camerer, C. (1995). Individual decision making. In *Handbook of Experimental Economics*. Princeton University Press.

Chen, Y. (2011). Perturbed communication games with honest senders and naive receivers. *Journal of Economic Theory 146*(2), 401–424.

Chen, Y., N. Kartik, and J. Sobel (2008). Selecting cheap-talk equilibria. *Econometrica 76*(1), 117–136.

Crawford, V. P. and J. Sobel (1982). Strategic information transmission. *Econometrica*, 1431–1451.

Dickhaut, J. W., K. A. McCabe, and A. Mukherji (1995). An experimental study of strategic information transmission. *Economic Theory 6*(3), 389–403.

Epstein, L. G. (2006). An axiomatic model of non-Bayesian updating. *Review of Economic Studies 73*(2), 413–436.

Eyster, E. and M. Rabin (2005). Cursed equilibrium. *Econometrica 73*(5), 1623–1672.

Forges, F. (1985). Correlated equilibria in a class of repeated games with incomplete information. *International Journal of Game Theory 14*(3), 129–149.

Geller, E. S. and G. F. Pitz (1968). Confidence and decision speed in the revision of opinion. *Organizational Behavior and Human Performance 3*(2), 190–201.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review 95*(1), 384–394.

Gneezy, U., A. Kajackaite, and J. Sobel (2018). Lying aversion and the size of the lie. *American Economic Review 108*(2), 419–53.

Goltsman, M., J. Hörner, G. Pavlov, and F. Squintani (2009). Mediation, arbitration and negotiation. *Journal of Economic Theory 144*(4), 1397–1420.

Gordon, S. (2010). On infinite cheap talk equilibria. Working Paper.

Grether, D. M. (1978). Recent psychological studies of behavior under uncertainty. *American Economic Review 68*(2), 70–74.

Grether, D. M. (1992). Testing bayes rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior & Organization 17*(1), 31–57.

Ivanov, M. (2010). Communication via a strategic mediator. *Journal of Economic Theory 145*(2), 869–884.

Jehiel, P. and F. Koessler (2008). Revisiting games of incomplete information with analogy-based expectations. *Games and Economic Behavior 62*(2), 533 – 557.

Kahneman, D. and A. Tversky (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology 3*(3), 430–454.

Kahneman, D. and A. Tversky (1973). On the psychology of prediction. *Psychological Review 80*(4), 237.

Kartik, N. (2009). Strategic communication with lying costs. *Review of Economic Studies 76*(4), 1359–1395.

Kartik, N., S. Lee, and D. Rappoport (2019). Single-crossing differences on distributions. Working Paper.

Kartik, N., M. Ottaviani, and F. Squintani (2007). Credulity, lies, and costly talk. *Journal of Economic theory 134*(1), 93–116.

Kawamura, K. (2015). Confidence and competence in communication. *Theory and Decision 78*(2), 233–259.

King, A. (2016). Science, politics and policymaking: Even though expert knowledge has become indispensible for policymaking, providing scientific advice to governments is not always easy. *EMBO reports 17*(11), 1510–1512.

Krishna, V. and J. Morgan (2004). The art of conversation: eliciting information from experts through multi-stage communication. *Journal of Economic Theory 117*(2), 147–179.

Lafky, J., E. K. Lai, and W. Lim (2022). Preferences vs. strategic thinking: An investigation of the causes of overcommunication. *Games and Economic Behavior 136*, 92–116.

Melumad, N. D. and T. Shibano (1991). Communication in settings with no transfers. *RAND Journal of Economics 22*(2), 173–198.

Myerson, R. B. (2013). *Game theory*. Harvard University Press.

O'Donoghue, T. and M. Rabin (1999, March). Doing it now or later. *American Economic Review 89*(1), 103–124.

O'Donoghue, T. and M. Rabin (2003). Studying optimal paternalism, illustrated by a model of sin taxes. *American Economic Review 93*(2), 186–191.

Ortoleva, P. (2012, October). Modeling the change of paradigm: Non-Bayesian reactions to unexpected news. *American Economic Review 102*(6), 2410–2436.

Ottaviani, M. and F. Squintani (2006). Naive audience and communication bias. *International Journal of Game Theory 35*(1), 129–150.

Pitz, G. F. (1969). An inertia effect (resistance to change) in the revision of opinion. *Canadian Journal of Psychology 23*(1), 24.

Pitz, G. F., L. Downing, and H. Reinhold (1967). Sequential effects in the revision of subjective probabilities. *Canadian Journal of Psychology 21*(5), 381.

Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature 36*(1), 11–46.

Szalay, D. (2012). Strategic information transmission and stochastic orders. Technical report, SFB/TR 15 Discussion Paper.

Wang, J. T.-y., M. Spezio, and C. F. Camerer (2010). Pinocchio's pupil: using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American Economic Review 100*(3), 984–1007.

Weible, C. M., T. Heikkila, P. deLeon, and P. A. Sabatier (2012). Understanding and influencing the policy process. *Policy Sciences 45*(1), 1–21.

Zhao, C. (2016). Representativeness and similarity. Working Paper, Princeton University.

Appendices for Online Publication

A Omitted Proofs

Proof of Proposition 1.

Proof. The "if" part of the first statement. Let $\theta^* < 0$. It follows that $y_S(\theta) \neq y_R(\theta)$ for all θ. Then, there exists $\epsilon > 0$ such that if y, y' are two distinct actions induced in equilibrium, $|y - y'| \geq \epsilon$ (see Crawford and Sobel (1982, Lemma 1) for a detailed proof). Hence, $N(a, b) < \infty$.

The "only if" part of the first statement. Now let $\theta^* \geq 0$. We have $[y_R(0), y_R(1)] \subseteq [y_S(0), y_S(1)]$; that is, the sender has an outward bias. See Gordon (2010, Theorem 4) for a detailed proof that if the bias is outward, there is at least one ∞-step equilibrium.

Now, we show that (1) $\theta^* + b$ is the unique limit point of the set of equilibrium actions, and (2) θ^* is the unique type that separates. Note that the existence of an ∞-step equilibrium implies that $\theta^* \geq 0$. We first show that (1) $\theta^* + b$ is the unique limit point of the set of equilibrium actions, and (2) θ^* is the unique type that separates. Let $y(\theta)$ be the action induced by θ in equilibrium and y be a limit point of $\{y(\theta)\}_{\theta\in[0,1]}$ and let $\{y_n\} \subseteq \{y(\theta)\}_{\theta\in[0,1]}$ be a monotone sequence that converges to y. For any n, there exists θ_n such that $y_S(\theta_n) = \theta_n + b$ is in between y_n and y_{n+1}. Otherwise, all $\theta \in [0, 1]$ would share the same strict preference over y_n and y_{n+1}, which implies that y_n and y_{n+1} cannot both be induced in equilibrium. By construction, θ_n converges to some $\widetilde{\theta}$. It is clear that $y_S(\widetilde{\theta}) = y$. By incentive compatibility, we must have $y(\widetilde{\theta}) = y$ in equilibrium. Otherwise, there always exists n such that a sender with type $\widetilde{\theta}$ prefers y_n to $y(\widetilde{\theta})$.

Now, we show that $\widetilde{\theta}$ must separate. Suppose not. Without loss of generality, assume that $\{\theta : y(\theta) = y\} = [\underline{\theta}, \overline{\theta}]$ with $\overline{\theta} > \underline{\theta}$. Then, by optimality of the receiver,

$$y = (1 - a)\gamma(\underline{\theta}, \overline{\theta}) + a\mu.$$

Then, by definition of an interval-partitional equilibrium, for all $\theta \in [0, \underline{\theta}) \cup (\overline{\theta}, 1]$,

$$y(\theta) \in [0, (1 - a)\underline{\theta} + a\mu) \cup ((1 - a)\overline{\theta} + a\mu, 1].$$

Because p_0 has a positive density everywhere, there cannot exist $\{y_n\} \subseteq \{y(\theta)\}_{\theta\in[0,1]}$ such that $y_n \to y$, leading to a contradiction. Hence, $\widetilde{\theta}$ must separate, and thus, $y(\widetilde{\theta}) = \widetilde{\theta} + b = (1 - a)\widetilde{\theta} + a\mu$, which implies that $\widetilde{\theta} = \theta^*$. Hence, $\theta^* + b$ is the unique limit point of $\{y(\theta)\}_{\theta\in[0,1]}$ and (1) is done.

Suppose $\theta \neq \theta^*$ separates. Then, there always exists $\theta' \neq \theta$ such that $y_S(\theta') = y_R(\theta)$. Hence, θ' would strictly prefer mimicking θ to sending his own equilibrium message. Thus, (2) is established.

By Theorem 5 of Gordon (2010), since there is one agreement type, the set of actions induced in equilibrium is at most countable. By (1), there can be at most a finite number of pooling intervals in $[0, \theta^* - \varepsilon] \cup [\theta^* + \varepsilon, 1]$, and a countably infinite number of pooling intervals in $[\theta^* - \varepsilon, \theta^* + \varepsilon]$ for any $\epsilon > 0$. Then the proof is completed by noting that θ^* cannot have a non-degenerate neighboring pooling interval on either side; otherwise, types that are very close to θ^* have an incentive to deviate to θ^*. □

Proof of Theorem 1.

Proof. First, if $\theta^* < 0$, we are back to the CS world, in which the sender's first-best action and the receiver's "full-information" best response do not coincide. Then, by Crawford and Sobel (1982, Lemma 3 and Theorem 3), Condition (M) implies (i) and (ii).

If $\theta^* \geq 0$, Condition (M*) implies that for any finite n, all n-step equilibria share the same set of boundary types and therefore are outcome equivalent. To show (i), it suffices to show that all ∞-step equilibria also share the same set of boundary types. Let $\{\theta_i\} \cup \{\theta^*\} \cup \{\theta_i'\}$ and $\{\tau_i\} \cup \{\theta^*\} \cup \{\tau_i'\}$ each characterizes an ∞-step equilibria in the manner of Proposition 1. If $\tau_1 \neq \theta_1$, by Condition (M*), τ_i and θ_i cannot both converge to θ^*—a contradiction. Hence, $\tau_i = \theta_i$ for all i. Similarly, $\tau_i' = \theta_i'$ for all i. Hence, the two equilibria are outcome equivalent.

Now, we show that ex ante, the receiver prefers the ∞-step equilibrium to any other equilibrium. First, we will establish a proxy for welfare. In any equilibrium, the joint distribution of action y and type θ determines welfare. Thus, by the optimality of the receiver, the ex ante payoff of the receiver is given by

$$\begin{aligned}
\mathrm{EU}_R &= \mathbb{E}[-(y-\theta)^2] \\
&= \mathbb{E}[-(y-\mathbb{E}[\theta]+\mathbb{E}[\theta]-\theta)^2] \\
&= -var(y) + 2cov(y,\theta) - var(\theta) \\
&= -var((1-a)\mathbb{E}[\theta|y] + a\mathbb{E}[\theta]) + 2cov((1-a)\mathbb{E}[\theta|y] + a\mathbb{E}[\theta], \mathbb{E}[\theta|y]) - var(\theta) \\
&= (1-a^2)var(\mathbb{E}[\theta|y]) - var(\theta) \\
&= (1-a^2)\mathbb{E}[\mathbb{E}[\theta|y]^2] - (1-a^2)\mathbb{E}[\theta]^2 - var(\theta)
\end{aligned}$$

where expectations, variances, and covariances are taken with respect to the joint distribution of actions and types. Clearly, EU_R is increasing in $\mathbb{E}[\mathbb{E}[\theta|y]^2]$. For any (finite dimensional or infinite dimensional) vector $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots)$ such that $\theta_{i-1} \leq \theta_i$ for all $i \geq 1$, define

$$W(\boldsymbol{\theta}) = \sum_{i \geq 1} \gamma(\theta_{i-1}, \theta_i)^2 (F(\theta_i) - F(\theta_{i-1})).$$

For any $\lambda \in [0,1]$, let $\boldsymbol{\theta} \wedge \lambda = (\min\{\theta_0, \lambda\}, \min\{\theta_1, \lambda\}, \dots)$ and $\boldsymbol{\theta} \vee \lambda = (\max\{\theta_0, \lambda\}, \max\{\theta_1, \lambda\}, \dots)$.

By Condition (M), for each finite n, any n-step equilibrium is characterized by the same set of boundary types $\{\theta_i\}_{i=0}^n$. For all such $\boldsymbol{\theta}$, we claim that

$$W(\boldsymbol{\theta}) \leq W(\boldsymbol{\theta} \wedge \theta^*) + W(\boldsymbol{\theta} \vee \theta^*). \tag{A.1}$$

In fact, $W(\boldsymbol{\theta})$ is the proxy for welfare in the equilibrium characterized by $\{\theta_i\}_{i\geq 0}$, while $W(\boldsymbol{\theta} \wedge \theta^*) + W(\boldsymbol{\theta} \vee \theta^*)$ is the proxy for welfare if we insert θ^* into the set of boundary types. To establish the claim, it suffices to show that for any $\tau < \tau' < \tau''$,

$$\gamma(\tau, \tau'')^2 (F(\tau'') - F(\tau)) \leq \gamma(\tau, \tau')^2 (F(\tau') - F(\tau)) + \gamma(\tau', \tau'')^2 (F(\tau'') - F(\tau'))$$

which reads

$$1 \leq \frac{x^2}{z} + \frac{(1-x)^2}{1-z},$$

where $x = \frac{\int_\tau^{\tau'} \theta p_0(d\theta)}{\int_\tau^{\tau''} \theta p_0(d\theta)}$ and $z = \frac{F(\tau') - F(\tau)}{F(\tau'') - F(\tau)}$. It is easy to see that the inequality is true if $z \in (0,1)$.

Now, we prove that the ∞-step equilibria Pareto-dominates any equilibrium with fewer steps. There are three steps to prove this. For any $\lambda \in (0,1)$ and $N \in \mathbb{Z}^+$, a sequence $\{\tau_i\}_{i=0}^N$ is a *λ-forward (λ-backward) solution* if it is a forward (backward) solution to (2.2) with $\tau_0 = 0$ ($\tau_0 = 1$) and $\tau_N \geq \lambda$ ($\tau_N \leq \lambda$).

Step 1: Let $\theta^* > 0$ and $\{\theta_i\} \subseteq [0,\lambda) \cup (\lambda, 1]$ be a λ-forward solution such that $\theta_1 < \lambda < \theta^*$. Then, $\frac{dW(\boldsymbol{\theta} \wedge \lambda)}{d\theta_1} < 0$.

First, note by Condition (M) that $W(\boldsymbol{\theta} \wedge \lambda)$ is completely pinned down by θ_1. Additionally, note that we exclude λ from the domain to avoid differentiability concerns when the cardinality of $\{\theta_i \wedge \lambda\}$ changes. This situation does not create any problem with our analysis, because $W(\boldsymbol{\theta} \wedge \lambda)$ is continuous in θ_1. For simplicity of exposition, let $K = \#\{i : \theta_i < \lambda\}$. Let $\tau_i = \theta_i$ for all $0 \leq i \leq K-1$ and $\tau_K = \lambda$. Thus, $W(\boldsymbol{\tau}) = W(\boldsymbol{\theta} \wedge \lambda)$.

$$
\begin{aligned}
\frac{dW(\boldsymbol{\theta} \wedge \lambda)}{d\theta_1} &= \frac{dW(\boldsymbol{\tau})}{d\tau_1} \\
&= \sum_{i=1}^{K} \left[2\gamma(\tau_{i-1},\tau_i)\frac{d\gamma(\tau_{i-1},\tau_i)}{d\tau_1}(F(\tau_i) - F(\tau_{i-1})) + \gamma(\tau_{i-1},\tau_i)^2 (f(\tau_i)\frac{d\tau_i}{d\tau_1} - f(\tau_{i-1})\frac{d\tau_{i-1}}{d\tau_1}) \right] \\
&= \sum_{i=1}^{K} \gamma(\tau_{i-1},\tau_i) \left[2\frac{d\gamma(\tau_{i-1},\tau_i)}{d\tau_1}(F(\tau_i) - F(\tau_{i-1})) + \gamma(\tau_{i-1},\tau_i)(f(\tau_i)\frac{d\tau_i}{d\tau_1} - f(\tau_{i-1})\frac{d\tau_{i-1}}{d\tau_1}) \right] \\
&= \sum_{i=1}^{K} \gamma(\tau_{i-1},\tau_i) \left[(2\tau_i - \gamma(\tau_{i-1},\tau_i)) f(\tau_i) \frac{d\tau_i}{d\tau_1} - (2\tau_{i-1} - \gamma(\tau_{i-1},\tau_i)) f(\tau_{i-1}) \frac{d\tau_{i-1}}{d\tau_1} \right] \\
&= \sum_{i=1}^{K-1} f(\tau_i)\frac{d\tau_i}{d\tau_1} \left[\gamma(\tau_{i-1},\tau_i)(2\tau_i - \gamma(\tau_{i-1},\tau_i)) - \gamma(\tau_i,\tau_{i+1})(2\tau_i - \gamma(\tau_i,\tau_{i+1})) \right] \\
&= \sum_{i=1}^{K-1} f(\tau_i)\frac{d\tau_i}{d\tau_1} (\gamma(\tau_i,\tau_{i+1}) - \gamma(\tau_{i-1},\tau_i))(\gamma(\tau_{i-1},\tau_i) + \gamma(\tau_i,\tau_{i+1}) - 2\tau_i),
\end{aligned}
$$

where the second-to-last equality is due to the fact that $\frac{d\tau_0}{d\tau_1} = \frac{d\tau_K}{d\tau_1} = 0$. By $\{\theta_i\}$ is a forward solution to (2.2), for $i = 1, 2, 3, \ldots, K-2$,

$$\tau_i = (1-a)\frac{\gamma(\tau_{i-1},\tau_i) + \gamma(\tau_i,\tau_{i+1})}{2} + a\theta^*.$$

Additionally, by Condition (M) and $\theta_K \geq \lambda$,

$$\tau_{K-1} \geq (1-a)\frac{\gamma(\tau_{K-2},\tau_{K-1}) + \gamma(\tau_{K-1},\lambda)}{2} + a\theta^*.$$

Condition (M) ensures that $\frac{d\tau_i}{d\tau_1} > 0$ for all $1 \leq i \leq K-1$. Therefore,

$$\frac{dW(\boldsymbol{\theta} \wedge \lambda)}{d\theta_1} \leq \frac{2a}{1-a} \sum_{i=1}^{K-1} f(\tau_i)\frac{d\tau_i}{d\tau_1}(\gamma(\tau_i,\tau_{i+1}) - \gamma(\tau_{i-1},\tau_i))(\tau_i - \theta^*) < 0.$$

Step 2: Let $\{\theta_i'\} \subseteq [0,\lambda) \cup (\lambda, 1]$ be a λ-backward solution such that $\theta_1' > \lambda > \theta^*$. Then $\frac{dW(\boldsymbol{\theta}' \wedge \lambda)}{d\theta_1'} > 0$.

The proof is symmetric to Step 1.

Step 3: Let $\{\eta_i\}$ and $\{\eta_i'\}$ be the sequences that characterize the set of ∞-step equilibria. Then,

$$
\begin{aligned}
W(\boldsymbol{\eta}) &\geq \sup_{\{\theta_i\} \text{ is a } \theta^*\text{-forward solution}} W(\boldsymbol{\theta} \wedge \theta^*), \\
W(\boldsymbol{\eta}') &\geq \sup_{\{\theta_i'\} \text{ is a } \theta^*\text{-backward solution}} W(\boldsymbol{\theta} \vee \theta^*).
\end{aligned}
$$

We will show the first inequality. The argument for the second is symmetric. First, if $\theta^* = 0$, there is nothing to prove. Suppose $\theta^* > 0$. Let $\{\theta_i\}$ be an arbitrary θ^*-forward solution. Note that $\theta_1 > \eta_1$. Otherwise, by Condition (M), $\theta_i \leq \eta_i < \theta^*$ for all i—a contradiction. Let $\lambda_i = \frac{\eta_i + \eta_{i+1}}{2}$ for $i \geq 1$. By Step 1, $W(\boldsymbol{\eta} \wedge \lambda_i) \geq W(\boldsymbol{\theta} \wedge \lambda_i)$ for all i, since $\theta_1 > \eta_1$. Let $i \to \infty$; then, $\lambda_i \to \theta^*$. It follows that $W(\boldsymbol{\eta}) \geq W(\boldsymbol{\theta} \wedge \theta^*)$.

Consider any N-step equilibrium $\{\theta_i\}_{i=0}^N$. Clearly, $\{\theta_i\}_{i=0}^N$ is a θ^*-forward solution, and $\{\theta_{N-i}\}_{i=N}^0$ is a θ^*-backward solution. Hence,

$$W(\boldsymbol{\theta}) \leq W(\boldsymbol{\theta} \wedge \theta^*) + W(\boldsymbol{\theta} \vee \theta^*) \leq W(\boldsymbol{\eta}) + W(\boldsymbol{\eta}').$$

Hence, any ∞-step equilibrium Pareto-dominates equilibria with fewer steps. □

Proof of Proposition 3.

Proof. Let f be the prior density of θ and F be the corresponding CDF. Let $\gamma_1 = \frac{\partial \gamma(\theta, \theta')}{\partial \theta}$ and $\gamma_2 = \frac{\partial \gamma(\theta, \theta')}{\partial \theta'}$. We first establish the following lemma.

Lemma 2. *If f is continuously differentiable and log-concave, then $\gamma_1(\underline{\theta}, \overline{\theta}) + \gamma_2(\underline{\theta}, \overline{\theta}) \leq 1$ for any $\underline{\theta} < \overline{\theta}$.*

Proof. It is clear that

$$\gamma_1(\underline{\theta}, \overline{\theta}) + \gamma_2(\underline{\theta}, \overline{\theta}) = \frac{[\overline{\theta} - \gamma(\underline{\theta}, \overline{\theta})] f(\overline{\theta})}{F(\overline{\theta}) - F(\underline{\theta})} + \frac{[\gamma(\underline{\theta}, \overline{\theta}) - \underline{\theta}] f(\underline{\theta})}{F(\overline{\theta}) - F(\underline{\theta})}.$$

Thus, it suffices to show that for any $\underline{\theta} < \overline{\theta}$,

$$[\overline{\theta} - \gamma(\underline{\theta}, \overline{\theta})] f(\overline{\theta}) + [\gamma(\underline{\theta}, \overline{\theta}) - \underline{\theta}] f(\underline{\theta}) \leq F(\overline{\theta}) - F(\underline{\theta}). \tag{A.2}$$

Let $G(\theta) = \int_0^\theta F(t) dt$. Intergrating by parts yields

$$\begin{aligned} \text{LHS} &= \frac{G(\overline{\theta}) - G(\underline{\theta}) - (\overline{\theta} - \underline{\theta}) F(\underline{\theta})}{F(\overline{\theta}) - F(\underline{\theta})} f(\overline{\theta}) + \frac{(\overline{\theta} - \underline{\theta}) F(\overline{\theta}) - [G(\overline{\theta}) - G(\underline{\theta})]}{F(\overline{\theta}) - F(\underline{\theta})} f(\underline{\theta}) \\ &= \frac{\left[G(\overline{\theta}) - G(\underline{\theta}) - (\overline{\theta} - \underline{\theta}) F(\underline{\theta})\right] (f(\overline{\theta}) - f(\underline{\theta}))}{F(\overline{\theta}) - F(\underline{\theta})} + (\overline{\theta} - \underline{\theta}) f(\underline{\theta}). \end{aligned}$$

Thus, inequality (A.2) reduces to

$$\frac{f(\overline{\theta}) - f(\underline{\theta})}{F(\overline{\theta}) - F(\underline{\theta})} \leq \frac{F(\overline{\theta}) - F(\underline{\theta}) - (\overline{\theta} - \underline{\theta}) f(\underline{\theta})}{G(\overline{\theta}) - G(\underline{\theta}) - (\overline{\theta} - \underline{\theta}) F(\underline{\theta})}.$$

Given any θ, define

$$g(\delta) \equiv (F(\theta + \delta) - F(\theta) - \delta f(\theta))(F(\theta + \delta) - F(\theta)) - (G(\theta + \delta) - G(\theta) - \delta F(\theta))(f(\theta + \delta) - f(\theta)).$$

Clearly $g(0) = 0$. To show inequality (A.2), it suffices to show that $g'(\delta) \geq 0$ for any $\delta \in [0, 1 - \theta]$. We have

$$g'(\delta) = (F(\theta + \delta) - F(\theta) - \delta f(\theta)) f(\theta + \delta) - (G(\theta + \delta) - G(\theta) - \delta F(\theta)) f'(\theta + \delta).$$

Clearly, $g'(0) = 0$. For $\delta > 0$, $g'(\delta) \geq 0$ is equivalent to

$$\frac{f'(\theta + \delta)}{f(\theta + \delta)} \leq \frac{F(\theta + \delta) - F(\theta) - \delta f(\theta)}{G(\theta + \delta) - G(\theta) - \delta F(\theta)}.$$

By the log-concavity of f, it suffices to show that there is $\xi \in [\theta, \theta+\delta]$ such that

$$\frac{f'(\xi)}{f(\xi)} = \frac{F(\theta+\delta) - F(\theta) - \delta f(\theta)}{G(\theta+\delta) - G(\theta) - \delta F(\theta)},$$

which is just a generalized version of the mean value theorem. Define

$$\begin{aligned} h(x) = \quad & (F(\theta+\delta) - F(\theta) - \delta f(\theta))[G(x) - G(\theta) - (x-\theta)F(\theta)] \\ & - (G(\theta+\delta) - G(\theta) - \delta F(\theta))[F(x) - F(\theta) - (x-\theta)f(\theta)]. \end{aligned}$$

Then, $h(\theta+\delta) = h(\theta) = 0$. By Rolle's theorem, there is $\eta \in (\theta, \theta+\delta)$ such that $h'(\eta) = 0$. Thus

$$\frac{f(\eta) - f(\theta)}{F(\eta) - F(\theta)} = \frac{F(\theta+\delta) - F(\theta) - \delta f(\theta)}{G(\theta+\delta) - G(\theta) - \delta F(\theta)}.$$

Then, by Cauchy's mean value theorem, there is $\xi \in (\theta, \eta)$ such that

$$\frac{f'(\xi)}{f(\xi)} = \frac{f(\eta) - f(\theta)}{F(\eta) - F(\theta)}$$

and we are done. □

Now we proceed to prove Proposition 3. Let $\{\theta_i\}_{i=0}^N$ and $\{y_i\}_{i=1}^N$ solve the following system of equations, with $\theta_{i-1} < \theta_i$ for all $i \geq 1$:

$$\begin{aligned} y_i &= (1-a)\gamma(\theta_{i-1}, \theta_i) + a\mu, \quad \text{for } i \geq 1 \\ \theta_i &= \frac{y_i + y_{i+1}}{2} - b, \quad \text{for } 1 \leq i < N. \end{aligned}$$

By definition, $\{\theta_i\}$ is a forward solution to (2.2).

Given θ_{i-1}, θ_i determines y_i through the first equation, which in turn determines y_{i+1} through the second equation and then θ_{i+1} through the first equation. Differentiating the first equation with respect to y_i, the second equation with respect to θ_i yields

$$\frac{1}{1-a} = \gamma_1(\theta_{i-1}, \theta_i)\frac{d\theta_i}{dy_i} + \gamma_2(\theta_{i-1}, \theta_i)\frac{d\theta_{i-1}}{dy_i}, \text{ for } i \geq 1; \tag{A.3}$$

$$2 = \frac{dy_i}{d\theta_i} + \frac{dy_{i+1}}{d\theta_i}, \text{ for } 1 \leq i < N. \tag{A.4}$$

By Lemma 2, $\frac{d\theta_0}{dy_1} = 0$ implies $\frac{d\theta_1}{dy_1} > 1$. It follows that $\frac{d\theta_1}{dy_2} < 1$ and, hence, $\frac{d\theta_2}{dy_2} > 1$. Inductively, $\frac{d\theta_i}{dy_i} \geq 1$ and $\frac{dy_i}{d\theta_{i-1}} \geq 1$ for any $i \geq 2$. It follows that $\frac{d\theta_i}{d\theta_1} \geq 1$ for any $i \geq 1$, which implies Condition (M*) with $\epsilon = \bar{\tau}_1 - \hat{\tau}_1$. □

Proof of Proposition 4.

Proof. In any equilibrium, the joint distribution of action y and type θ determines welfare. Furthermore, from the proof of Theorem 1, we know that

$$\mathrm{EU}_R(a,b) = (1-a^2)var(\mathbb{E}[\theta|y]) - var(\theta) = \frac{1+a}{1-a}var(y) - var(\theta).$$

in which the expectation and the variances are taken with respect to the joint distribution of actions and types in the $N(a,b)$-step equilibrium. By Proposition 1, it is clear that $var(y)>0$ in any ∞-step equilibrium. Hence, for any $b\in(0,\mu)$, we always have $\mathrm{EU}_R(\frac{b}{\mu},b)=\frac{\mu+b}{\mu-b}var(y)-var(\theta)>-var(\theta)$.

The second step is to show that $\mathrm{EU}_R(0,b)=-var(\theta)$ for any $b\in[\frac{\mu}{2},\mu)$. For any $\theta\in[0,1]$, define

$$V(\theta)\equiv\frac{\gamma(0,\theta)+\gamma(\theta,1)}{2}-\theta.$$

Clearly, $V(0)=\frac{\mu}{2}$ and $V(1)=\frac{\mu-1}{2}$. Furthermore, $V(\theta)$ is continuously differentiable, since γ is continuously differentiable. Log-concavity of the density function ensures that $\gamma_1(\theta,\theta')+\gamma_2(\theta,\theta')\le 1$ for any $\theta<\theta'$. Hence, $\gamma_i(\theta,\theta')<1$ for any $i\in\{1,2\}$ and $\theta<\theta'$. Thus,

$$V'(\theta)=\frac{\gamma_2(0,\theta)+\gamma_1(\theta,1)}{2}-1<0.$$

It then follows that $V(\theta)<\frac{\mu}{2}$ for any $\theta\in(0,1]$. Since $V(0,\theta,\theta'|0,b)\le V(\theta)-\frac{\mu}{2}<0$ for any $\theta<\theta'$ and $b\in[\frac{\mu}{2},\mu)$, we conclude that only the babbling equilibrium exists in standard cheap talk when $b\in[\frac{\mu}{2},\mu)$, and thus $\mathrm{EU}_R(0,b)=-var(\theta)$ when $b\in[\frac{\mu}{2},\mu)$.

Note that log-concavity ensures that the original monotonicity condition in Crawford and Sobel (1982), i.e. Condition (M) with $a=0$, is satisfied. Let $\hat\theta$ be the unique solution to $\gamma(\theta,1)=2\theta$ and let $\hat b=\gamma(0,\hat\theta)/2$. Clearly, $\hat\theta<\frac{\mu}{2}$. Furthermore,

$$V(0,0,\hat\theta|0,\hat b)=V(0,\hat\theta,1|0,\hat b)=0.$$

Hence when $b\in(\hat b,\frac{\mu}{2})$, the most informative equilibrium in standard cheap talk features two pooling intervals. Therefore, for any $b\in(\hat b,\frac{\mu}{2})$,

$$\mathrm{EU}_R(0,b)=-F(\theta)(\gamma(0,\theta)-\mu)^2-(1-F(\theta))(\gamma(\theta,1)-\mu)^2-var(\theta)$$

in which θ satisfies $V(\theta)=b$. Since V is strictly decreasing and continuously differentiable, by the implicit function theorem, it is clear that $\mathrm{EU}_R(0,b)$ is continuous at $b=\frac{\mu}{2}$.

Similarly, log-concavity ensures that the welfare in any n-step equilibrium when $a=\frac{b}{\mu}$, denoted as $\mathrm{EU}_R^n(\frac{b}{\mu},b)$, is continuous in b for any finite n. Note that since $\theta^*=0$, by Theorem 2 in Gordon (2010) and Proposition 1, an n-step equilibrium always exists for any $n\in\mathbb{Z}^+$ along the path. Thus, since $\mathrm{EU}_R(\frac{1}{2},\frac{\mu}{2})>\mathrm{EU}_R(0,\frac{\mu}{2})$, there exists $n\in\mathbb{Z}^+$ such that $\mathrm{EU}_R^n(\frac{1}{2},\frac{\mu}{2})>\mathrm{EU}_R(0,\frac{\mu}{2})$. The claim is then established by noting that $\mathrm{EU}_R^n(\frac{b}{\mu},b)-\mathrm{EU}_R(0,b)$ is continuous in b at $b=\frac{\mu}{2}$ and applying Theorem 1. □

Proof of Theorem 2.

Proof. Consider an arbitrary equilibrium. Without loss of generality, assume that the equilibrium messages are simply recommended actions. Let $y_B\equiv\mathbb{E}[\theta|y]$, which is the Bayesian response to recommendation y. Note that $y=(1-a)y_B+a\mathbb{E}[\theta]$, which implies that $\mathbb{E}[y]=\mathbb{E}[\theta]$. We also know that

$$cov(\mathbb{E}[y|\theta],\theta)=cov(y,\theta)=cov(\mathbb{E}[\theta|y],y)=cov(y_B,y).$$

Thus, from the receiver's point of view,

$$\mathrm{EU}_R = \mathbb{E}[-(y-\theta)^2] = -var(y) + 2cov(y,\theta) - var(\theta) = (1-a^2)\cdot var(y_B) - \frac{1}{12}. \tag{A.5}$$

On the other hand, from the sender's point of view, by the envelope theorem,

$$\begin{aligned}\mathrm{EU}_R = \mathrm{EU}_S + b^2 &= \mathrm{IU}_S(0) + 2\int_0^1 E[y|\theta](1-\theta-b)d\theta + b^2 - \frac{1}{3} && \text{(A.6)}\\ &= \mathrm{IU}_S(0) - 2\int_0^1 E[y|\theta]\theta d\theta + 1 - b + b^2 - \frac{1}{3} && \text{(A.7)}\\ &= \mathrm{IU}_S(0) - 2(1-a)var(y_B) - b + b^2 + \frac{1}{6}. && \text{(A.8)}\end{aligned}$$

See Goltsman et al. (2009) for the detailed argument behind equation (A.6). It follows from equations (A.5) and (A.8) that

$$\mathrm{EU}_R = \frac{1+a}{3+a}\left[\mathrm{IU}_S(0) + \left(\frac{1}{2}-b\right)^2\right] - \frac{1}{12}. \tag{A.9}$$

Now, let $a = 2b$. By Proposition 1, $N(a,b) = \infty$ and type 0 separates in the ∞-step equilibrium. It is clear that in this case, $\mathrm{IU}_S(0) = 0$, which implies that

$$\mathrm{EU}_R(2b,b) = \frac{1+2b}{3+2b}\left(\frac{1}{2}-b\right)^2 - \frac{1}{12} > \frac{1}{3}\left(\frac{1}{2}-b\right)^2 - \frac{1}{12} = \mathrm{EU}_R^{\mathrm{B}}(b).$$

Now, it suffices to show that $\mathrm{EU}_R(a,b)$ is continuous in a at $a = 2b$.

We first show that $\lim_{a\downarrow 2b}\mathrm{EU}_R(a,b) = \mathrm{EU}_R(2b,b)$. For any $a > 2b$, it is clear that $\theta^* > 0$, and thus, $N(a,b) = \infty$. By Proposition 8 in Appendix C, in any ∞-step equilibrium, we have

$$\theta_1 = \left(\frac{1}{2}-\frac{b}{a}\right)\left(\frac{2\sqrt{a}}{1+\sqrt{a}}\right),$$

Thus,

$$\mathrm{IU}_S(0) = -\left((1-a)\left(\frac{1}{2}-\frac{b}{a}\right)\left(\frac{\sqrt{a}}{1+\sqrt{a}}\right) + \frac{a}{2} - b\right)^2 = -a\left(\frac{1}{2}-\frac{b}{a}\right)^2,$$

which goes to 0 as $a \downarrow 2b$. It follows from (A.9) that $\lim_{a\downarrow 2b}\mathrm{EU}_R(a,b) = \mathrm{EU}_R(2b,b)$.

Then, we show that $\lim_{a\uparrow 2b}\mathrm{EU}_R(a,b) = \mathrm{EU}_R(2b,b)$. When $a < 2b$, we have $N(a,b) < \infty$. In the $N(a,b)$–step equilibrium, by Proposition 7 in Appendix C,

$$\theta_1 = \frac{(1-\theta^*)(\alpha^1 - \alpha^{-1}) - \theta^*(\alpha^{N(a,b)-1} - \alpha^{1-N(a,b)})}{\alpha^{N(a,b)} - \alpha^{-N(a,b)}} + \theta^*.$$

in which $\alpha = \frac{1+\sqrt{a}}{1-\sqrt{a}}$ and $N(a,b)$ is the largest integer such that

$$\frac{-\alpha-1}{\alpha^n + \alpha^{1-n} - \alpha - 1} < \theta^*.$$

It follows that

$$N(a,b) = \lceil \log_\alpha \frac{(\alpha+1)(1-\theta^*) + \sqrt{(\alpha+1)^2(1-\theta^*)^2 - 4\alpha\theta^{*2}}}{-2\theta^*} - 1\rceil.$$

It is clear that $\lim_{a\uparrow 2b} N(a,b) = \infty$, and thus, $\lim_{a\uparrow 2b}\theta_1 = 0$. It follows that $\lim_{a\uparrow 2b}\mathrm{IU}_S(0) = 0$, and thus, $\mathrm{EU}_R(a,b)$

is continuous in a at $a = 2b$. In fact, it is easy to see that $\mathrm{EU}_R(a,b)$ is continuous in a for any $a \in (0,1)$. $\square$

Proof of Lemma 1.

Proof. By the properties of the conditional expectation and the fact that $\mathbb{E}[X|Z] = Z$, we have

$$\mathbb{E}[X\mathbf{1}_{Z\le t}] = \mathbb{E}[Z\mathbf{1}_{Z\le t}]$$

for any $t \in \mathbb{R}$. It then follows that

$$\begin{aligned}\mathbb{E}[Z\mathbf{1}_{X\le t}] - \mathbb{E}[X\mathbf{1}_{X\le t}] &= \mathbb{E}[Z\mathbf{1}_{X\le t}] - \mathbb{E}[X\mathbf{1}_{X\le t}] - \mathbb{E}[Z\mathbf{1}_{Z\le t}] + \mathbb{E}[X\mathbf{1}_{Z\le t}] \\ &= \mathbb{E}[(Z-X)(\mathbf{1}_{X\le t} - \mathbf{1}_{Z\le t})] \ge 0\end{aligned}$$

for all $t \in \mathbb{R}$, since $(Z-X)(\mathbf{1}_{X\le t} - \mathbf{1}_{Z\le t})$ is always nonnegative. $\square$

Proof of Theorem 3.

Proof. We first prove the case when $\theta^* \ge 0$. Consider the following envelope condition implied by (IC_S):

$$\mathrm{IU}_S(\theta^*) = \mathrm{IU}_S(0) + \int_0^{\theta^*} 2(\mathbb{E}_\nu[y|\theta] - (\theta + b))d\theta. \tag{A.10}$$

Now, let $y_B \equiv \mathbb{E}_\nu[\theta|y]$, the Bayesian response to recommendation y. Since $\mathbb{E}_\nu[\theta|y_B] = y_B$, Lemma 1 implies that

$$\int_0^{\theta^*} \mathbb{E}_\nu[y_B|\theta]d\theta = \mathbb{E}_\nu[y_B\mathbf{1}_{\theta\le\theta^*}] \ge \mathbb{E}_\nu[\theta\mathbf{1}_{\theta\le\theta^*}] = \int_0^{\theta^*} \theta d\theta. \tag{A.11}$$

Combining inequality (A.11), equation (A.10), and $y = (1-a)y_B + a\cdot\frac{1}{2}$ yields

$$\mathrm{IU}_S(\theta^*) - \mathrm{IU}_S(0) \ge \int_0^{\theta^*} 2a(\theta^* - \theta)d\theta. \tag{A.12}$$

By $\mathrm{IU}_S(\theta^*) \le 0$,

$$\mathrm{IU}_S(0) \le -\int_0^{\theta^*} 2a(\theta^* - \theta)d\theta.$$

Note that equation (A.9) also holds with mediation. Hence,

$$\begin{aligned}\mathrm{EU}_R &\le \frac{1+a}{3+a}\left[\left(\frac{1}{2} - b\right)^2 - 2a\int_0^{\theta^*}(\theta^* - \theta)d\theta\right] - \frac{1}{12} \\ &= \frac{1+a}{3+a}\left[\left(\frac{1}{2} - b\right)^2 - a{\theta^*}^2\right] - \frac{1}{12} \\ &= \frac{1-a^2}{3+a}\left(\frac{1}{4} - \frac{b^2}{a}\right) - \frac{1}{12},\end{aligned} \tag{A.13}$$

in which the maximum of EU_R is attained if and only if (i) $\mathrm{IU}_S(\theta^*) = 0$, and (ii) inequality (A.11) is binding. The proof of Lemma 1 implies that (A.11) is binding if and only if

$$(y_B - \theta)(\mathbf{1}_{\theta\le\theta^*} - \mathbf{1}_{y_B\le\theta^*}) = 0 \quad \text{almost surely,}$$

which is equivalent to

$$\mathbf{1}_{\theta \le \theta^*} = \mathbf{1}_{y_B \le \theta^*} \quad \text{almost surely.} \tag{A.14}$$

Then, (ii) is a direct consequence of (A.14) and $y = (1-a)y_B + a\mu$.

Now, suppose $\theta^* < 0$. This case is similar to Goltsman et al. (2009). First note that condition (ii) is automatically satisfied. Second, by equation (A.9) and $\mathrm{IU}_S(0) \le 0$,

$$\mathrm{EU}_R \le \frac{1+a}{3+a}\left(b - \frac{1}{2}\right)^2 - \frac{1}{12}. \tag{A.15}$$

The inequality above is binding if and only if $\mathrm{IU}_S(0) = 0$. It suffices to show that $\mathrm{IU}_S(0) = 0$ is possible whenever $\frac{a}{2} < b < \frac{1}{2}$, which is given by Proposition 10 in Appendix E. □

Proof of Theorem 4.

Proof. When $b \in (0, \frac{a}{2}]$, it suffices to show that the $N(a,b)$-step equilibrium exactly achieves the upper bound given in (A.13). Recall that in any interval-partitional equilibrium of the direct-talk game,

$$\mathrm{EU}_R = \frac{1+a}{3+a}\left[\mathrm{IU}_S(0) + \left(\frac{1}{2} - b\right)^2\right] - \frac{1}{12}.$$

By Proposition 8, θ_1, the first cutoff in any ∞-step equilibrium, is $\theta^*(1-\alpha^{-1})$. Hence, the welfare of the ∞-step equilibrium is given by

$$\begin{aligned} \mathrm{EU}_R &= \frac{1+a}{3+a}\left[-\left[(1-a)\frac{\theta^*(1-\alpha^{-1})}{2} + a\theta^*\right]^2 + \left(\frac{1}{2} - b\right)^2\right] - \frac{1}{12} \\ &= \frac{1+a}{3+a}\left[\left(\frac{1}{2} - b\right)^2 - a\left(\frac{1}{2} - \frac{b}{a}\right)^2\right] - \frac{1}{12} \\ &= \frac{1-a^2}{3+a}\left(\frac{1}{4} - \frac{b^2}{a}\right) - \frac{1}{12}. \end{aligned} \tag{A.16}$$

When $b \in (\frac{a}{2}, \frac{1}{2})$, it suffices to show that the welfare bound given by (A.15) cannot be achieved by direct talk. By Theorem 3, the welfare bound is achieved if and only if $\mathrm{IU}_S(0) = 0$. Thus, in order to achieve the bound, action b must be induced by the first interval in the equilibrium. By equation (C.1), in an n-step equilibrium,

$$\theta_1 = \frac{(1-\theta^*)(\alpha - \alpha^{-1}) - \theta^*(\alpha^{n-1} - \alpha^{1-n})}{\alpha^n - \alpha^{-n}} + \theta^*, \tag{A.17}$$

in which $\alpha = \frac{1+\sqrt{a}}{1-\sqrt{a}}$ and $\theta^* = \frac{1}{2} - \frac{b}{a}$. Thus,

$$\begin{aligned}
y_1 - b &= \frac{(1-a)\theta_1}{2} + \frac{a}{2} - b \\
&= \frac{(1-a)(1-\theta^*)(\alpha-\alpha^{-1})}{2(\alpha^n - \alpha^{-n})} + \frac{(1-a)\theta^*(\alpha^n - \alpha^{-n} - \alpha^{n-1} + \alpha^{1-n})}{2(\alpha^n - \alpha^{-n})} + a\theta^* \\
&= \frac{(1-a)(1-\theta^*)(\alpha-\alpha^{-1})}{2(\alpha^n - \alpha^{-n})} + \frac{(1-a)\theta^*(\alpha^{n-1} + \alpha^{-n})(\alpha-1)}{2(\alpha^n - \alpha^{-n})} + a\theta^* \\
&= \frac{(1-a)(1-\theta^*)(\alpha-\alpha^{-1})}{2(\alpha^n - \alpha^{-n})} + \frac{\sqrt{a}(1+\sqrt{a})\theta^*(\alpha^{n-1} + \alpha^{-n})}{\alpha^n - \alpha^{-n}} + a\theta^* \\
&= \frac{(1-a)(1-\theta^*)(\alpha-\alpha^{-1})}{2(\alpha^n - \alpha^{-n})} + \frac{\sqrt{a}\theta^*(\alpha^{n-1} + \alpha^{-n})}{\alpha^n - \alpha^{-n}} + \frac{a\theta^*(\alpha^{n-1} + \alpha^{-n})}{\alpha^n - \alpha^{-n}} + a\theta^* \\
&= \frac{(1-a)(1-\theta^*)(\alpha-\alpha^{-1})}{2(\alpha^n - \alpha^{-n})} + \frac{\sqrt{a}\theta^*(\alpha^{n-1} + \alpha^{-n})}{\alpha^n - \alpha^{-n}} + \frac{a\theta^*(\alpha^{n-1} + \alpha^{n})}{\alpha^n - \alpha^{-n}} > 0,
\end{aligned}$$

since $a < 1, \theta^* < 0$, and $\alpha > 1$. Hence, in any equilibrium of the direct-talk game, the welfare bound cannot be achieved if $b > \frac{a}{2}$. □

Proof of Proposition 5.

Proof. It is clear from the proof of Theorem 2 that $\mathrm{EU}_R = (1-a^2)\cdot var(y_B) - var(\theta)$. Now consider EU_R^*. We have

$$\mathrm{EU}_R^* = (1-a)\mathrm{EU}_R + a\mathbb{E}_{q'}[-(\rho(m)-\theta)^2],$$

where $q' \equiv p_0(\theta)\int_\Theta p(m, d\theta)$. Hence,

$$\begin{aligned}
\mathrm{EU}_R^* &= (1-a)[(1-a^2)\cdot var(y_B) - var(\theta)] + a[-var(y) - var(\theta)] \\
&= (1+a)(1-a)^2\cdot var(y_B) - a\cdot(1-a)^2 var(y_B) - var(\theta) \\
&= (1-a)^2 var(y_B) - var(\theta),
\end{aligned}$$

where the second equality follows from $y = (1-a)y_B + a\mu$.

Thus,

$$(\mathrm{EU}_R^* + var(\theta))\frac{1+a}{1-a} = \mathrm{EU}_R + var(\theta),$$

and Proposition 5 follows. □

B The NITS Criterion

For the purpose of equilibrium refinement, Chen et al. (2008) propose the *no-incentive-to-separate* (hereafter NITS) condition, which requires that the worst type does not want to deviate by separating himself. In this section, we show that generically, only the most informative equilibrium survives the NITS refinement. Our notion differs slightly from the original NITS condition, in that the worst type in our setting may not always be type 0. More precisely, a type $\theta \in [0,1]$ is said to be *worst* if $U_S(y_R(\theta'), \theta') \geq U_S(y_R(\theta), \theta')$ for any $\theta' \in [0,1]$; that is, θ is a worst type if no other type has strict incentive to mimic θ. Hence, when $\theta^* \geq 0$, the unique worst type is θ^*.

Definition. *An equilibrium* $\{\sigma, \rho\}$ *satisfies the NITS condition if* $U_S(\rho(\sigma(\theta)), \theta) \geq U_S(y_R(\theta), \theta)$ *for any worst type* θ.

The following proposition shows that under mild assumptions on the prior, generically, only the most informative equilibrium satisfies the NITS condition.

Proposition 6. *If* p_0 *has continuously differentiable and log-concave density, then given generic values of* a *and* b*, an interval-partitional equilibrium satisfies the NITS condition if and only if it is an* $N(a,b)$*-step equilibrium.*

Proof. First, note that $\max\{0, \theta^*\}$ is the unique worst type of the game. If $\theta^* < 0$, then Proposition 3 in Chen et al. (2008) applies. Henceforth, we will assume that $\theta^* \geq 0$. Fix finite N and $1 \leq k \leq N$. Given $a \in (0,1)$, let $\{\theta_i\}_{i=0}^N$ satisfy the following system of equations,

$$\begin{cases} \theta_i = (1-a)\frac{\gamma(\theta_{i-1},\theta_i)+\gamma(\theta_i,\theta_{i+1})}{2} + a\theta^*, \text{ for } 1 \leq i \leq N-1, \\ \theta_0 = 0, \\ \theta_n = 1, \\ \gamma(\theta_{k-1}, \theta_k) = \theta^*. \end{cases}$$

If the fourth equation holds, in the interval partitional equilibrium characterized by $\{\theta_i\}_{i=0}^N$, the equilibrium message of type θ^* exactly induces the action $(1-a)\theta^* + a\mu = \theta^* + b$, which implies that a sender of type θ^* does not have an incentive to separate. Substituting the fourth equation into the first yields

$$\begin{cases} \theta_i = (1-a)\frac{\gamma(\theta_{i-1},\theta_i)+\gamma(\theta_i,\theta_{i+1})}{2} + a\gamma(\theta_{k-1}, \theta_k), \text{ for } 1 \leq i \leq N-1, \\ \theta_0 = 0, \\ \theta_n = 1. \end{cases} \tag{B.1}$$

Let $M \subseteq (0,1)^{N-1}$ be the set of all increasing sequences in $(0,1)^{N-1}$. Define $f : M \to \mathbb{R}^{N-1}$ as follows:

$$f_i(\tau_1, \tau_2, \ldots, \tau_{N-1}) = (1-a)\frac{\gamma(\tau_{i-1}, \tau_i) + \gamma(\tau_i, \tau_{i+1})}{2} + a\gamma(\tau_{k-1}, \tau_k)$$

for any $\tau \equiv (\tau_1, \tau_2, \ldots, \tau_{N-1}) \in M$ where $\tau_0 = 0$ and $\tau_N = 1$. It is clear that any solution to (B.1) is a fixed point of f. Note that $f(\tau) \in M$ if $\tau \in M$, since the γ function is increasing in both arguments.

We will show that f has at most one fixed point in M. It then follows that (B.1) can have at most one solution in M. Thus, for any $a \in (0,1)$, there exists at most a countable number of θ^*s such that type θ^* has no incentive to separate in an equilibrium with finite steps. The proof is then completed by noting that θ^* depends only on a and b.

To show that f has at most one fixed point in M, we first show the following lemma.

Lemma 3. *If* p_0 *has continuously differentiable and log-concave density, then* $x_1 < x_2$ *and* $y_1 < y_2$ *implies that*

$$|\gamma(x_1, x_2) - \gamma(y_1, y_2)| \leq \max\{|x_1 - y_1|, |x_2 - y_2|\},$$

where equality is achieved only if $x_1 - y_1 = x_2 - y_2$.

Proof. WLOG assume that $\gamma(x_1, x_2) - \gamma(y_1, y_2) \geq 0$. Let $\epsilon = \max\{x_1 - y_1, x_2 - y_2\}$. If γ is strictly increasing in both arguments, $\epsilon \geq 0$. If $\epsilon = 0$, then $x_1 = y_1$ and $x_2 = y_2$. In this case, clearly the lemma holds. Now suppose $\epsilon > 0$. Note that

$$\gamma(x_1, x_2) - \gamma(y_1, y_2) \leq \gamma(y_1 + \epsilon, x_2) - \gamma(y_1, x_2 - \epsilon),$$

where equality is achieved if and only if $\epsilon = x_1 - y_1 = x_2 - y_2$. Note that by construction, $x_2 > y_1 + \epsilon$ and $x_2 - \epsilon > y_1$.

If p_0 has continuously differentiable and log-concave density, by (A.2), for any $x < y$, we have $\frac{\partial \gamma(x,y)}{\partial x} + \frac{\partial \gamma(x,y)}{\partial y} \leq 1$. By the mean value theorem, if $\epsilon > 0$, then

$$\frac{\gamma(y_1 + \epsilon, x_2) - \gamma(y_1, x_2 - \epsilon)}{\epsilon} \leq 1,$$

which implies that

$$\gamma(x_1, x_2) - \gamma(y_1, y_2) \leq \epsilon.$$

If equality is achieved, then $\epsilon = x_1 - y_1 = x_2 - y_2$. $\square$

Suppose $\tau, \theta \in M$ are distinct fixed points of f. Then

$$\begin{aligned}
& \max_i |f_i(\tau) - f_i(\theta)| \\
= \; & \max_i \left| (1-a)\frac{\gamma(\tau_{i-1}, \tau_i) + \gamma(\tau_i, \tau_{i+1})}{2} + a\gamma(\tau_{k-1}, \tau_k) - (1-a)\frac{\gamma(\theta_{i-1}, \theta_i) + \gamma(\theta_i, \theta_{i+1})}{2} - a\gamma(\theta_{k-1}, \theta_k) \right| \\
\leq \; & \max_i \left| (1-a)\frac{\max\{|\tau_{i-1} - \theta_{i-1}|, |\tau_i - \theta_i|\} + \max\{|\tau_i - \theta_i|, |\tau_{i+1} - \theta_{i+1}|\}}{2} \right| + a\max\{|\tau_{k-1} - \theta_{k-1}|, |\tau_k - \theta_k|\} \\
\leq \; & \max_i |\tau_i - \theta_i| \\
= \; & \max_i |f_i(\tau) - f_i(\theta)|.
\end{aligned}$$

To achieve equality, we must have

$$|\gamma(\tau_{k-1}, \tau_k) - \gamma(\theta_{k-1}, \theta_k)| = \max\{|\tau_{k-1} - \theta_{k-1}|, |\tau_k - \theta_k|\} = \max_i |\tau_i - \theta_i|,$$

which implies that $\tau_{k-1} - \theta_{k-1} = \tau_k - \theta_k$. WLOG assume that $\tau_k - \theta_k \equiv \epsilon > 0$. Then, $\max_i |\tau_i - \theta_i| = \epsilon$.

By $f(\tau) - f(\theta) = \tau - \theta$,

$$(1-a)\frac{\gamma(\tau_{k-1}, \tau_k) + \gamma(\tau_k, \tau_{k+1})}{2} + a\gamma(\tau_{k-1}, \tau_k) - (1-a)\frac{\gamma(\theta_{k-1}, \theta_k) + \gamma(\theta_k, \theta_{k+1})}{2} - a\gamma(\theta_{k-1}, \theta_k) = \tau_k - \theta_k.$$

It follows that

$$\begin{aligned}
\frac{\gamma(\tau_{k-1}, \tau_k) + \gamma(\tau_k, \tau_{k+1})}{2} - \frac{\gamma(\theta_{k-1}, \theta_k) + \gamma(\theta_k, \theta_{k+1})}{2} &= \epsilon \\
\gamma(\tau_k, \tau_{k+1}) - \gamma(\theta_k, \theta_{k+1}) &= \epsilon \\
&= \max\{|\tau_k - \theta_k|, |\tau_{k+1} - \theta_{k+1}|\}.
\end{aligned}$$

It follows that $\tau_{k+1} - \theta_{k+1} = \tau_k - \theta_k = \epsilon$. Inductively, $\tau_n - \theta_n = \epsilon$ for all $k \leq n \leq N-1$. However, by $f_{N-1}(\tau) - f_{N-1}(\theta) =$

$\tau_{N-1} - \theta_{N-1}$, one has

$$\begin{aligned}
\frac{\gamma(\tau_{N-2},\tau_{N-1}) + \gamma(\tau_{N-1},1)}{2} - \frac{\gamma(\theta_{N-2},\theta_{N-1}) + \gamma(\theta_{N-1},1)}{2} &= \epsilon \\
\gamma(\tau_{N-1},1) - \gamma(\theta_{N-1},1) &= \epsilon \\
&= \max\{|\tau_{N-1} - \theta_{N-1}|, 0\},
\end{aligned}$$

which implies that $\tau_{N-1} - \theta_{N-1} = 0$, is a contradiction. Hence, f can have at most one fixed point. □

C Equilibria Characterization in the Uniform Case

We first characterize the set of finite interval-partitional equilibria. Let the set of pooling intervals in equilibrium be $\{[\theta_{i-1}, \theta_i]\}_{i=1}^n$, where $0 = \theta_0 < \theta_1 < \cdots < \theta_n = 1$, and let y_i be the induced action for $\theta \in (\theta_{i-1}, \theta_i)$. Unimodality of the utility function ensures that $y_1 < y_2 < \cdots < y_n$.

Proposition 7. *For $n \in \mathbb{Z}^+$, the following set of equations characterizes an n-step equilibrium if and only if $\theta^* > \underline{\theta}(n)$:*

$$\theta_i = \frac{(1-\theta^*)(\alpha^i - \alpha^{-i}) - \theta^*(\alpha^{n-i} - \alpha^{i-n})}{\alpha^n - \alpha^{-n}} + \theta^* \quad \text{for } i = 0, 1, 2, \ldots, n, \tag{C.1}$$

in which $\alpha = \frac{1+\sqrt{a}}{1-\sqrt{a}}$, $\theta^* = \frac{1}{2} - \frac{b}{a}$, and $\underline{\theta}(n) = \frac{-\alpha-1}{\alpha^n + \alpha^{1-n} - \alpha - 1}$.

Proof. In the uniform case, (2.2) reduces to

$$\theta_{i+1} = \frac{2+2a}{1-a}\theta_i - \theta_{i-1} + \frac{4b-2a}{1-a} \tag{C.2}$$

The characteristic function of the second-order difference equation given by (C.2) is

$$x^2 - \frac{2+2a}{1-a}x + 1 = 0.$$

By $a \in (0,1)$, the characteristic function has two real roots: $\frac{1+\sqrt{a}}{1-\sqrt{a}}$ and $\frac{1-\sqrt{a}}{1+\sqrt{a}}$. Hence, for all i,

$$\theta_i = c\left(\frac{1+\sqrt{a}}{1-\sqrt{a}}\right)^i + c'\left(\frac{1-\sqrt{a}}{1+\sqrt{a}}\right)^i + \frac{1}{2} - \frac{b}{a}.$$

Equation (C.1) then follows from the fact that $\theta_0 = 0$ and $\theta_n = 1$.

To show that $\{\theta_i\}_{i=0}^n$ to characterize an $n-$step equilibrium, we need to verify only that $\theta_i < \theta_{i+1}$ for all i. Let $f(i) = (1-\theta^*)(\alpha^i - \alpha^{-i}) - \theta^*(\alpha^{n-i} - \alpha^{i-n})$. Then,

$$f'(i) \propto (1 - \theta^* + \theta^*\alpha^{-n})\alpha^{2i} + (1 - \theta^* + \theta^*\alpha^n).$$

By $\theta^* \leq \frac{1}{2}$ and $\alpha > 1$, the coefficient of α^{2i} is always positive. If $\theta^* > \frac{1}{1-\alpha^n}$, $f'(i)$ is always positive, and hence, $\{\theta_i\}_{i=1}^n$ characterizes an equilibrium. If $\theta^* \leq \frac{1}{1-\alpha^n}$, we only need $\theta_1 > 0$. Hence,

$$\frac{(1-\theta^*)(\alpha - \alpha^{-1}) - \theta^*(\alpha^{n-1} - \alpha^{1-n})}{\alpha^n - \alpha^{-n}} + \theta^* > 0, \tag{C.3}$$

which reads $\theta^* > \underline{\theta}(n)$.

It is easy to verify that $\underline{\theta}(n) < \frac{1}{1-\alpha^n}$. Therefore, equation (C.1) characterizes an n-step equilibrium if and only if $\theta^* > \underline{\theta}(n)$. The induced actions in any $n-$step equilibrium can differ only at each θ_i, so all $n-$step equilibria are outcome equivalent. The "only if" direction is trivial. □

Now, we characterize the ∞-step equilibrium.

Proposition 8. *If $\theta^* = \frac{1}{2} - \frac{b}{a} \geq 0$, i.e., $b \leq \frac{a}{2}$, then any ∞-step equilibrium is characterized by the set of boundary types $\{\theta_i\}_{i=0}^{\infty} \cup \{\theta^*\} \cup \{\theta_i'\}_{i=0}^{\infty}$ in which $\theta_i = \theta^*(1-\alpha^{-i})$ and $\theta_i' = \theta^* + (1-\theta^*)\alpha^{-i}$.*

Proof. Let $\{\sigma, \rho\}$ be a ∞-step equilibrium. By Proposition 1, θ^* is the unique type that separates. Moreover, the action induced by θ^* is the unique limit point of the set of induced actions. It follows that the set of actions induced in an ∞-step equilibrium is countable.

We first show that for any $\epsilon > 0$, infinitely many actions are induced within $(y(\theta^*), y(\theta^*) + \epsilon)$. Suppose there exists a small $\epsilon > 0$ such that no action is induced within $(y(\theta^*), y(\theta^*) + \epsilon)$. Then, consider a sender of type $\theta^* + \epsilon/2$. Quadratic utility implies that the sender must prefer $y(\theta^*)$ to any other induced actions, which is contradictory to the fact that θ^* separates. Similarly, when $\theta^* > 0$, for any $\epsilon > 0$, infinitely many actions are induced within $(y(\theta^*) - \epsilon, y(\theta^*))$. Hence, we can assume WLOG that $\{y(\theta)\}_{\theta \in [0,1]} = \{y_i\}_{i=1}^{\infty} \cup \{y(\theta^*)\} \cup \{y_i'\}_{i=1}^{\infty}$ where $y_n \uparrow y(\theta^*)$ and $y_n' \downarrow y(\theta^*)$.

Let $\theta_0 = 0$ and θ_i be such that the sender type is indifferent between y_i and y_{i+1}. Then, equation (C.2) holds for any integer $i \geq 1$. The set of solutions is given by

$$\theta_i = c\alpha^i + c'\alpha^{-i} + \theta^* \tag{C.4}$$

for some constants c, c'. We know that $\theta_0 = 0$ and $\theta_i \to \theta^*$. It follows that $\theta_i = \theta^*(1 - \alpha^{-i})$.

Let $\theta_0' = 1$ and θ_i' be such that the sender type is indifferent between y_i' and y_{i+1}'. Still,

$$\theta_i' = c\alpha^i + c'\alpha^{-i} + \theta^*$$

for some constants c and c'. We know $\theta_0' = 1$ and $\theta_i' \to \theta^*$. Thus, $\theta_i' = (1-\theta^*)\alpha^{-i} + \theta^*$.

Now, it suffices to verify that the sender with type θ^* does not want to deviate, which is clear since $y(\theta^*)$ is the most preferred action for type θ^*. □

D Welfare Results for Close-to-Uniform Distributions

Let f be the prior density of θ and F be the corresponding CDF. Consider the differential equation

$$\frac{1 - F(\theta)}{f(\theta)} = \gamma\theta + \delta$$

for some $\gamma, \delta \in \mathbb{R}$. When $\theta = 1$, $\gamma\theta + \delta = 0$, which implies that $\gamma = -\delta$. Thus, at any $\theta \in (0, 1)$,

$$\frac{d \ln(1 - F(\theta))}{d\theta} = \frac{1}{\gamma(1-\theta)}.$$

Since $1 - F(\theta)$ is strictly decreasing in θ, we need $\gamma < 0$. Thus,

$$\ln(1 - F(\theta)) = -\frac{1}{\gamma}\ln(\gamma(\theta - 1)) + c$$

which implies that $1 - F(\theta) = (1-\theta)^{-\frac{1}{\gamma}}$. Let $\beta = -\frac{1}{\gamma}$. We have that $f(\theta) = \beta(1-\theta)^{\beta-1}$ with $\beta > 0$, which is indeed a Beta distribution with $\alpha = 1$.

Proposition 9. *Given any $b \in (0,\mu)$ and $\beta > 0$, there exists $\varepsilon > 0$ such that $\frac{1-\theta}{\beta} - \varepsilon \le \frac{1-F(\theta)}{f(\theta)} \le \frac{1-\theta}{\beta} + \varepsilon$ for any $\theta \in [0,1]$ implies $EU_R(\frac{b}{\mu}, b) > EU_R^B(b)$.*

Proof. Consider the class of distributions such that $\frac{1-\theta}{\beta} - \varepsilon \le \frac{1-F(\theta)}{f(\theta)} \le \frac{1-\theta}{\beta} + \varepsilon$ for some $\beta, \varepsilon > 0$. One one hand, it is clear from the proof of Theorem 2 that

$$\mathrm{EU}_R = (1-a^2)var(y_0) - var(\theta). \tag{D.1}$$

On the other hand, a similar argument to Lemma 3 of Goltsman et al. (2009) yields

$$\mathrm{EU}_R = \mathrm{IU}_S(0) + 2\int_0^1 \mathbb{E}[y|\theta](1 - F(\theta) - bf(\theta))d\theta - \mathbb{E}[\theta^2] + b^2. \tag{D.2}$$

Hence,

$$\mathrm{EU}_R = \mathrm{IU}_S(0) + 2\int_0^1 \mathbb{E}[y|\theta]\left(\frac{1-F(\theta)}{f(\theta)} - b\right)f(\theta)d\theta - \mathbb{E}[\theta^2] + b^2. \tag{D.3}$$

Observe that

$$\begin{aligned}
&\mathrm{IU}_S(0) + 2\int_0^1 \mathbb{E}[y|\theta]\left(\frac{1-\theta}{\beta} - b\right)f(\theta)d\theta - \mathbb{E}[\theta^2] + b^2\\
&= \mathrm{IU}_S(0) - \frac{2}{\beta}\mathbb{E}[y\theta] + 2\left(\frac{1}{\beta} - b\right)\mu - \mathbb{E}[\theta^2] + b^2\\
&= \mathrm{IU}_S(0) - \frac{2}{\beta}(cov(y,\theta) + \mu^2) + 2\left(\frac{1}{\beta} - b\right)\mu - \mathbb{E}[\theta^2] + b^2\\
&= \mathrm{IU}_S(0) - \frac{2}{\beta}((1-a)var(y_0) + \mu^2) + 2\left(\frac{1}{\beta} - b\right)\mu - \mathbb{E}[\theta^2] + b^2\\
&= \mathrm{IU}_S(0) - \frac{2(1-a)}{\beta}var(y_0) - var(\theta) + \hat{g}(\mu,\beta,b)\\
&= \mathrm{IU}_S(0) - \frac{2(1-a)}{\beta}\left(\frac{\mathrm{EU}_R + var(\theta)}{1-a^2}\right) - var(\theta) + \hat{g}(\mu,\beta,b)\\
&= \mathrm{IU}_S(0) - \frac{2}{\beta}\left(\frac{\mathrm{EU}_R + var(\theta)}{1+a}\right) - var(\theta) + \hat{g}(\mu,\beta,b)
\end{aligned}$$

in which $\hat{g}(\mu,\beta,b) = -\frac{2+\beta}{\beta}\mu^2 + 2\left(\frac{1}{\beta} - b\right)\mu + b^2$. Hence,

$$\frac{\beta(1+a)}{2+\beta(1+a)}[\mathrm{IU}_S(0) + \hat{g}(\mu,\beta,b) - 2\mu\varepsilon] - var(\theta) \le \mathrm{EU}_R \le \frac{\beta(1+a)}{2+\beta(1+a)}[\mathrm{IU}_S(0) + \hat{g}(\mu,\beta,b) + 2\mu\varepsilon] - var(\theta).$$

Note that

$$\left(\frac{1-\theta}{\beta} - \varepsilon\right)f(\theta) \le 1 - F(\theta) \le \left(\frac{1-\theta}{\beta} + \varepsilon\right)f(\theta).$$

Integrating over $[0,1]$ yields

$$\frac{1-\mu}{\beta}-\varepsilon \le \mu \le \frac{1-\mu}{\beta}+\varepsilon.$$

It is easy to see that as $\varepsilon \to 0$, $\mu \to \frac{1}{1+\beta}$, and $\hat{g}(\mu,\beta,b) \to (b-\frac{1}{1+\beta})^2$. Thus

$$\lim_{\varepsilon\to 0} \mathrm{EU}_R(a,b) = \frac{(1+a)\beta}{2+(1+a)\beta}\left[\mathrm{IU}_S(0)+\left(b-\frac{1}{1+\beta}\right)^2\right]-var(\theta).$$

The function $g_1(a)-g_2(a)$ in our argument corresponds to $\frac{(1+a)\beta}{2+(1+a)\beta}$, which is strictly increasing in a. Clearly

$$\lim_{\varepsilon\to 0}[\mathrm{EU}_R(b/\mu,b)-\mathrm{EU}_R^B(b)] \ge \left(\frac{(1+b/\mu)\beta}{2+(1+b/\mu)\beta}-\frac{\beta}{2+\beta}\right)\left(b-\frac{1}{1+\beta}\right)^2 > 0$$

and we are done. □

E Optimal Mediation Rule when $b \in (\frac{a}{2},\frac{1}{2})$

Proposition 10. *For any $b \in (\frac{a}{2},\frac{1}{2})$, the following mediation rule is optimal: When $\theta \in [0,\theta_1]$, the mediator recommends action b. For $i=2,\dots,n$, when $\theta \in (\theta_{i-1},\theta_i]$, with probability π, he recommends action b, and with probability $1-\pi$ he recommends action $(1-a)\frac{\theta_{i-1}+\theta_i}{2}+\frac{a}{2}$, where*

$$\begin{aligned}
&\theta_0 = 0, \quad \theta_1 = \frac{(\alpha-1)[2+\theta^*(\alpha^{n-1}+\alpha^{1-n}-2)]}{2(\alpha^n-\alpha^{1-n})},\\
&\theta_i = \frac{(1-\theta^*)(\alpha^{i-1}-\alpha^{1-i})+(\theta_1-\theta^*)(\alpha^{n-i}-\alpha^{i-n})}{\alpha^{n-1}-\alpha^{1-n}}+\theta^*, \quad i=2,\dots,n,\\
&\pi = 1-\frac{1-2b}{(1-\theta_1)(1-a)}
\end{aligned}$$

and n such that

$$\log_\alpha\left(1-\frac{1+\sqrt{1-2\theta^*}}{\theta^*}\right) \le n \le \log_\alpha\left(1-\frac{1+\sqrt{1-2\theta^*}}{\theta^*}\right)+1$$

in which $\alpha = \frac{1+\sqrt{a}}{1-\sqrt{a}}$ and $\theta^* = \frac{1}{2}-\frac{b}{a}$.

Proof. Our construction is similar in spirit to the optimal mediation rule in Goltsman et al. (2009): The mediator recommends action b for the interval containing type 0 and randomizes between b and another action for other intervals.

First, b should be the recommended action for the first interval. Thus,

$$b = (1-a)\left((1-\pi)\frac{\theta_1}{2}+\frac{\pi}{2}\right)+\frac{a}{2}.$$

Let $\delta \equiv 1-(1-a)(1-\pi)$. The above equation reads

$$\theta_1 = \frac{2b-\delta}{1-\delta}.$$

IC_S for θ_1 and IC_R for y_2 together imply

$$\theta_1 = \frac{b + (1-a)\frac{\theta_1+\theta_2}{2} + \frac{a}{2}}{2} - b,$$

which reads

$$(3+a)\theta_1 = (1-a)\theta_2 + 2a\theta^*. \tag{E.1}$$

Note that by construction $\{\theta_i\}_{i\geq 1}$ must be a forward solution to (2.2). Thus, there exist $c, c' \in \mathbb{R}$ such that

$$\theta_i = c\alpha^{i-1} + c'\alpha^{1-i} + \theta^*.$$

Given θ_1 and $\theta_n = 1$, we have

$$\begin{cases} c + c' = \theta_1 - \theta^* \\ c\alpha^{n-1} + c'\alpha^{1-n} = 1 - \theta^*, \end{cases}$$

which yields

$$\begin{cases} c = \frac{1-\theta^*-(\theta_1-\theta^*)\alpha^{1-n}}{\alpha^{n-1}-\alpha^{1-n}} \\ c' = \frac{-1+\theta^*+(\theta_1-\theta^*)\alpha^{n-1}}{\alpha^{n-1}-\alpha^{1-n}}. \end{cases}$$

Thus,

$$\theta_2 = (1-\theta^*)\frac{\alpha-\alpha^{-1}}{\alpha^{n-1}-\alpha^{1-n}} + (\theta_1-\theta^*)\frac{\alpha^{n-2}-\alpha^{2-n}}{\alpha^{n-1}-\alpha^{1-n}} + \theta^*. \tag{E.2}$$

Combining equations (E.1) and (E.2) yields

$$\frac{2\sqrt{a}}{\alpha^{n-1}-\alpha^{1-n}}(1-\theta^*) - \frac{\sqrt{a}(\alpha^{n-1}+\alpha^{1-n})}{\alpha^{n-1}-\alpha^{1-n}}(\theta_1-\theta^*) = \theta_1. \tag{E.3}$$

Hence,

$$\theta_1 = \frac{2b-\delta}{1-\delta} = \frac{\sqrt{a}(2(1-\theta^*)+\theta^*(\alpha^{n-1}+\alpha^{1-n}))}{\alpha^{n-1}-\alpha^{1-n}+\sqrt{a}(\alpha^{n-1}+\alpha^{1-n})},$$

which yields

$$1-\delta = (1-2b)\frac{\alpha^{n-1}-\alpha^{1-n}+\sqrt{a}(\alpha^{n-1}+\alpha^{1-n})}{\alpha^{n-1}-\alpha^{1-n}+\sqrt{a}(1-\theta^*)(\alpha^{n-1}+\alpha^{1-n}-2)}. \tag{E.4}$$

To ensure that $\{\theta_i\}_{i=0}^n$ is an equilibrium, first, we need $\delta \leq 2b$—which, by equation (E.4), implies that

$$\theta^*(\alpha^{n-1}+\alpha^{1-n}) + 2(1-\theta^*) \geq 0,$$

which yields

$$\alpha^{n-1} \leq 1 - \frac{1+\sqrt{1-2\theta^*}}{\theta^*}, \tag{E.5}$$

since $\theta^* < 0$ and $\alpha > 1$. Second, we also need $\delta \geq a$, which requires

$$\begin{aligned}
(1-2b)\frac{\alpha^{n-1}-\alpha^{1-n}+\sqrt{a}(\alpha^{n-1}+\alpha^{1-n})}{\alpha^{n-1}-\alpha^{1-n}+\sqrt{a}(1-\theta^*)(\alpha^{n-1}+\alpha^{1-n}-2)} &\leq 1-a \\
\theta^*\left(\alpha^{n-1}\frac{2\sqrt{a}}{1-\sqrt{a}}-\alpha^{1-n}\frac{2\sqrt{a}}{1+\sqrt{a}}\right) &\leq -2-\theta^*(\alpha^{n-1}+\alpha^{1-n}-2) \\
\theta^*\left(\alpha^{n-1}(\alpha-1)+\alpha^{1-n}(\alpha^{-1}-1)\right) &\leq -2-\theta^*(\alpha^{n-1}+\alpha^{1-n}-2) \\
\theta^*(\alpha^n+\alpha^{-n})+2(1-\theta^*) &\leq 0.
\end{aligned}$$

Since $\theta^* < 0$, the inequality above yields

$$\alpha^n \geq 1-\frac{1+\sqrt{1-2\theta^*}}{\theta^*}. \tag{E.6}$$

Combining inequalities (E.4) and (E.6) yields

$$\log_\alpha\left(1-\frac{1+\sqrt{1-2\theta^*}}{\theta^*}\right) \leq n \leq \log_\alpha\left(1-\frac{1+\sqrt{1-2\theta^*}}{\theta^*}\right)+1. \tag{E.7}$$

Then, it is easy to verify that our construction in the proposition is indeed a feasible mediation rule. □