Divergent Interpretation and Divergent Prediction in Communication*

Shintaro Miura[†] Takuro Yamashita[‡]

May 28, 2018

Abstract

We consider a cheap-talk game à la Crawford and Sobel (1982) with almost-common interest players. The sender's bias parameter is only approximately common knowledge. Compared to the standard case where the structure of the bias parameter is (exactly) common

*This paper was previously circulated as "On the Possibility of Information Transmission". We are grateful to Attila Ambrus, Sidartha Gordon, Tadashi Hashimoto, Daisuke Hirata, Kazumi Hori, Hideshi Itoh, Atsushi Kajii, Shinsuke Kambe, Navin Kartik, Kohei Kawamura, Frederic Koessler, Botond Kőszegi, Takashi Kunimoto, Christoph Kuzmics, Michel Le Breton, Ming Li, Shih En Lu, George Mailath, Kota Murayama, Marco Ottaviani, Daisuke Oyama, Di Pei, Takashi Shimizu, Joel Sobel, Satoru Takahashi, Takashi Ui, Yuichi Yamamoto, and all seminar participants at CTWE (Hitotsubashi), EEA-ESEM 2014 (Toulouse), 20th DC Conference (Fukuoka), ASSET (Aix-Marseille), Game Theory Workshop (Kyoto), and Bocconi, for their helpful comments. We thank the editors and anonymous referees for insightful suggestions. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 714693), and from JPSP Grant-in-Aid for Young Scientists (B: 16K17093). All remaining errors are our own.

[†]Department of Economics, Kanagawa University, smiura@kanagawa-u.ac.jp

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

knowledge, communication between the players is subject to *divergent interpretation* of the sender's messages by the receiver, and *divergent prediction* of the receiver's reaction by the sender. We show that the complementary nature of these phenomena can result in significant welfare consequences even with a "small" (in a certain sense) departure from (exact) common knowledge.

1 Introduction

In real-life communication, mutual miscommunication due to *divergent interpretation* of messages and *divergent prediction* of reactions to messages (and their interactions) are often observed. For example, consider the potential employer of a job market candidate who has a recommendation letter saying that he is "very good". The interpretation of this message is based on the employer's subjective belief about the recommender's characteristics. If the potential employer believes that the recommender is sincere, then this candidate may be hired; otherwise, the employer may discount the recommendation and may not hire him. In this sense, different employers may interpret the same message differently. We call this phenomenon *divergent interpretation.*

Likewise, different recommenders may have different subjective beliefs about the employer's reaction to their messages. We call this phenomenon *divergent prediction.* How to recommend a potential employee is based on the recommender's belief about the employer's interpretation, which can be a source of exaggeration of the recommendation. For example, if the sincere recommender thinks that the employer interprets his message at face value, then he may genuinely say "very good". However, if the recommender believes that the employer may discount his message, then he may say "extremely good" instead of saying "very good", hoping that, after the message is discounted, the employer essentially interprets it (and takes an action) in the way the recommender intends.

Importantly, such an exaggeration of messages owing to divergent prediction may exacerbate the divergent interpretation by further blurring the link between the face values of messages and their intended meaning. In this sense, divergent interpretation and divergent prediction are complementary, and because of this complementarity, these phenomena may have a significant effect on the nature of information that is communicated between players. These phenomena can be observed, not only in the recommendation context, but also in many other contexts, such as advertisements, political campaigns, and so on.

The main objective of this paper is to clarify welfare consequences of miscommunication due to the complementarity of divergent interpretation and divergent prediction, in an environment where the game's payoff structure is only approximately common knowledge. To reach our objective, we consider a situation in which some parameters are not commonly known. Particularly, in the following sense, the relaxation of the common knowledge assumptions in the standard framework is necessary in order to appropriately understand these phenomena and their complementary natures. First, in the standard model, where the payoff structure is common knowledge, in any fixed equilibrium, the intended meaning of each message is interpreted by the receiver exactly the way in which the sender intends (and this feature is common knowledge). In this sense, there is no divergent interpretation. Second, because of the first point, there is also no possibility of divergent prediction. As a consequence, the potential welfare effects of these phenomena are underestimated.[1]

In the approximate common knowledge scenario studied in this paper, it is common knowledge that the bias parameter is in $[-\varepsilon, \varepsilon]$, but that the players do not necessarily agree on the exact value of the bias parameter. In this sense, whenever $\varepsilon > 0$, the players' beliefs about the "true model" are not perfectly aligned, as opposed to the standard common-knowledge setting. Thus, even in an equilibrium, different "types" of the receiver may interpret each message differently. Likewise, different "types" of the sender may predict the receiver's reaction to each message differently. Therefore, even if ε is close to zero so that the model is almost common knowledge, divergent interpretation and divergent prediction (and their interactions) may occur.

We show that these phenomena could significantly distort communication. As a base model, consider a standard cheap-talk model where the preferences of the sender and the receiver are perfectly aligned (i.e., there is no "bias").

[1]This claim is formalized in Section 4.

In this case, it may be reasonable to imagine that the players play a fully revealing equilibrium, which is most informative and optimal in terms of the players' welfare if the model is common knowledge. However, if the model is only imperfectly common knowledge, then the players would be subject to divergent interpretation and divergent prediction, which may significantly confuse information transmission in this equilibrium. More precisely, for any $\varepsilon > 0$, there is a Harsanyi type space that satisfies the following. First, it is common knowledge that the bias is in $[-\varepsilon, \varepsilon]$ in this type space. Second, for any equilibrium such that "full revelation occurs whenever zero bias is commonly believed", *any message may be sent in any state* by some type of sender, and therefore, *any action may be played in any state* by some type of receiver. We observe that this equilibrium exhibits both divergent interpretation and divergent prediction (to be defined later), and their complementarity results in *maximal miscommunication* in the sense that the set of possible equilibrium actions is always unbounded, which is the main result of this paper.[2]

We also show that if an equilibrium does not exhibit divergent interpretation or divergent prediction (i.e., at least one of them), then the set of possible

[2] In Theorem 1, this result is obtained by a type space without common prior about the bias parameter. One may wonder if the same result would be obtained with a type space that is consistent with some common prior, which would be particularly relevant if one is interested in the "ex ante" welfare effects of such miscommunication. Although weaker in a formal sense, we obtain a similar result (Proposition 1) of maximal miscommunication under additional structures of the model, which suggests that absence of a common prior itself is not the crucial source of this phenomenon, but rather relaxation of the common knowledge of the bias parameter is.

equilibrium actions is never unbounded, and furthermore, under reasonable technical assumptions, it is small (in a certain probabilistic sense formally defined later). Thus, the combination of divergent interpretation and divergent prediction is necessary for significant effects of miscommunication.

1.1 Related literature

Our notion of divergent interpretation and divergent prediction is closely related to the concept of *language barriers*. Blume and Board (2013) consider a common-interest cheap-talk game with *language types* of the players, which determine the set of messages that the players can send and understand. They show that, because each player only knows his/her own language type, equilibria exhibit *indeterminacy of indicative* (resp. *imperative*) *meaning*, which indicates that the receiver's (resp. sender's) equilibrium behavior may not be ex post optimal due to uncertainty about language types. Giovannoni and Xiong (2017) and Blume (2018) investigate welfare consequences of language barriers, and point out that language barriers are not impediments to efficient communication if the message space is sufficiently rich (in some sense). More specifically, the former considers an environment à la Crawford and Sobel (1982) in which both players' language types matter. The latter considers the sender's language type in common-interest games, but where its higher-order uncertainty is relevant.[3]

[3] In a broad sense, *noisy-talk games* (Blume, Board, and Kawamura, 2007; Blume and Board, 2014; Lightle, 2014) also represent misspecification of the opponent's behaviors because the receiver may not correctly recognize the sent message due to noise structures.

This paper can be regarded as a complement of these studies. As in the previous works introducing language types as additional uncertainty, we add uncertainty about the bias, which might prevent efficient communication, and study its impacts. In particular, our two distortion properties (divergent interpretation/prediction) imply negative welfare impacts as by the indeterminacy of indicative/imperative meaning of Blume and Board (2013).[4] In contrast, our scope is quite different from theirs. The objective of our paper is to investigate the interactions of the two distortion properties, which is not well discussed in the literature. In fact, the existing works study those properties separately or focus on the environment where they are negligible. Furthermore, our welfare implications about the additional uncertainty is opposite to those in the literature. Specifically, Blume (2018) shows that if uncertainty about the language type is sufficiently small (in some sense), then its impact is negligible (Proposition 3). However, we show that even though the additional uncertainty is sufficiently small, its impact to the welfare is significant.

The impact of uncertain bias in cheap-talk games is also studied in the literature. For example, Morgan and Stocken (2003), Dimitrakas and Sarafidis (2006), Li and Madarász (2008), and Gordon (2010, 2011) analyze environments where the bias parameter is the sender's private information.[5] How-

[4]Divergent interpretation/prediction is defined as disagreement of beliefs about the opponent's behavior. In contrast, indeterminacy of indicative/imperative meaning is defined as deviations from the ex post optimality (due to uncertainty about language types), which does not necessitate disagreed beliefs. As a consequence, we have different implications about these properties, which is discussed later.

[5]These papers study one-shot communication where the sender's preference is state

ever, these papers assume that the bias parameter follows a specific distribution which itself is common knowledge, and in this sense, these papers only consider "low-order" uncertainty. As a result, the complementarity of divergent interpretation and divergent prediction does not arise in those environments, while this complementarity is the main driving force of our welfare implication.[6]

We follow the literature on robust prediction in games with respect to players' high-order beliefs, such as Monderer and Samet (1989), Kajii and Morris (1997), and Morris and Ui (2005), to model approximate common knowledge. These papers perturb players' beliefs such that the base model is common p-belief with p close to (but not) one ("approximate common knowledge").[7] More precisely, we follow its generalized version by Oury and Tercieux (2007), allowing for a small misspecification of the model, possibly with a high probability.[8] These studies investigate whether or not the predicted behavior in the base model is "robust" in the sense that a simi-

dependent. See also Sobel (1985), Benabou and Laroque (1992) and Morris (2001) who study the impact of uncertain bias in repeated interactions, and Diehl and Kuzmics (2014) for the model in which the sender's preference is state independent.

[6]Pei (2015) is close to our setting in that he studies the impact of higher-order uncertainty on the bias parameter. However, in his setting, the bias parameter can take either 0 or $+\infty$, which differs from our setting where it is close to zero. Furthermore, the complementarity between divergent interpretation and divergent prediction is beyond the scope of his paper.

[7]The main differences to those papers are as follows. First, we do not necessarily assume a common prior in modeling the players' perturbed beliefs, while most of the papers in the literature consider common-prior type spaces; and second, we only consider "parametrized" uncertainty (more specifically, only uncertainty about the sender's bias parameter), while other papers usually consider non-parametric uncertainty.

[8]Similar notions of perturbation appear also in Chassang and Takahashi (2011) and Meyer-ter-Vehn and Morris (2011).

lar behavior is an equilibrium even if the model is not common knowledge, but is almost common knowledge in the above sense. For various classes of finite games, they provide conditions for such robustness. Our focus is on an important class of continuous games, specifically cheap-talk games à la Crawford and Sobel (1982), and we show that equilibrium prediction with approximate common knowledge (of the bias parameter) can be very different from the one with (full) common knowledge.[9]

In the literature, another, but a distinct, popular approach to model "nearby" common-knowledge situations exhibits k-th order mutual belief for arbitrary finite (though not infinite) k, such as Rubinstein (1989), Carlsson and van Damme (1993), Weinstein and Yildiz (2007), and so on.[10] Establishing formal relationships among those different approaches in the current context would be an interesting future direction, but it should be noted that, in a general class of games, these versions of nearby common knowledge perturbation is typically strictly more permissive than the "approximate" common knowledge approaches discussed in the last paragraph (which our approach is closer to). For example, in Weinstein and Yildiz (2007), no (nontrivial) event is common knowledge, because at some level k (finite, how large it is), a player's type is *arbitrarily* different from the baseline payoff type, which is crucial to obtain their result that any rationalizable action

[9]Lu (2017) studies the robustness of a fully-revealing equilibrium in the context of multiple-sender models, and obtains a negative implication for its validity.

[10]Chen, Takahashi, and Xiong (2014) study this approach with arbitrary payoff uncertainty, and show that a Cournot oligipoly with three or more firms admits a large set of quantity choices each as uniquely rationalizable outcomes.

of the player can be his unique equilibrium action in a perturbed (in their sense) model. In contrast, our "approximate" common knowledge approach considers a less permissive perturbation (and in this sense "closer" to the exact common knowledge situation) in that the nontrivial event that the payoff parameter is close to the baseline case is indeed common knowledge. Hence, in principle, the set of equilibrium actions in our case can be much smaller than that of all rationalizable actions predicted by Weinstein and Yildiz (2007).

The paper is structured as follows. Section 2 introduces the model where the players face high-order uncertainty on the bias parameter. Section 3 demonstrates that the maximal miscommunication could occur in this environment, and Section 4 clarifies that it is a consequence of the complementary nature of divergent interpretation and divergent prediction. Section 5 concludes the paper. All proofs appear in Appendix A, and an extension beyond the quadratic-loss preferences is discussed in Appendix B.

2 Model

Consider a cheap-talk model à la Crawford and Sobel (1982) with quadratic-loss utilities. There are two players, a sender ($i = 1$) and a receiver ($i = 2$).[11] The sender knows the true payoff-state, $\theta \in \Theta$, while the receiver does not. We consider a message game where the sender sends a message $m \in M$ in

[11] As a convention, we treat the sender as male and the receiver as female throughout this paper.

the first stage, and then the receiver takes an action $a \in A$ in the second stage after observing m. In the following, we assume $\Theta = A = M = \mathbb{R}$ to simplify the argument.[12] The sender's utility is $u(a, \theta, d) = -(a - \theta - d)^2$, and the receiver's utility is $v(a, \theta) = -(a - \theta)^2$, where $d \in \mathbb{R}$ represents the difference in their preferences, called the *bias* parameter.

In the standard model, we assume that d is common knowledge, and θ follows a specific common-knowledge distribution. Instead, we assume that the model is only approximately common knowledge. Specifically, we only assume that it is common knowledge between the players that "d is close to zero", that is, there exists $\varepsilon > 0$ such that it is common knowledge that $|d| \leq \varepsilon$. Let $D = [-\varepsilon, \varepsilon]$. This assumption may be reasonable in a setting where it is common knowledge that each player has a statistically consistent estimator (perhaps based on data) for the bias parameter. As a result of consistency, each players' estimate is close to the true value (with a high probability), and this itself is common knowledge. Nevertheless, it is not necessarily the case that those estimates coincide.

Even with these assumptions, the players' beliefs, represented by a Harsanyi type space denoted by $\mathcal{T} = (T_1, T_2, b_1, b_2)$, could have a rich structure. Specifically, for each $i = 1, 2$, player i's type is an element t_i of a measurable space T_i. His belief mapping is a measurable mapping $b_i : T_i \to \Delta(T_{-i})$, that is, given his type t_i, his belief over T_{-i} is represented by $b_i(t_i) \in \Delta(T_{-i})$.

[12] A similar (though slightly weaker) result holds in a more standard case with bounded Θ and A with appropriate modifications, available from the authors upon request.

Because the sender knows the bias parameter d and the true state θ, let $d(t_1) \in D$ and $\theta(t_1) \in \Theta$ denote, respectively, the bias and the true state when the sender's type is t_1. These mappings $d(\cdot)$ and $\theta(\cdot)$ are assumed to be measurable. We denote by $T = T_1 \times T_2$ the set of type profiles. Given $t_2 \in T_2$, the receiver's marginal belief about d is denoted by $b_2^D(t_2) \in \Delta(D)$, that is, for each measurable set $E \subseteq D$:

$$b_2^D(E|t_2) = \int_{T_1} 1\{d(t_1) \in E\} db_2(t_1|t_2). \tag{1}$$

To represent the common knowledge assumptions introduced above, for every t_2, we assume that $b_2^D(D|t_2) = 1$. Let $\mathbb{T}^{\varepsilon}$ represent the class of the type spaces satisfying those conditions.

In the message game, given $\mathcal{T}$, let $\sigma_1 : T_1 \to M$ denote the sender's pure strategy, $\sigma_2 : T_2 \times M \to A$ denote the receiver's pure strategy,[13] and $\psi : T_2 \times M \to \Delta(\Theta)$ be the receiver's posterior beliefs on the payoff-state. The solution concept is perfect Bayesian equilibrium (PBE, hereafter) defined as follows.

Definition 1. A triple $(\sigma_1^*, \sigma_2^*, \psi^*)$ is a PBE if it satisfies the following conditions:

(i) for any $t_1 \in T_1$,

$$\sigma_1^*(t_1) \in \arg\max_{m \in M} \int_{t_2 \in T_2} u(\sigma_2^*(t_2, m), \theta(t_1), d(t_1)) db_1(t_2|t_1), \tag{2}$$

[13] We focus on pure (and measurable) strategies throughout the paper.

(ii) for any $t_2 \in T_2$ and $m \in M$,

$$\sigma_2^* \in \arg\max_{a \in A} \int_{\theta \in \Theta} v(a, \theta) d\psi^*(\theta | t_2, m); \tag{3}$$

(iii) for any $t_2 \in T_2$ and $m \in M$, $\psi^*(\cdot | t_2, m)$ is a regular conditional probability measure given σ_1^*.[14]

To economize notation, we simply represent equilibrium $(\sigma_1^*, \sigma_2^*, \psi^*)$ as $\sigma^* = (\sigma_1^*, \sigma_2^*)$ unless it is confusing. Let $A(\theta|\sigma^*)$ denote the set of equilibrium actions of the receiver when the true state is θ in equilibrium σ^*, that is:

$$A(\theta|\sigma^*) = \{\, a \in A \mid \exists (t_1, t_2) \in T \text{ s.t. } \theta(t_1) = \theta,\ a = \sigma_2^*\left(t_2, \sigma_1^*(t_1)\right) \,\}. \tag{4}$$

This $A(\theta|\sigma^*)$ is our main object of interest for the remainder of this paper. If $A(\theta|\sigma^*)$ contains an action that is far from θ, we interpret it as a consequence of *miscommunication*. We study how each player's divergent belief about the other player, which can be arbitrarily "small", maps to the *size of miscommunication*, which is defined by $\sup A(\theta|\sigma^*) - \inf A(\theta|\sigma^*)$.[15]

[14] $\Theta = \mathbb{R}$ guarantees that such a regular conditional probability measure exists for any t_2 and m (the *regular conditional probability property*). As is well known, a regular conditional probability measure may not be unique. In particular, given any "off-path" message, any probability measure over Θ would work.

[15] Because we want to compare, for example, $A(\theta|\sigma') = [\theta - \varepsilon, \theta + \varepsilon]$ and $A(\theta|\sigma'') = A$, the definition based on its cardinality is inappropriate.

3 Maximal miscommunication

In the base model where $d = 0$ is common knowledge, it is reasonable to imagine that a fully revealing equilibrium is played.[16] To formalize the idea, we introduce the following additional notation. Given a type space $\mathcal{T} = (T_1, T_2, b_1, b_2) \in \mathbb{T}^\varepsilon$, consider a subset of types $\tilde{T}_i \subseteq T_i$ for each i such that, for each i and $t_i \in \tilde{T}_i$, $b_i(\tilde{T}_{-i}|t_i) = 1$, that is, $\tilde{T} = \tilde{T}_1 \times \tilde{T}_2$ is a *belief-closed* subset. If, in addition, $d(t_1) = 0$ for every $t_1 \in \tilde{T}_1$, we say that *zero bias is commonly believed in* $\tilde{T}$.

Definition 2. A PBE σ^* given $\mathcal{T}$ has *Property FR0* if the following is satisfied: for every belief-closed subset $\tilde{T}$ such that zero bias is commonly believed, every $\theta \in \Theta$, every $t_1 \in \tilde{T}_1$ such that $\theta(t_1) = \theta$, and every $t_2 \in \tilde{T}_2$, we have $\sigma_1^*(t_1) = \theta$ and $\sigma_2^*(t_2, \sigma_1^*(t_1)) = \theta$.

That is, whenever $d = 0$ is commonly believed among the players, full revelation of information occurs. Throughout this section, we fix an equilibrium with Property FR0, denoted by σ^*. It is worthwhile to remark that if $d = 0$ is common knowledge (in $\mathcal{T}$), then Property FR0 immediately implies that $A(\theta|\sigma^*) = \{\theta\}$ for any θ. We are interested in the effects of slightly relaxing the common knowledge assumption in the sense that only $|d| \leq \varepsilon$ is common knowledge. In fact, its effect can be significant, as stated in the following theorem, which is the main result of this paper.

[16] In other words, we focus on the maximum-partitioned equilibrium if $d = 0$ is common knowledge. It is often an implicit convention in cheap-talk games to focus on the maximum-partitioned equilibrium, based on the justification given by the "no incentive to separate" (NITS) condition of Chen, Kartik, and Sobel (2008).

Theorem 1. For each $\varepsilon > 0$, there exists $\mathcal{T} \in \mathbb{T}^{\varepsilon}$ such that: (i) a PBE with Property FR0 exists; and (ii) for any PBE σ^* with Property FR0, we have $A(\theta|\sigma^*) = A$ for every θ.

Theorem 1 shows that if we slightly relax the common knowledge assumption on the bias parameter and focus on a "seemingly natural" equilibrium in which full revelation occurs for those who commonly believe $d = 0$, then we must admit any action choice at any θ. In this sense, the equilibrium exhibits the "maximal" level of miscommunication. Recall that, in our model, it is common knowledge that the bias parameter lies in an arbitrarily small interval, and in this sense, the amount of the players' disagreement about it is arbitrarily small. Nevertheless, this disagreement (no matter how small) implies maximal miscommunication. As we discuss in Section 4, this is a consequence of the complementary nature of *divergent interpretation* (of the sender's messages by different types of the receiver) and *divergent prediction* (of the receiver's reactions by different types of the sender), which we define formally in Section 4.

3.1 Sketch of the proof

Here, we outline the main idea of the proof. The proof basically comprises the following two steps.

First, we construct a Harsanyi type space $\mathcal{T} \in \mathbb{T}^{\varepsilon}$ analogous to the level-k theory, heuristically illustrated in Figure 1. T_i^k in the figure represents the

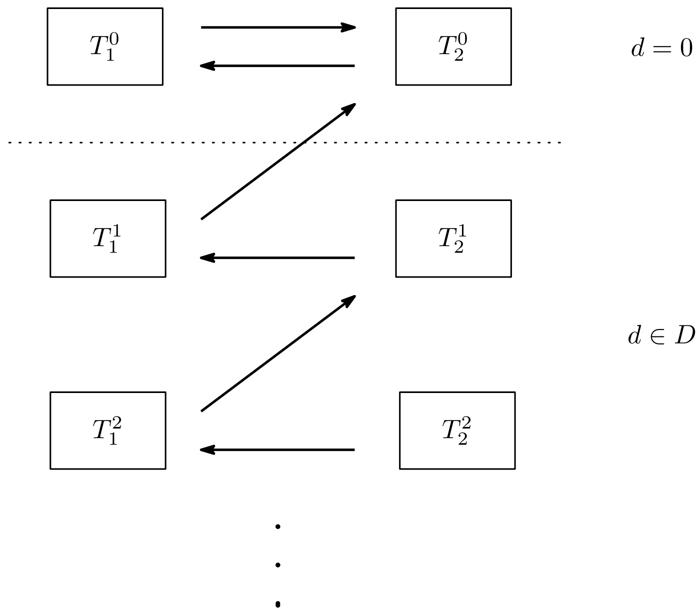

Figure 1: Construction of Harsanyi type space $\mathcal{T}$

set of "level-k types" of player i. First, we define the set of level-0 types of the players, $T^0 = T_1^0 \times T_2^0$, as a belief-closed subset in which zero bias is commonly believed. That is, any level-0 type of sender has no bias and certainly believes that the type of receiver is in T_2^0. Also, any level-0 type of receiver certainly believes that the type of sender is in T_1^0, and therefore, believes that $d = 0$ for certain.

We then define the successive levels as follows. Each level-1 type of sender has bias $d \in D$, and certainly believes that the type of receiver is in T_2^0; that is, the sender could be biased, but he certainly believes that the receiver believes $d = 0$. On the other hand, any level-1 type of receiver certainly believes that the type of sender is in T_1^1; that is, the receiver believes that the sender: (i) could be biased; and (ii) believes that the receiver has a level-0 type. Likewise, any level-2 type of sender could be biased and certainly believes that the type of receiver is in T_2^1, and any level-2 type of receiver certainly believes that the sender's type is in T_1^2, and so on. We then define $T_i = \bigcup_{k=0}^{\infty} T_i^k$. One interpretation may be that a level-0 type is the "naive" type who believes that there is no conflict in their preferences. A type of sender in T_1^k tries to best respond to a type of receiver in T_2^{k-1}, and a type of receiver in T_2^k tries to best respond to a type of sender in T_1^k.[17]

As the second step, given the Harsanyi type space $\mathcal{T}$ constructed above, we construct a PBE with Property FR0 (hence, we obtain existence). Furthermore, we show that $A(\theta|\sigma^*) = A$ for any θ in any PBE σ^* with Property FR0.

The idea of the proof of this second step is as follows. Let $A^k(\theta|\sigma^*)$ denote the set of actions that the receiver in T_2^k can play in an equilibrium if she receives message $m = \theta$. First, because zero bias is commonly believed in T_0, Property FR0 implies that the sender with a level-0 type reports θ truthfully, and then the receiver with a level-0 type takes $a = \theta$. Hence, $A^0(\theta|\sigma^*) = \{\theta\}$ for any θ.

[17] Although our construction of the type space is analogous to that often used in the level-k theory (see Stahl and Wilson (1994, 1995), and Nagel (1995); see Crawford (2003) for its application to strategic communication), interpreting the hierarchical levels as the players' strategic sophistication may not be sensible. For example, the players with type 0 in our type space play an equilibrium, and in this sense they are strategically sophisticated. Rather, the level here may be interpreted as a measure of distance from the base model where $d = 0$ is common knowledge. Strzalecki (2014) adopts the level-k construction to represent each player's higher-order beliefs about the opponents' depth of reasoning.

Now, for the sender with a level-1 type, if he is biased, then he may report untruthfully. In particular, because the sender believes that the receiver has a level-0 type, he reports θ plus his bias. Given this, because the receiver with a level-1 type believes that the sender is of such a type, she adjusts the action according to the bias that she believes, that is, her action is the reported message minus the believed amount of the bias. As a result, $A^1(\theta|\sigma^*) = [\theta - \varepsilon, \theta + \varepsilon]$ for any θ. But then, because the sender with a level-2 type expects such a discounting by the receiver, he would exaggerate his message even more so that the message he sends is the true payoff-state plus his bias plus the discount made by the receiver. Because the receiver with a level-2 type expects such an exaggeration, she discounts more. As a result, $A^2(\theta|\sigma^*) = [\theta - 2\varepsilon, \theta + 2\varepsilon]$ for any θ. By induction, it is shown that $A^k(\theta|\sigma^*) = [\theta - k\varepsilon, \theta + k\varepsilon]$ for any θ. Because $A = \bigcup_k A^k(\theta|\sigma^*) \subseteq A(\theta|\sigma^*)$, $A(\theta|\sigma^*) = A$ for any θ.[18]

Remark 1. Note that, on the equilibrium path, the receiver's belief (on Θ) and actions vary with respect to received messages. In this sense, communication plays a nontrivial role in σ^*, especially *given* the type of the receiver being fixed. In this sense, potential failure of communication in our construction is fundamentally different from a "babbling equilibrium" (i.e., an equilibrium where the sender sends the same message regardless of θ, and

[18]The same result holds even if each level-k type of a player has a non-degenerate belief over any type of the opponent with level k' where $k' < k$, because the best response given such a belief is represented by a convex combination of the best responses with degenerate beliefs.

the receiver plays $a = E(\theta)$ for any received message), where the receiver's belief and action are invariant, given whatever messages are observed.

Remark 2. As demonstrated in Theorem 1, messages sent by the sender with level-k types in equilibrium σ^* are more exaggerated compared with that which the level-0 types send as k increases. This phenomenon may be related to *language inflation* of Kartik, Ottaviani, and Squintani (2007) and Kartik (2009). In their papers, cheap-talk games are transformed into costly-signaling games by introducing naive receivers or lying costs, which makes each message have its own "intrinsic" meaning. They then define language inflation based on this intrinsic meaning. Our work, in contrast, is about pure cheap-talk games where a message does not have any intrinsic meaning, and thus, it is difficult to formally define language inflation and discuss a connection to their work. Nevertheless, our exercise demonstrates that (a broad sense of) language inflation can be a link between divergent interpretation and divergent prediction, and make them complementary. To the best of our knowledge, this feature of language inflation is not well investigated in the literature.

Remark 3. While one might think of Theorem 1 as reminiscent of Weinstein and Yildiz (2007), we would like to emphasize that our result is not a direct implication from their result. Weinstein and Yildiz (2007) consider "nearby" common knowledge situations that are different from the "approximate" common knowledge of ours. Specifically, in their type space, there exists some finite k where each player's k-th order belief is *arbitrarily* different from the baseline common-knowledge case. This k corresponds to their notion of proximity to the baseline common-knowledge case, but notice that k cannot be infinite in any "nearby" type space. Indeed, such arbitrariness at some finite order k is essential for their result that any rationalizable action can be a uniquely equilibrium action for any nearby type space in their sense.

However, clearly, such a type space is *not* in $\mathbb{T}^{\varepsilon}$. Our "approximate" common knowledge situation is more restricted in the sense that any acceptable beliefs must be consistent with the common knowledge that the bias parameter is sufficiently closed to zero. Such an event cannot be common knowledge in the construction of Weinstein and Yildiz (2007). Thus, in principle, the set of predictable actions in our approach could be much smaller than that under the approach by Weinstein and Yildiz (2007).[19]

3.2 Type spaces with common priors

While Theorem 1 shows that the set of equilibrium actions is unbounded, one may have some concerns about it. First, the maximal miscommunication seems to heavily depend on the particular Harsanyi type space con-

[19]To further clarify the difference, it may be useful to consider some different games from cheap-talk games studied in this paper, where these two approaches generate different predictions. For example, imagine a classical "battle-of-the sexes" game, where each agent earns a positive payoff if they play the same action, while failing to do so results in zero payoff. In a type space considered in Weinstein and Yildiz (2007), any action of each player is rationalizable, and hence can be some type's unique equilibrium action. However, in our case, with a sufficiently small payoff perturbation, both players' coordinating on a (strict) Nash equilibrium of the baseline game is an equilibrium. This is precisely due to the property that the nontrivial event that the payoff parameter is close to the baseline case is indeed common knowledge.

structed above. In particular, the construction of our Harsanyi type space is analogous to level-k theory, and we assign infinite belief hierarchies involving systematically wrong beliefs, which may seem a very special type space. Second, because we do not assume a common prior over the type space, Theorem 1 is silent on which actions in the obtained unbounded support are "more likely" than others. In this sense, it is ambiguous how far the equilibrium exhibiting the maximal miscommunication is from the fully revealing equilibrium focused on the baseline scenario. To answer these questions, we consider an environment where there is a common prior over the type space, and demonstrate that the maximal miscommunication could happen even in such a common prior framework. Furthermore, we show that the welfare measured by the receiver's ex ante expected utility is significantly different from that of the fully revealing equilibrium even though we consider only small departure from (exact) common knowledge about the bias is slightly.

Throughout this section, the baseline model is modified as follows. We assume that Harsanyi type space $\mathcal{T} = (T_1, T_2, b_1, b_2)$ admits a common prior $\phi \in \Delta(T)$. That is, for any i, $b_i(\cdot|t_i) \in \Delta(T_{-i})$ coincides with the conditional probability distribution $\phi(\cdot|t_i) \in \Delta(T_{-i})$. The analysis of this model is more complicated than that of the baseline model for the following reasons. First, the receiver's belief updating is two-dimensional: with a message from the sender, the receiver updates her belief jointly about the true payoff-type and the bias. Second, Property FR0 does not have much bite with a common prior: if zero bias is commonly believed in a subset of types, then it is indeed probability zero that a receiver with such a type faces with a biased sender, which induces nothing interesting. The following modification is made in order to circumvent these issues. First, we assume that the payoff-state follows an improper uniform distribution $\mu \in \Delta(\Theta)$, which drastically simplifies the updating of the state part.[20] Second, we adopt a weaker version of Property FR0 defied as follows.

Definition 3. For $\eta > 0$, a PBE σ^* given $\mathcal{T}$ has *Property FR-η* if the following is satisfied: for any $t_1 \in T_1$ and $t_2 \in T_2$ such that (i) $\theta(t_1) = \theta$ and (ii) zero bias is commonly-η believed, we have $\sigma^*(t_1) = \theta$ and $\sigma_2^*(t_2, m) = m$.

The following proposition shows that a similar (though weaker version of) maximal miscommunication could happen even in this common prior framework, which suggests that absence of a common prior itself is not the crucial source of this phenomenon, but rather relaxation of the common knowledge of the bias parameter is. Define $\delta(\varepsilon) = (1-\varepsilon)/\varepsilon^2$.

Proposition 1. For any $\varepsilon > 0$, there exists $\mathcal{T} \in \mathcal{T}^\varepsilon$ which admits a common prior $\phi \in \Delta(T)$ under which there exists a PBE with Property FR-$(1/(1+\varepsilon^2))$ such that (i) $A(\theta|\sigma^*) = [\theta - \delta(\varepsilon), \theta + \delta(\varepsilon)]$, and (ii) the receiver's ex ante expected utility is $V(\sigma^*) = -(1-\varepsilon)^2/(3\varepsilon)$.[21]

The sketch of the proof is as follows. Consider the following type space $\tilde{\mathcal{T}}$. Given $\varepsilon > 0$, let $\tilde{T}_1^0 = \Theta \times \{0\}$, $\tilde{T}_1^1 = \Theta \times (D \backslash \{0\})$, $\tilde{T}_2^0 = \{0\}$, and

[20] Admittedly it is a restrictive assumption, and in this sense, the result in this section is not more than a suggestive exercise. Whether we obtain similar results with other μ is important but well beyond the scope of the paper.

[21] Note that, as $\varepsilon \to 0$, we have $\delta(\varepsilon) \to \infty$ and $-(1-\varepsilon)^2/(3\varepsilon) \to -\infty$.

$\tilde{T}_2^1 = D\backslash\{0\}$. For any i, define $\tilde{T}_i = \tilde{T}_i^0 \cup \tilde{T}_i^1$. The belief of each type is defined as follows. The sender with type $t_1 = (\theta, 0)$ has no bias, knows payoff-state θ, and certainly believes that the receiver is type $t_2 = 0$. Likewise, the sender with type $t_1 = (\theta, d)$ has bias $d \neq 0$, knows payoff-state θ, and believes that the receiver's type is $t_2 = 0$ with probability $\varepsilon^3/(1-\varepsilon)$, and and $t_2 = d$ with probability $(1-\varepsilon-\varepsilon^3)/(1-\varepsilon)$, respectively. In contrast, the receiver with type $t_2 = 0$ believes that (i) the payoff-state follows improper uniform distribution $\mu \in \Delta(\Theta)$, and independent of that, (ii) the bias is 0 with probability $1/(1+\varepsilon)^2$, and it is uniformly distributed over $D\backslash\{0\}$ with the remaining probability. Also, the receiver of type $t_2 = d (\neq 0)$ believes that the payoff-state follows distribution μ, but she certainly believes that the bias is d. With simple algebra, it is easy to confirm that this type space admits a common prior. In particular, conditional on θ, the distribution of types is summarized as in Table 1.

We can show that the following is a PBE in this type space by the standard argument.

$$\sigma_1^*(\theta, d) = \theta + \frac{1-\varepsilon}{\varepsilon^3} d, \tag{5}$$

$$\sigma_2^*(t_2, m) = m - \frac{1-\varepsilon}{\varepsilon^3} t_2. \tag{6}$$

Notice that for the sender of type $t_1 = (\theta, 0)$ and the receiver of type $t_2 = 0$, zero bias is commonly-$(1/(1+\varepsilon^2))$ believed, and the sender reports θ truthfully and the receiver who observes message m takes action $a = m$ in

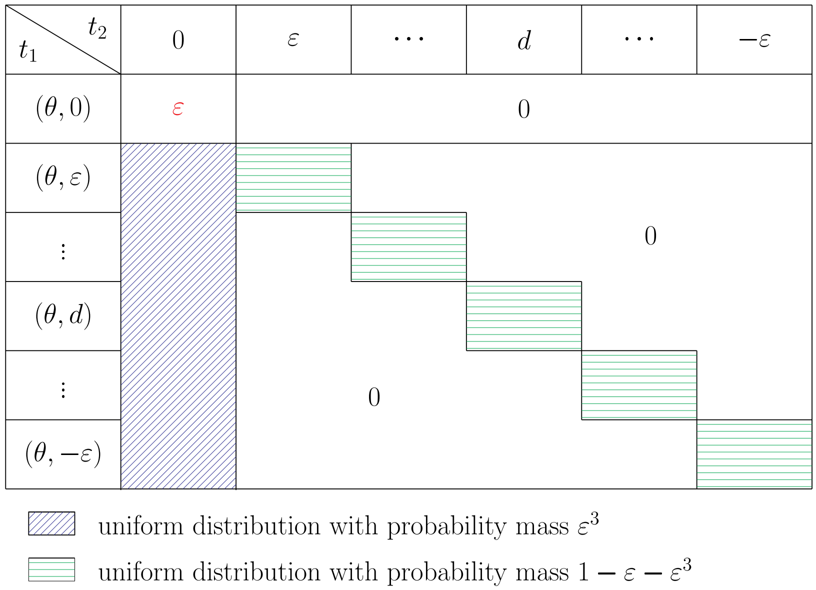

Table 1: The type distribution conditional on θ

this equilibrium. That is, this equilibrium satisfies Property FR-$1/(1 + \varepsilon^2)$.

Because, conditional on θ, the sender of type $t_1 = (\theta, d)$ plays with the receiver of type $t_2 = 0$ with density $\varepsilon^2/2$ for any $d \in D\backslash\{0\}$, action $a = \theta + (1-\varepsilon)d/\varepsilon^3$ is played in this equilibrium. With the remaining probability, the receiver correctly infers the correct payoff-state, and then $a = \theta$ is played. Therefore, we obtain that $A(\theta|\sigma^*) = [\theta - \delta(\varepsilon), \theta, \theta + \delta(\varepsilon)]$, and the receiver's ex ante expected utility $V(\sigma^*) = -(1-\varepsilon)^2/(3\varepsilon)$.

The main message of this proposition is similar to that of Theorem 1. As for the equilibrium constructed in Theorem 1, this equilibrium also exhibits the maximal miscommunication in the sense that $A(\theta|\sigma^*)$ tends to the entire action set A and the receiver's ex ante expected utility tends to $-\infty$ as

$\varepsilon \to 0$. That is, this proposition also shows that communication is significantly distorted even though the disagreement over the bias is arbitrarily small. In particular, the equilibrium constructed here is quite different from the fully revealing equilibrium in terms of the ex ante welfare. In the fully revealing equilibrium, the receiver's ex ante expected utility is 0, which is the best scenario for the receiver. In contrast, in the equilibrium constructed in Proposition 1, the ex ante expected utility tends to $-\infty$, which is the worst case for the receiver. This difference demonstrates that our equilibrium is opposite to the fully revealing equilibrium in this probabilistic framework.[22]

4 The complementarity of divergent interpretation and divergent prediction

4.1 Definition

We have thus far discussed the possibility of maximal miscommunication that arises because of the lack of common priors over the bias. As mentioned above, this phenomenon seems to be related to the complementary nature of divergent interpretation and divergent prediction. In this section, we investigate this observation. To formally define our notion of divergent interpreta-

[22]However, Proposition 1 is weaker than Theorem 1 because the type space varies with ε and improper uniform distribution μ is necessary. Furthermore, this proposition does not assure that any equilibrium satisfying Property FR-η exhibits the maximal miscommunication. Because the comprehensive analysis for common-prior type spaces is well beyond the scope of this paper, it is left for future research.

tion and divergent prediction, we introduce the following additional notation and technical assumptions. Hereafter, we assume the payoff-state follows a full-support common prior distribution $\mu \in \Delta(\Theta)$ with a finite mean. That is, for any $t_2 \in T_2$, $b_2^\Theta(\cdot|t_2) = \mu(\cdot)$ holds, where $b_2^\Theta \in \Delta(\Theta)$ is the receiver's marginal belief about θ.[23] Fix some $\varepsilon > 0$, a Harsanyi type space $\mathcal{T} \in \mathbb{T}^\varepsilon$, and a (measurable, pure-strategy) equilibrium σ. Given σ_1, the receiver with type t_2 can form a joint belief over $\Theta \times D \times M$, based on $b_2(t_2)$. Let $b_2^M(t_2)$ denote its marginal over M, i.e., for each measurable $E \subseteq M$:

$$b_2^M(E|t_2) = \int_{T_1} 1\{\sigma_1(t_1) \in E\} db_2(t_1|t_2). \tag{7}$$

Similarly, given σ_2, for each $t_1 \in T_1$, $m \in M$, and measurable $E' \subseteq A$, let:

$$b_1^A(E'|t_1, m) = \int_{T_2} 1\{\sigma_2(t_2, m) \in E'\} db_1(t_2|t_1), \tag{8}$$

denote the sender's belief about the receiver's action choices when the sender has type t_1 and sends message m. Let $\tilde{M} = \{m \in M | \exists t_1 \in T_1 \text{ s.t. } \sigma_1(t_1) = m\}$ be the set of on-path messages in σ. In the following, for each $S \subseteq \Theta$ with $\mu(S) > 0$, $E[\theta|\theta \in S]$ represents the conditional expected value of θ given $\theta \in S$ *with respect to* μ, that is, $E[\theta|\theta \in S] = \left(\int_S \theta d\mu\right)/\mu(S) \in \mathbb{R}$.

[23] Formally, for each measurable set $E' \subseteq \Theta$:

$$b_2^\Theta(E'|t_2) = \int_{T_1} 1\{\theta(t_1) \in E'\} db_2(t_1|t_2).$$

For simplicity, we only consider equilibria satisfying the following conditions: (i) for each $t_2 \in T_2$, the support of $b_2^M(t_2)$ coincides with $\tilde{M}$; and (ii) there is *no redundant message*, that is, for any $m, m' \in \tilde{M}$ with $m \neq m'$, there exists type $t_1 \in T_1$ such that:

$$\int_{T_2} (\sigma_2(t_2, m) - \theta(t_1) - d(t_1))^2 db_1(t_2|t_1) \neq \int_{T_2} (\sigma_2(t_2, m') - \theta(t_1) - d(t_1))^2 db_1(t_2|t_1). \tag{9}$$

Condition (i) guarantees that the receiver can form a conditional belief over $\Theta \times D$ given any $m \in \tilde{M}$, which we denote by $b_2^{\Theta \times D}(t_2, m)$. Condition (ii) is imposed in order to avoid unnecessary complexity because of the sender's complete indifference. We believe that this non-redundancy condition is not very demanding, in that violation of (ii) implies that the sender is indifferent between m and m' *regardless of his type*.

Definition 4. An equilibrium σ exhibits *divergent interpretation* if there exist types $t_2, t_2' \in T_2$ of the receiver and $m \in \tilde{M}$ such that:

$$b_2^{\Theta \times D}(t_2, m) \neq b_2^{\Theta \times D}(t_2', m).$$

That is, there exist two types of the receiver who have divergent beliefs over the state of nature (not the sender's types), even if they observe the same on-path message.[24] For example, in equilibrium σ with divergent in-

[24]Such divergence in interpretation may be because, for example, some types of the receiver have "wrong" beliefs. However, we take an agnostic position regarding the source

terpretation, t_2 may think that m is sent because θ is high and d is low, while t_2' may think that it is because θ is low and d is high. Obviously, the equilibrium σ^* constructed in Theorem 1 exhibits divergent interpretation.

Definition 5. An equilibrium σ exhibits *divergent prediction* if there exist types $t_1, t_1' \in T_1$ of the sender and $m \in \tilde{M}$ such that $b_1^A(t_1, m) \neq b_1^A(t_1', m)$.

That is, there exist two types of the sender who have divergent beliefs over the receiver's reaction to an on-path message. For example, in equilibrium σ with divergent prediction, t_1 may think that his message "good" is taken literally as "good", while t_1' may think that it is much discounted. Again, equilibrium σ^* constructed in Theorem 1 exhibits divergent prediction.

Remark 4. As mentioned in the literature review, divergent interpretation/prediction is related to indeterminacy of indicative/imperative meaning defined by Blume and Board (2013) in the sense that both of these concepts capture some aspects of communication failure due to additional uncertainty. More specifically, indeterminacy of indicative/imperative meaning describes ex post suboptimality of equilibrium behaviors, which arises because of uncertainty about the opponent's language type. That is, because the opponent's language type is unknown (though its distribution is known) from one's perspective, her/his decision is based on its expectation, which implies ex post suboptimality. Likewise, divergent interpretation/prediction describes that

of this divergence, and focus on its consequence. The same remark applies to the concept of divergent prediction below.

different types have different beliefs about the opponent's behavior. Due to this disagreement, some types may play ex post suboptimal behavior.

While divergent interpretation/prediction is defined as disagreement of beliefs, such disagreements are unnecessary for indeterminacy of indicative/imperative meaning because it is defined directly as deviations from the ex post optimality due to uncertainty about language types. To be more precise, for example, consider a common-interest cheap-talk game in which only the sender has nontrivial language types governed by a common prior. In this model, as shown by Blume and Board (2013), indeterminacy of indicative meaning might occur, but divergent interpretation/prediction do not. Because the receiver's belief should be based on the common prior over the sender's language type, there is no room for belief disagreement.

4.2 Results

Recall that, in the standard cheap-talk setting, $A(\theta|\sigma)$ is a singleton for (almost) every θ in any equilibrium σ. In this sense, the size of this set, for each θ or in expectation with respect to θ, captures the consequence of divergent interpretation or divergent prediction. As demonstrated in Theorem 1, $A(\theta|\sigma^*)$ could be unbounded if an equilibrium exhibits both divergent interpretation and divergent prediction. However, if one of the properties is absent in equilibrium σ, then $A(\theta|\sigma)$ is never unbounded.

Theorem 2. For any type space $\mathcal{T} \in \mathbb{T}^{\varepsilon}$ and equilibrium σ that does not exhibit divergent interpretation or divergent prediction, $A(\theta|\sigma)$ is bounded for every θ.

Theorem 2 says that, as opposed to the case considered in Theorem 1, the "infinite" size of miscommunication does not occur unless there exist both divergent prediction and divergent interpretation. Notice that the statement includes not only cases of certain bias, but also those of uncertain bias (with or without a common prior for the bias parameter). Even in such cases, the "infinite" size of miscommunication does not occur. In this sense, complementarity of divergent interpretation and divergent prediction is crucial. Intuitively, the boundedness arises from the fact that the effect of higher-order uncertainty is drastically limited in the absence of one of the properties.

Although the formal proof is in Appendix A, one of the key observations to obtain this result is that, with the combination of Condition (ii) and the lack of divergent interpretation or divergent prediction, the equilibrium message of the sender with type t_1 becomes only a function of $x = \theta(t_1) + d(t_1)$. Hence, any payoff-irrelevant information contained in t_1 (such as his higher-order beliefs) does not play any role. This is in stark contrast to the equilibrium constructed in Theorem 1, where higher-order beliefs play a central role for unbounded $A(\theta|\sigma^*)$. Given this observation, the boundedness of $A(\theta|\sigma)$ is derived by a similar argument as in the standard cheap-talk games.

In general, the size of $A(\theta|\sigma)$ can be large (even though bounded) at least for some θ. For example, even in the standard setting without uncertain bias, imagine a two-partition equilibrium and type θ who is indifferent between the two messages. However, in this standard setting, such indifferent types are "rare" with respect to the probability measure μ, as long as μ is absolutely continuous. Similarly, in our setting with uncertain bias, the μ-weighted average size of $A(\theta|\sigma)$ is small (in the sense made explicit below), under the following regularity conditions.

Assumption 1. μ admits a (full-support) Lipschitz-continuous density function f with convex tails.[25]

Theorem 3. Suppose that Assumption 1 holds. Then, for any $\delta > 0$, there exists $\varepsilon > 0$ such that for any type space $\mathcal{T} \in \mathbb{T}^{\varepsilon}$ and any equilibrium σ that does not exhibit either divergent interpretation or divergent prediction, there exist $c_1, c_2 \in \mathbb{R}$, such that:

$$\mu([c_1, c_2]) > 1 - \delta, \text{ and} \tag{10}$$

$$E[\sup A^*(\theta) - \inf A^*(\theta)|\theta \in [c_1, c_2]] < \delta. \tag{11}$$

Theorem 3 says that the expected size of $A(\theta|\sigma)$ (except for a set of θ in the tails whose probability can be made arbitrarily small) can be bounded with an arbitrary small bound $\delta > 0$, as long as μ satisfies the regularity condition. Therefore, we conclude that, even if the players face (high-order)

[25] f has convex tails if there exists $c \geq 0$ such that f is convex on $(-\infty, -c)$ and (c, ∞). Many popular parametric density functions satisfy this property (as well as Lipschitz continuity). See, for example, Lam and Mottet (2016).

uncertainty, if only divergent interpretation or divergent prediction occurs, it has little impact on communication.

Remark 5. While our focus is the complementary nature of divergent interpretation and divergent prediction itself, the characterization of each phenomenon seems an interesting open question. An obvious sufficient condition (on the primitive) for not having these phenomena is to assume a common prior, denoted $\phi \in \Delta(\Theta \times D)$, over the payoff-state and the bias, as in the standard cheap-talk models. Note that, with such a common prior, there is no higher-order uncertainty (more precisely, conditional on the players' first-order beliefs, higher-order beliefs are degenerated). This is different from the type spaces in Theorem 1 and in Proposition 1, where at least second-order beliefs are non-degenerated.

Claim 1. If there is common prior $\phi \in \Delta(\Theta \times D)$ over $\Theta \times D$, then any equilibrium exhibits neither divergent interpretation nor divergent prediction.

Notice that when d is common knowledge or it follows a common prior even if it is uncertain, the receiver's reaction is uniquely determined independent of her type. Thus, no equilibrium exhibits divergent prediction. Furthermore, because divergent prediction never appears, the receiver's behavior is uniquely specified, which means that it is unnecessary for the sender to distinguish the receiver's types. As a consequence, no equilibrium exhibits divergent prediction. Because the comprehensive characterization of these phenomena is well beyond the scope of this paper, it is left for future research.

5 Concluding remarks

This paper studies a certain class of cheap-talk games à la Crawford and Sobel (1982), where it is common knowledge that the bias parameter is close to zero. We show the following results. First, there exists a type space in which for any PBE where full revelation occurs among the types of players who commonly believe the bias to be zero, *any* action may be played by some type of receiver in *any* payoff-state. This result may be interpreted as maximal miscommunication. We also demonstrate that a similar phenomenon could happen even if a common prior over the type space is assumed, where the equilibrium is quite different from the fully revealing equilibrium in terms of the ex ante welfare. Thus, we insist that the main reason of such a miscommunication is the high-order uncertainty for the bias parameter. Second, we show that such significant miscommunication is a consequence of the complementary nature of divergent interpretation and divergent prediction. That is, if one of these properties is absent, then the size of miscommunication is much smaller even if players face higher-order uncertainty. These results shed light on the potential welfare impacts of divergent interpretation and divergent prediction in cheap-talk communication.

There are two concluding remarks. First, our first result is shown in a rather unique setting. For example, the sets of states and actions are unbounded, and $d = 0$ (rather than some nonzero value) is fixed as a "benchmark" case. It is left open for future research as to how much our result generalizes to alternative specifications in this manner.

Additionally, we focus on the fully revealing equilibrium among many other equilibria when $d = 0$ is common knowledge. It would be interesting to study if similar results hold even if we assume other (non-fully revealing) equilibrium behaviors in the base model. If the conclusion ends up varying across different assumptions on the equilibrium behaviors, the methodology offered in this paper could be interpreted as providing a new dimension in which to compare equilibria in the base model, based on behavior in nearby environments. This may be useful for arguing equilibrium selection and the ranking of equilibria.

A related question is whether or not a similar interaction of divergent interpretation and divergent prediction occurs in other environments. This paper focuses on a cheap-talk environment, but the fundamental concept of this paper also applies to various other environments. For example, in a companion paper, Miura and Yamashita (2014) show a similar result in a costly-signaling example, à la Spence (1973). More specifically, if we assume that a fully-separating equilibrium occurs when the model parameters are common knowledge, any outcome is possible in any payoff-state in nearby environments. Because the comprehensive analysis in a general environment is beyond the scope of this paper, it is left for future research.

Appendix A: Proofs

A.1 Proof of Theorem 1

We first construct $\mathcal{T} = (T_1, T_2, b_1, b_2)$ as follows. First, let $T_1^0 = \{t_1^0(\theta)|\theta \in \Theta\}$ be a subset of types of the sender, which we refer to as "level-0" types, where for each θ, $t_1^0(\theta)$ is a type of the sender who:(i) has $d = 0$; (ii) knows the payoff-state θ; and (iii) believes that the receiver's type is in T_2^0, that is:

$$d(t_1^0(\theta)) = 0, \ \theta(t_1^0(\theta)) = \theta, \text{ and } b_1(T_2^0|t_1^0(\theta)) = 1. \tag{12}$$

Let $T_2^0 = \{t_2^0\}$, where t_2^0 is a "level-0" type of the receiver, who believes that the sender's type is in T_1^0 (i.e., $b_2(T_1^0|t_2^0) = 1$).[26] Note that $d = 0$ is commonly believed among them.

Next, for each $d \in D$, let $T_1^1(d) = \{t_1^1(d, \theta)|\theta \in \Theta\}$ be another subset of types of the sender ("level-1" types), where for each θ, $t_1^1(d, \theta)$ is a type of the sender who: (i) has the bias d; (ii) knows the payoff-state θ; and (iii) believes that the receiver's type is t_2^0 for certain, that is:

$$d(t_1^1(d, \theta)) = d, \ \theta(t_1^1(d, \theta)) = \theta, \text{ and } b_1(T_2^0|t_1^1(d, \theta)) = 1. \tag{13}$$

Let $T_1^1 = \bigcup_{d \in D} T_1^1(d)$. Likewise, let $T_2^1 = \{t_2^1(d)|d \in D\}$ be a set of "level-1" types of the receiver, where for each $d \in D$, $t_2^1(d)$ believes that the sender's

[26] By assumption, t_2^0 believes that the sender's type is certainly $t_1^0(\theta)$ when the payoff-state is θ.

type is in $T_1^1(d)$ (i.e., $b_2(T_1^1(d)|t_2^1(d)) = 1$).[27]

Inductively, given T_2^k for each $k = 1, 2, \ldots$, let T_1^{k+1} be another subset of the sender's types ("level-$(k+1)$" types) as follows. First, for each $d \in D$ and $t_2 \in T_2^k$, let $T_1^{k+1}(d, t_2) = \{t_1^{k+1}(d, \theta, t_2)|\theta \in \Theta\}$ be a subset of types of the sender, where for each θ, $t_1^{k+1}(d, \theta, t_2)$ is a type of the sender who: (i) has the bias d; (ii) knows the payoff-state θ; and (iii) believes that the receiver's type is t_2 for certain, that is:

$$d(t_1^{k+1}(d, \theta, t_2)) = d, \ \theta(t_1^{k+1}(d, \theta, t_2)) = \theta, \ \text{and } b_1(t_2|t_1^{k+1}(d, \theta, t_2)) = 1. \quad (14)$$

Let $T_1^{k+1} = \bigcup_{d \in D, t_2 \in T_2^k} T_1^{k+1}(d, t_2)$.

Similarly, let $T_2^{k+1} = \{t_2^{k+1}(d, t_2)|d \in D, t_2 \in T_2^k\}$ be another subset of the receiver's types ("level-$(k+1)$" types), where for each $d \in D$ and $t_2 \in T_2^k$, $t_2^{k+1}(d, t_2)$ believes that the sender's type is in $T_1^{k+1}(d, t_2)$ (i.e., $b_2(T_1^{k+1}(d, t_2)|t_2^{k+1}(d, t_2)) = 1$).[28] We complete the description of the type space by defining $T_i = \bigcup_{k=0}^{\infty} T_i^k$ for each i.

Now we consider any PBE σ^* given this type space with Property FR0, and then show that $A(\theta|\sigma^*) = A$ for any $\theta \in \Theta$. Let $A^k(\theta|\sigma^*)$ denote the set of actions that the receiver in T_2^k can play in equilibrium σ^* if she receives

[27] By assumption, $t_2^1(d)$ believes that the sender's type is $t_1^1(d, \theta)$ for certain when the payoff-state is θ.

[28] By assumption, $t_2^{k+1}(d, t_2)$ believes that the sender's type is $t_1^{k+1}(d, \theta, t_2)$ for certain when the payoff-state is θ.

message $m = \theta$, which is defined by:

$$A^k(\theta|\sigma^*) = \left\{a \in A \middle| \exists t_2 \in T_2^k \text{ s.t. } a = \sigma_2^*\big(t_2|\theta\big)\right\}. \tag{15}$$

Notice that $\bigcup_k A^k(\theta|\sigma^*) \subseteq A(\theta|\sigma^*)$ for every θ.

Without loss of generality, we assume that $\sigma_1^*(t_1^0(\theta)) = \theta$ for each $\theta \in \Theta$, and $\sigma_2^*(t_2^0, m) = m$ for each $m \in M$. Obviously, $A^0(\theta|\sigma^*) = \{\theta\}$ for each θ.

For the sender with $t_1^1(d, \theta) \in T_1^1$, because he believes the receiver's type is t_2^0, his unique best response is to send $\sigma_1^*(t_1^1(d, \theta)) = \theta + d$. Given this, consider the receiver with type $t_2^1(d) \in T_2^1$ where $d \in D$. Because the receiver believes the sender's type is one of those in $T_1^1(d)$, her unique best response is $\sigma_2^*(t_2^1, m) = m - d$. Hence, $A^1(\theta|\sigma^*) = [\theta - \varepsilon, \theta + \varepsilon]$ for any θ.

By induction, suppose that, for each $k = 1, 2, \ldots$, and for each $\delta_k \in [-k\varepsilon, k\varepsilon]$, there exists $t_2 \in T_2^k$ such that $\sigma_2^*(t_2, m) = m - \delta_k$ for each $m \in M$. Consider the sender with type $t_1^{k+1}(d, \theta, t_2) \in T_1^{k+1}(d, t_2)$ for some $d \in D$ and $\theta \in \Theta$. Because he believes that the receiver's type is t_2, his unique best response is to send $\sigma_1^*(t_1^{k+1}(d, \theta, t_2)) = \theta + d + \delta_k$. Given this, consider the receiver with type $t_2^{k+1}(d', t_2) \in T_2^{k+1}$ where $d' \in D$. Because she believes the sender's type is one of those in $T_1^{k+1}(d', t_2)$, her unique best response is $\sigma_2^*(t_2^{k+1}(d', t_2), m) = m - d' - \delta_k \in [m - (k+1)\varepsilon, m + (k+1)\varepsilon]$. Hence, $A^{k+1}(\theta|\sigma^*) = [\theta - (k+1)\varepsilon, \theta + (k+1)\varepsilon]$ for any θ.

Therefore, $\bigcup_k A^k(\theta|\sigma^*) = \mathbb{R}(= A) \subseteq A(\theta|\sigma^*)$ for every θ. That is, $A(\theta|\sigma^*) = A$ for every θ. $\square$

A.2 Proof of Proposition 1

Fix $\varepsilon > 0$, arbitrarily, and we first construct type space $\tilde{\mathcal{T}} = (\tilde{T}_1, \tilde{T}_2, \tilde{b}_1, \tilde{b}_2)$ as follows. Let $\tilde{T}_1^0 = \Theta \times \{0\}$, $\tilde{T}_1^1 = \Theta \times D \backslash \{0\}$, $\tilde{T}_2^0 = \{0\}$, and $\tilde{T}_2^1 = D \backslash \{0\}$. For any i, define $\tilde{T}_i = \tilde{T}_i^0 \cup \tilde{T}_i^1$.

Each type's belief is described as follows. For the sender of $t_1 = (\theta, 0)$, he (i) has bias $d = 0$; (ii) knows payoff-state θ; and (iii) certainly believes that the receiver's type is $t_2 = 0$. That is:

$$d(\theta, 0) = d, \ \theta(\theta, 0) = \theta, \ \text{and } \tilde{b}_1(\tilde{T}_2^0 | (\theta, 0)) = 1. \tag{16}$$

For the receiver of type $t_2 = 0$, she believes that (i) the payoff-state follows improper uniform distribution $\mu \in \Delta(\Theta)$, and independent of that, (ii) the bias is 0 with probability $1/(1+\varepsilon^2)$, and it is uniformly distributed over $D \backslash \{0\}$ with the remaining probability. That is:

$$\begin{aligned} &\tilde{b}_2^{\Theta}(\cdot | t_2 = 0) = \mu(\cdot | t_2 = 0), \\ &\tilde{b}_2^{D}(d | t_2 = 0) = \begin{cases} (\varepsilon^2 + d\varepsilon)/[2(1+\varepsilon^2)] & \text{if } d \in [-\varepsilon, 0), \\ (2+\varepsilon^2)/[2(1+\varepsilon^2)] & \text{if } d = 0, \\ (2+\varepsilon^2 + d\varepsilon)/[2(1+\varepsilon^2)] & \text{if } d \in (0, \varepsilon], \end{cases} \end{aligned} \tag{17}$$

where $\tilde{b}_2^D(t_2 = 0)$ represents a cumulative distribution function over D.

For the sender of type $t_1 = (\theta, d)$ with $d \in D \backslash \{0\}$, he (i) has bias d; (ii) knows payoff-state θ; and (iii) believes that the receiver's type is 0 and d

with probabilities $\varepsilon^3/(1-\varepsilon)$ and $(1-\varepsilon-\varepsilon^3)/(1-\varepsilon)$, respectively. That is:

$$
\begin{aligned}
&d(\theta, d) = d, \;\; \theta(\theta, d) = \theta, \\
&\tilde{b}_1(t_2) = \begin{cases} \varepsilon^3/(1-\varepsilon) & \text{if } t_2 = 0, \\ (1-\varepsilon-\varepsilon^3)/(1-\varepsilon) & \text{if } t_2 = d, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \tag{18}
$$

For the receiver of type $t_2 = d$ with $d \in D \backslash \{0\}$, she believes that (i) the payoff-state follows improper uniform distribution μ, and (ii) the sender's bias is d for certain. That is,

$$
\begin{aligned}
&\tilde{b}_2^{\Theta}(\cdot | t_2 = d) = \mu(\cdot), \\
&\tilde{b}_2^{D}(d' | t_2 = d) = \begin{cases} 0 & \text{if } d' \in [-\varepsilon, d), \\ 1 & \text{otherwise.} \end{cases} .
\end{aligned} \tag{19}
$$

Notice that this type space admits a common prior. In particular, one can show that the distribution over the type space conditional one θ is summarized as in Table 1.

Next, we show that the following strategies constitute a PBE satisfying Property FR-$(1/(1+\varepsilon^2))$.

$$
\sigma_1^*(\theta, d) = \theta + \frac{1-\varepsilon}{\varepsilon^3} d, \tag{20}
$$

$$
\sigma_2^*(t_2, m) = m - \frac{1-\varepsilon}{\varepsilon^3} t_2. \tag{21}
$$

Notice that the sender with type $t_1 = (\theta, 0)$ and the receiver with type $t_2 = d \in D\backslash\{0\}$ essentially "knows" the opponent type. Hence, given the opponent strategy, it is obvious that their behaviors are the best responses. For the sender of type $t_1 = (\theta, d)$ with $d \in D\backslash\{0\}$, he faces either type $t_2 = 0$ (who plays $a = m$) with probability $\varepsilon^3/(1-\varepsilon)$ or $t_2 = d$ (who plays $a = m - (1-\varepsilon)d/\varepsilon^3$) with the remaining probability. Hence, his expected utility from message m is

$$-\frac{\varepsilon^3}{1-\varepsilon}(m-\theta-d)^2 - \frac{1-\varepsilon-\varepsilon^3}{1-\varepsilon}\left(m - \frac{1-\varepsilon}{\varepsilon^3}d - \theta - d\right)^2. \tag{22}$$

By the first-order condition, his best response is $m = \theta + (1-\varepsilon)d/\varepsilon^3$.

Finally, consider the best response of the receiver with type $t_2 = 0$. Because we assume improper uniform distribution μ as a prior over the payoff-state, the receiver's belief updating about θ is straightforward. Hence, given message m, the receiver's ex ante expected utility from action a is

$$-\frac{1}{1+\varepsilon^2}(a-m)^2 - \int_{-\varepsilon}^{\varepsilon}\left(a - m + \frac{1-\varepsilon}{\varepsilon^3}\tilde{d}\right)^2 \frac{\varepsilon}{2(1+\varepsilon^2)}d\tilde{d}. \tag{23}$$

By the first-order condition, her best response is $a = m$. Thus, these strategies constitute a PBE. Furthermore, for the sender of type $t_1 = (\theta, 0)$ and the receiver of type $t_2 = 0$, zero bias is commonly $1/(1+\varepsilon^2)$ believed. Hence, PBE σ^* satisfies Property FR-$(1/(1+\varepsilon^2))$.

In this equilibrium, conditional on θ, the sender of bias d plays with the receiver of type $t_2 = 0$ with density $\varepsilon^2/2$ for any $d \in D\backslash\{0\}$. In this scenario, action $a = \theta + (1+\varepsilon)d/\varepsilon^3$ is played. Because the receiver correctly infer the true payoff-state with the remaining probability, the equilibrium action is $a = \theta$. As a result, $A(\theta|\sigma^*) = [\theta - \delta(\varepsilon), \theta + \delta(\varepsilon)]$. Therefore, conditional on θ, the receiver's ex ante expected utility is

$$-(1-\varepsilon^3)(\theta-\theta)^2 - \int_{-\varepsilon}^{\varepsilon}\left(\theta + \frac{1-\varepsilon}{\varepsilon^3}\tilde{d} - \theta\right)^2 \frac{\varepsilon^2}{2} d\tilde{d} = -\frac{(1-\varepsilon)^2}{3\varepsilon}. \tag{24}$$

Because this value is independent of payoff-state θ, the receiver's ex ante expected utility is given by $V(\sigma^*) = -(1-\varepsilon)^2/(3\varepsilon)$. $\square$

A.3 Proof of Theorem 2

Fix arbitrary Harsanyi type space T and equilibrium σ that does not exhibit either divergent interpretation or divergent prediction. Then, for any on-path message $m \in \tilde{M}$, the sender of any type believes the same reaction of the receiver, i.e., $b_1^A(t_1, m)$ does not vary with $t_1 \in T_1$.[29] Therefore, we denote it by $b_1^A(m)$.

By Condition (ii), for any $m, m' \in \tilde{M}$, there exists t_1 satisfying (9), which, together with the no-divergent-prediction condition, implies:

$$\int_A a \; db_1^A(a|m) \neq \int_A a \; db_1^A(a|m'), \tag{25}$$

[29] If an equilibrium does not exhibit divergent interpretation, then every type of the receiver responds in the same way. Hence, the difference in the sender's belief does not matter.

that is, the mean actions are different.[30] Without loss of generality, in what follows, we identify each message $m \in \tilde{M}$ with the mean action induced by it, that is, $m = \int_A a \ db_1^A(m)$. Furthermore, the no-divergent-prediction condition implies that the sender's types t_1, t_1' with $x = \theta(t_1) + d(t_1) = \theta(t_1') + d(t_1')$ have the same preference ordering over $\tilde{M}$. Thus, let $\tilde{M}(x)$ denote the set of best-response messages for the sender with $x = \theta(t_1)+d(t_1)$. We call this x a "virtual type" of the sender with t_1. Notice that $\tilde{M}(x)$ is non-empty for any x, and $\inf \tilde{M}(x)$ and $\sup \tilde{M}(x)$ are increasing in x.

For each $m \in \tilde{M}$, let $X(m) = \{x \in \mathbb{R} | m \in \tilde{M}(x)\}$ denote the set of virtual types who prefer message m. By the single-crossing condition, $X(m)$ is a convex set. Also, for each $m < m'$, we have $\sup X(m) \leq \inf X(m')$; otherwise, there are $x \neq x'$ for whom m, m' are both optimal, but then this contradicts that $m \neq m'$. Note that this implies $\inf \tilde{M}(x), \sup \tilde{M}(x) \in \mathbb{R}$ (i.e., $\tilde{M}(x)$ is bounded) for any $x \in \mathbb{R}$; for example, if $\sup \tilde{M}(x) = \infty$ for some x, then for any $x' > x$ and $m' \in \tilde{M}(x')$, there exists $m \in \tilde{M}(x)$ such that $m > m'$, but then the single-crossing preference implies that x' strictly prefers m to m', contradicting $m' \in \tilde{M}(x')$.

For each $m \in \tilde{M}$, let $\hat{A}(m) = \{\sigma_2(t_2, m) | t_2 \in T_2\}$ be the set of possible equilibrium actions given message m. Then, we show that set $\hat{A}(m)$ is bounded for any $m \in \tilde{M}$ as in the following lemma.

[30] More precisely, the no-divergent-prediction condition implies either different mean actions (as above) or (the same mean but) different variances. However, the second case means that either m or m' is never played, which contradicts to $m, m' \in \tilde{M}$.

Lemma 1. For each $m \in \tilde{M}$, define:

$$\underline{\zeta}(m) = \begin{cases} \inf X(m) - \varepsilon & \text{if } \inf X(m) > -\infty, \\ E[\theta|\theta \leq \sup X(m) - \varepsilon] & \text{otherwise,} \end{cases} \tag{26}$$

$$\overline{\zeta}(m) = \begin{cases} \sup X(m) + \varepsilon & \text{if } \sup X(m) < \infty, \\ E[\theta|\theta \geq \inf X(m) + \varepsilon] & \text{otherwise.} \end{cases} \tag{27}$$

Then, $\inf \hat{A}(m) \geq \underline{\zeta}(m)$ and $\sup \hat{A}(m) \leq \overline{\zeta}(m)$ for any $m \in \tilde{M}$.

Proof. (of the lemma) Fix $m \in \tilde{M}$. Suppose that $\inf X(m) > -\infty$. Because type t_1 of the sender does not send m if $\theta(t_1) + d(t_1) < \inf X(m)$, this message m is not sent given any $\theta < \inf X(m) - \varepsilon$. Therefore, $\inf \hat{A}(m) \geq \inf X(m) - \varepsilon$. When $\inf X(m) = -\infty$, there is a tighter (and finite) lower bound on $\inf \hat{A}(m)$. Because type t_1 of the sender sends m if $\theta(t_1) + d(t_1) < \sup X(m)$, this message m is sent given any $\theta < \sup X(m) - \varepsilon$. Therefore, given any type t_2 of the receiver, her action is (weakly) higher than $E[\theta|\theta \leq \sup X(m) - \varepsilon]$.[31] The upper bounds on $\sup \hat{A}(m)$ are analogously obtained as in the statement. □

Therefore, by Lemma 1 for each θ:

$$\begin{aligned} A(\theta|\sigma) &\subseteq \{a \in A | \exists x \in [\theta - \varepsilon, \theta + \varepsilon], \exists m \in \tilde{M}(x),\ a \in \hat{A}(m)\} \\ &\subseteq [\underline{\zeta}(\inf \tilde{M}(\theta - \varepsilon)), \overline{\zeta}(\sup \tilde{M}(\theta + \varepsilon))], \end{aligned} \tag{28}$$

[31] If $X(m) = \mathbb{R}$, then σ is a babbling equilibrium, which implies that $\inf \hat{A}(m) = E[\theta] = \int_\Theta \theta d\mu$.

which is bounded.[32] $\square$

A.4 Proof of Theorem 3

Fix arbitrary $\delta > 0$ and $\varepsilon > 0$. Also, fix arbitrary Harsanyi type space $\mathcal{T} \in \mathbb{T}^{\varepsilon}$ and equilibrium σ that does not exhibit either divergent interpretation or divergent prediction.

By the same argument used in Theorem 2, the sender's equilibrium message depends only on his virtual type. In the following, because the case with m such that $X(m) = \mathbb{R}$ (i.e., a babbling equilibrium) is trivial, we only consider the other case where $X(m) \subsetneq \mathbb{R}$. By the Lipschitz-continuity of density f, there exists $\lambda > 0$ such that $|f(\theta) - f(\theta')| \leq \lambda|\theta - \theta'|$ for any θ, θ'. Let $c_1, c_2 \in \mathbb{R}$ be such that: (i) $\Pr(\theta < c_1), \Pr(\theta > c_2) \in (0, \delta/2)$; (ii) for all $m \in \tilde{M}$ with $\inf X(m) > -\infty$, $c_1 < \inf X(m)$; (iii) for all $m \in \tilde{M}$ with $\sup X(m) < \infty$, $c_2 > \sup X(m)$; (iv) f is convex and increasing on $(-\infty, c_1)$, and is convex and decreasing on (c_2, ∞); (v) $c_1 < E[\theta|\theta < 0]$ and $c_2 > E[\theta|\theta > 0]$; and finally, (vi) for $m \in \tilde{M}$ with $\inf X(m) = -\infty$, $\sup X(m) + \varepsilon < c_2$, and for $m \in \tilde{M}$ with $\sup X(m) = +\infty$, $\inf X(m) - \varepsilon > c_1$.[33]

Let $\overline{f} = \max_{x \in [c_1, c_2]} f(x)$ and $\underline{f} = \min_{x \in [c_1, c_2]} f(x)$.

By construction, it is obvious that (10) holds. Hence, hereafter, we show

[32] Notice that $\underline{\zeta}(m)$ and $\overline{\zeta}(m)$ are increasing in m.

[33] These conditions hold if $|c_1|$ and $|c_2|$ are sufficiently large.

the second part of the statement. Define $M^* = \{m \in \tilde{M} | \sup X(m) \geq c_1 - \varepsilon \text{ and } \inf X(m) \leq c_2 + \varepsilon\}$. Note that any message not in M^* can be sent only if $\theta \notin [c_1, c_2]$. Let $\underline{a}(m) = \inf \hat{A}(m)$ and $\overline{a}(m) = \sup \hat{A}(m)$. First, we show the following lemmas.

Lemma 2. There exists $\alpha \in \mathbb{R}_{++}$ such that, for each $m \in M^*$:

$$\overline{a}(m) - \underline{a}(m) \leq \alpha\varepsilon. \tag{29}$$

Proof. Fix $m \in M^*$, arbitrarily, and let $\overline{x} = \sup X(m)$ and $\underline{x} = \inf X(m)$.

The following three cases are considered: (i) $\underline{x} \geq c_1$ and $\overline{x} \leq c_2$; (ii) $\underline{x} = -\infty$ and $\overline{x} \leq c_2 - \varepsilon$; and (iii) $\underline{x} \geq c_1 + \varepsilon$ and $\overline{x} = \infty$.

First, suppose that $\underline{x} \geq c_1$ and $\overline{x} \leq c_2$. That is, $X(m)$ is bounded, and also $X(m) \subseteq [c_1, c_2]$.

Case 1: $\overline{x} - \underline{x} < 3\varepsilon$.

In this case, by Lemma 1:

$$\overline{a}(m) - \underline{a}(m) < \overline{\zeta}(m) - \underline{\zeta}(m) = \overline{x} + \varepsilon - (\underline{x} - \varepsilon) < 5\varepsilon. \tag{30}$$

Case 2: $\overline{x} - \underline{x} \geq 3\varepsilon$.

Let $\eta = E[\theta | \theta \in [\underline{x} + \varepsilon, \overline{x} - \varepsilon]] \in [c_1, c_2]$. Note that $\mu([\underline{x} + \varepsilon, \overline{x} - \varepsilon]) \geq$

$(\overline{x} - \underline{x} - 2\varepsilon)\underline{f}$. In this case, the following holds:

$$\overline{a}(m) \leq \frac{2(\overline{x} + \varepsilon)(f_2 + 2\varepsilon\lambda)\varepsilon + \eta\mu([\underline{x} + \varepsilon, \overline{x} - \varepsilon])}{2(f_2 + 2\varepsilon\lambda)\varepsilon + \mu([\underline{x} + \varepsilon, \overline{x} - \varepsilon])}, \tag{31}$$

$$\underline{a}(m) \geq \frac{2(\underline{x} - \varepsilon)(f_1 + 2\varepsilon\lambda)\varepsilon + \eta\mu([\underline{x} + \varepsilon, \overline{x} - \varepsilon])}{2(f_1 + 2\varepsilon\lambda)\varepsilon + \mu([\underline{x} + \varepsilon, \overline{x} - \varepsilon])}, \tag{32}$$

where $f_1 = f(\underline{x} + \varepsilon)$ and $f_2 = f(\overline{x} - \varepsilon)$.

Now, we consider the following two sub-cases. First, suppose that $f_2 > f_1$. In this sub-case:

$$\begin{aligned}
\overline{a}(m) - \underline{a}(m) &\leq \frac{2\varepsilon[\overline{x}f_2 - \underline{x}f_1 + \varepsilon(f_1 + f_2) + 2\varepsilon\lambda(\overline{x} - \underline{x} + 2\varepsilon)]}{(f_1 + 2\varepsilon\lambda)\varepsilon + (\overline{x} - \underline{x} - 2\varepsilon)\underline{f}} \\
&\leq \frac{2\varepsilon(\overline{x} - \underline{x})}{(\overline{x} - \underline{x} - 2\varepsilon)\underline{f}}\left[f_1 + \frac{\overline{x}}{\overline{x} - \underline{x}}(f_2 - f_1) + 2\varepsilon\lambda\right] + \frac{4\varepsilon\overline{f} + 8\varepsilon^2\lambda}{(\overline{x} - \underline{x} - 2\varepsilon)\underline{f}}\varepsilon \\
&\leq \frac{2(\overline{x} - \underline{x})(\overline{f} + \lambda\overline{x} + 2\varepsilon\lambda) + 4\varepsilon\overline{f} + 8\varepsilon^2\lambda}{(\overline{x} - \underline{x} - 2\varepsilon)\underline{f}}\varepsilon \qquad (33) \\
&\leq \frac{2(\overline{x} - \underline{x})}{\overline{x} - \underline{x} - 2\varepsilon}\frac{\overline{f} + \lambda\overline{x} + 2\varepsilon\lambda}{\underline{f}}\varepsilon + \frac{4\overline{f} + 8\varepsilon\lambda}{\underline{f}}\varepsilon \\
&\leq \left[\frac{6(\overline{f} + \lambda\overline{x} + 2\varepsilon\lambda) + 4\overline{f} + 8\varepsilon\lambda}{\underline{f}}\right]\varepsilon,
\end{aligned}$$

where the second inequality is by $(f(\underline{x}+\varepsilon)+2\varepsilon\lambda)\varepsilon > 0$ and $\overline{f} \geq f_2$; the third inequality is by the Lipschitz-continuity; the fourth inequality is by $\overline{x}-\underline{x}-2\varepsilon \geq \varepsilon$; and the last inequality is by $(\overline{x}-\underline{x})/(\overline{x}-\underline{x}-2\varepsilon) \leq 3$.

Next, suppose that $f_2 \leq f_1$. Let $\phi_1 = \mu([\underline{x}+\varepsilon, \underline{x}-\varepsilon])/2(f_1+2\varepsilon\lambda)\varepsilon$ and

$\phi_2 = \mu([\underline{x}+\varepsilon, \underline{x}-\varepsilon])/2(f_2+2\varepsilon\lambda)\varepsilon$. Then:

$$\begin{aligned}
\overline{a}(m) - \underline{a}(m) &\leq \frac{\overline{x}+\varepsilon}{1+\phi_2} - \frac{\underline{x}-\varepsilon}{1+\phi_1} + \frac{2(f_1-f_2)}{(\overline{x}-\underline{x}-2\varepsilon)\underline{f}}\eta\varepsilon \\
&\leq \frac{\overline{x}-\underline{x}+2\varepsilon}{\phi_1} + \frac{2(\overline{x}-\underline{x})\lambda}{(\overline{x}-\underline{x}-2\varepsilon)\underline{f}}\eta\varepsilon \\
&\leq \frac{2(\overline{x}-\underline{x}+2\varepsilon)}{(\overline{x}-\underline{x}-2\varepsilon)\underline{f}}(\overline{f}+2\varepsilon\lambda)\varepsilon + \frac{6\eta\lambda}{\underline{f}}\varepsilon \qquad (34) \\
&\leq \left[\frac{10(\overline{f}+2\varepsilon\lambda)+6\eta\lambda}{\underline{f}}\right]\varepsilon,
\end{aligned}$$

where the second inequality is by $1+\phi_2 \geq 1+\phi_1 \geq \phi_1$ and the Lipschitz-continuity; the third inequality is by $\mu([\underline{x}+\varepsilon, \overline{x}-\varepsilon]) \geq (\overline{x}-\underline{x}-2\varepsilon)\underline{f}$, $f_1 \leq \overline{f}$, and $(\overline{x}-\underline{x})/(\overline{x}-\underline{x}-2\varepsilon) \leq 3$; and the last inequality is by $(\overline{x}-\underline{x}+2\varepsilon)/(\overline{x}-\underline{x}-2\varepsilon) \leq 5$.

Next suppose that $\underline{x} = -\infty$ and $(c_1+\varepsilon \leq)\overline{x} \leq c_2-\varepsilon$. In this case, $\underline{a}(m) = \eta = E[\theta|\theta < \overline{x}-\varepsilon] > E[\theta|\theta < c_1]$, and:

$$\overline{a}(m) \leq \frac{2(\overline{x}+\varepsilon)(f_2+2\varepsilon\lambda)\varepsilon + \eta\mu((-\infty, \overline{x}-\varepsilon])}{2(f_2+2\varepsilon\lambda)\varepsilon + \mu((-\infty, \overline{x}-\varepsilon])}. \qquad (35)$$

Thus:

$$\begin{aligned}
\overline{a}(m) - \underline{a}(m) &\leq \frac{2(\overline{x}+\varepsilon-\eta)(f_2+2\varepsilon\lambda)\varepsilon}{2(f_2+2\varepsilon\lambda)\varepsilon + \mu((-\infty, \overline{x}-\varepsilon])} \\
&\leq \left[\frac{2(c_2 - E[\theta|\theta < c_1])(\overline{f}+2\varepsilon\lambda)}{\mu((-\infty, c_1])}\right]\varepsilon. \qquad (36)
\end{aligned}$$

Finally, Case (iii) is similar to Case (ii), and hence omitted. Therefore, it suffices to set α as the supremum of the threshold coefficients of ε for Cases (i)-(iii). $\square$

Lemma 3. For each $\theta \in [c_1, c_2]$ and each $m, m' \in \bigcup_{x \in [\theta-\varepsilon, \theta+\varepsilon]} \tilde{M}(x)$:

$$\overline{a}(m') - \underline{a}(m) \leq \alpha\varepsilon + m' - m. \tag{37}$$

Proof. For each such m, m', we have:

$$\begin{aligned}\overline{a}(m') - \underline{a}(m) &= \overline{a}(m') - \underline{a}(m') + \overline{a}(m) - \underline{a}(m) + \underline{a}(m') - \overline{a}(m) \\ &\leq 2\alpha\varepsilon + m' - m,\end{aligned} \tag{38}$$

where the inequality is because of Lemma 2 and because $\underline{a}(m') \leq m'$ and $\overline{a}(m) \geq m$. $\square$

Lemma 4. There exist $\varepsilon^* > 0$ and $\beta > 1$ such that, for each $\varepsilon \in (0, \varepsilon^*)$, $\theta \in [c_1, c_2]$ and $m, m' \in \bigcup_{x \in [\theta-\varepsilon, \theta+\varepsilon]} \tilde{M}(x)$, we have:

$$(m' - m)\mu\left(\left[\frac{m+m'}{2} - \varepsilon, \frac{m+m'}{2} + \varepsilon\right]\right) \leq \beta \int_m^{m'} \mu([y - \varepsilon, y + \varepsilon])dy. \tag{39}$$

Proof. First, note that m and m' in the statement are necessarily in M^*; otherwise, it cannot be in $\tilde{M}(x)$ for any $x \in [\theta - \varepsilon, \theta + \varepsilon]$ and $\theta \in [c_1, c_2]$. Therefore, we never have $m, m' < c_1 - \varepsilon$ simultaneously; otherwise, $\sup X(m) < c_1 - \varepsilon$, and hence $m \notin M^*$. Similarly, we never have $m, m' > c_2 + \varepsilon$ simultaneously.

Furthermore, by taking ε^* sufficiently small, we do not simultaneously have $m < c_1 - \varepsilon$ and $m' > c_2 + \varepsilon$, for $\varepsilon \in (0, \varepsilon^*)$. To see this, take $\varepsilon^* < \min\{(c_2 - c_1)/2, E[\theta|\theta < 0] - c_1, c_2 - E[\theta|\theta > 0]\}$. Then, $m < c_1 - \varepsilon$ implies $m < E[\theta|\theta < 0]$, and $m' > c_2 + \varepsilon$ implies $m' > E[\theta|\theta > 0]$. Let $\psi, \psi' > 0$ be such that $m = E[\theta|\theta < -\psi]$ and $m' = E[\theta|\theta > \psi']$, and redefine $\varepsilon^* < \min\{(c_2 - c_1)/2, E[\theta|\theta < 0] - c_1, c_2 - E[\theta|\theta > 0], \psi, \psi'\}$.[34] Then, for $\varepsilon \in (0, \varepsilon^*)$, there do not exist θ and $m, m' \in \bigcup_{x \in [\theta - \varepsilon, \theta + \varepsilon]} \tilde{M}(x)$. Now, fix $\varepsilon \in (0, \varepsilon^*)$, and let $f_- = \min_{x \in [c_1 - 2\varepsilon, c_2 + 2\varepsilon]} f(x)$ and $f_+ = \max_{x \in [c_1 - 2\varepsilon, c_2 + 2\varepsilon]} f(x)$. Hereafter, we focus on the following three cases: (i) $m, m' \in [c_1 - \varepsilon, c_2 + \varepsilon]$; (ii) $m < c_1 - \varepsilon$ and $m' \in [c_1 - \varepsilon, c_2 + \varepsilon]$; and (iii) $m \in [c_1 - \varepsilon, c_2 + \varepsilon]$ and $m' > c_2 + \varepsilon$.

First, suppose that $m, m' \in [c_1 - \varepsilon, c_2 + \varepsilon]$. In this case:

$$(m' - m)\mu\left(\left[\frac{m + m'}{2} - \varepsilon, \frac{m + m'}{2} + \varepsilon\right]\right) \leq 2\varepsilon(m' - m)f_+, \tag{40}$$

and:

$$\int_m^{m'} \mu([y - \varepsilon, y + \varepsilon])dy \geq \int_m^{m'} 2\varepsilon f_- dy = 2\varepsilon(m' - m)f_-. \tag{41}$$

Thus, it is sufficient to take $\beta > f_+/f_-$.

Next, suppose that $m < c_1 - \varepsilon$ and $m' \in [c_1 - \varepsilon, c_2 + \varepsilon]$. For $z \in (-\varepsilon, \varepsilon)$,

[34] Because of the continuity of μ, there exist such ψ and ψ'.

define function $\xi : (m+z, m'+z) \to \mathbb{R}$ by:

$$\xi(y) = \begin{cases} f(y) & \text{if } y \in (m+z, c_1-\varepsilon], \\ f(c_1-\varepsilon) + (y-c_1+\varepsilon)\lambda & \text{if } y \in (c_1-\varepsilon, m'+z). \end{cases} \tag{42}$$

Notice that ξ is convex and weakly above $f(y)$ for $y \in (m+z, m'+z)$. Hence, we have:

$$\begin{aligned} \int_{m+z}^{m'+z} \xi(y)dy &\geq (m'-m)\xi\left(\frac{m+m'}{2}+z\right) \\ &\geq (m'-m)f\left(\frac{m+m'}{2}+z\right). \end{aligned} \tag{43}$$

Let β be greater than or equal to $(f(c_1-\varepsilon)+(c_2-c_1+2\varepsilon)\lambda)/f_-$. Then:

$$\begin{aligned} \beta \int_{m}^{m'} \mu([y-\varepsilon, y+\varepsilon])dy &= \beta \int_{-\varepsilon}^{\varepsilon}\int_{m+z}^{m'+z} f(y)dydz \\ &\geq \beta \int_{-\varepsilon}^{\varepsilon}\int_{m+z}^{m'+z} f_- dydz \\ &\geq \int_{-\varepsilon}^{\varepsilon}\int_{m+z}^{m'+z} (f(c_1-\varepsilon)+(c_2-c_1+2\varepsilon)\lambda)dydz \\ &\geq \int_{-\varepsilon}^{\varepsilon}\int_{m+z}^{m'+z} \xi(y)dydz \\ &\geq \int_{-\varepsilon}^{\varepsilon} (m'-m)f\left(\frac{m+m'}{2}+z\right) \\ &= (m'-m)\mu\left(\left[\frac{m+m'}{2}-\varepsilon, \frac{m+m'}{2}+\varepsilon\right]\right). \end{aligned} \tag{44}$$

Finally, Case (iii) can be shown by the similar argument used in Case (ii), and hence it is omitted. Therefore, we complete the proof by taking β

greater than the maximum of the conditions above. $\square$

Recall that, for each x, $\tilde{M}(x)$ contains either one or two elements, and furthermore, $x < x'$ implies $m \leq m'$ for any $m \in \tilde{M}(x)$ and $m' \in \tilde{M}(x')$; otherwise, it violates single-crossing. This implies that there are only countably many x such that $\tilde{M}(x)$ contains two elements.[35]

In the following, for each x, fix $m^*(x) \in \tilde{M}(x)$ arbitrarily. Note that this $m^*(x)$ is non-decreasing, has at most countably many discontinuities, as discussed above. If m^* is discontinuous at x, then the sender of type t_1 with $\theta(t_1) + d(t_1) = x$ is indifferent between sending m and m' that are included in $\tilde{M}(x)$. Define $m = m^*(x^-)$ and $m' = m^*(x^+)$.

Lemma 5.

(i) If $m^*(x)$ is discontinuous at x, then $x \in [(m^*(x^+) + m^*(x^-))/2 - \alpha\varepsilon, (m^*(x^+) + m^*(x^-))/2 + \alpha\varepsilon]$; and

(ii) if $m^*(x)$ is continuous at x, then either m^* is locally constant around x, or $m^*(x) \in [x - \varepsilon, x + \varepsilon]$.

Proof. (i) Suppose that $m^*(x)$ is discontinuous at x. Given that $b_2^A(m) \in \Delta([\underline{a}(m), \overline{a}(m)])$ and $b_2^A(m') \in \Delta([\underline{a}(m'), \overline{a}(m')])$, it is necessary that $x \in [(\underline{a}(m^*(x^+)) + \underline{a}(m^*(x^-)))/2, (\overline{a}(m^*(x^+)) + \overline{a}(m^*(x^-)))/2]$. Finally, because

[35] To see this, let $r(x)$ be a rational number between $(\min \tilde{M}(x), \max \tilde{M}(x))$ if $\tilde{M}(x)$ contains two elements (and hence this open set is non-empty). For any $x < x'$ such that both $\tilde{M}(x)$ and $\tilde{M}(x')$ contain two elements each, because $\max \tilde{M}(x) \leq \min \tilde{M}(x')$, we have $r(x) < r(x')$. This implies that the set of x such that $\tilde{M}(x)$ contains two elements cannot have a cardinality greater than that of the rational numbers.

$m^*(y) \in [\underline{a}(m^*(y)), \overline{a}(m^*(y))]$ for each y, together with Lemma 2, we obtain $x \in [(m^*(x^+) + m^*(x^-))/2 - \alpha\varepsilon, (m^*(x^+) + m^*(x^-))/2 + \alpha\varepsilon]$.

(ii) Suppose that $m^*(x)$ is continuous at x, and that $m^*(x)$ is not locally constant, that is, there exists a sequence $\{x_k\}_{k=1}^{\infty}$ such that, for each k, $|x - x_k| < \frac{1}{k}$ and $m^*(x) \neq m^*(x_k)$. Without loss of generality, we assume that $\{x_k\}_{k=1}^{\infty}$ is an increasing sequence, and hence $m^*(x_k) < m^*(x)$ for all k.[36]

Suppose, in contrast, that $x < m^*(x_k) - \varepsilon (< m - \varepsilon)$ for some k. Because $m^*(x_k) = \int_A a db_1^A(m^*(x_k))$, there exists some virtual type x'_k such that $m^*(x_k) = x'_k$, and then it is obvious that $m^*(x_k) \in \tilde{M}(x'_k)$. By the supposition, we have $x < m^*(x_k) - \varepsilon < x'_k$. Furthermore, by construction, x'_k weakly prefers $m^*(x_k)$ to $m^*(x)$, and x weakly prefers $m^*(x)$ to $m^*(x_k)$ (while one of them must be a strict preference; otherwise $m^*(x_k) = m^*(x)$). However, $m^*(x_k) < m^*(x)$ implies a contradiction to the single-crossing condition. Therefore, we conclude $x \geq m^*(x_k) - \varepsilon \geq m^*(x) - 1/k - \varepsilon$ for all k, and thus, $x \geq m^*(x) - \varepsilon$.

On the other hand, for each k, existence of x'_k as above implies that $x_{k-1} < x'_k < m^*(x_k) + \varepsilon$; if $x_{k-1} \geq x'_k$, then $m^*(x_{k-1}) \notin \tilde{M}(x_{k-1})$ by the single-crossing condition, which is a contradiction. Thus, $x = \lim_k x_k \leq \lim_k m^*(x_{k+1}) + \varepsilon = m^*(x) + \varepsilon$.

Therefore, we conclude $m^*(x) \in [x - \varepsilon, x + \varepsilon]$. □

[36] The other case can be shown by a similar argument.

Let X_0 denote the set of x at which $m^*(x)$ is discontinuous. Then:

$$\begin{aligned}
&\frac{1}{\mu([c_1,c_2])}\int_{c_1}^{c_2} \sup A^*(\theta) - \inf A^*(\theta) d\mu \\
\leq\ & \frac{1}{\mu([c_1,c_2])}\int_{c_1}^{c_2} \overline{a}(m^*(\theta+\varepsilon)) - \underline{a}(m^*(\theta-\varepsilon)) d\mu \qquad (45)\\
\leq\ & \frac{1}{\mu([c_1,c_2])}\left[\alpha\varepsilon + \int_{c_1}^{c_2} m^*(\theta+\varepsilon) - m^*(\theta-\varepsilon) d\mu\right],
\end{aligned}$$

where the first inequality is by definition of $\overline{a}$ and $\underline{a}$; and the second inequality is by Lemma 3. Now, the inside of the bracket can be transformed as follows:

$$\begin{aligned}
&\alpha\varepsilon + \int_{c_1}^{c_2} \overline{a}(m^*(\theta+\varepsilon)) - \underline{a}(m^*(\theta-\varepsilon)) d\mu \\
&= \alpha\varepsilon + \int_{\inf \tilde{M}}^{\sup \tilde{M}} \left[\sum_{x\in X_0} 1\{m \in (m^*(x^-), m^*(x^+))\}\mu([x-\varepsilon, x+\varepsilon])\right] \\
&\quad + 1\{m \notin (m^*(x^-), m^*(x^+)),\ \forall x \in X_0\}\mu([m-2\varepsilon, m+2\varepsilon]) dm \\
&= \alpha\varepsilon + \sum_{x\in X_0} \mu([x-\varepsilon, x+\varepsilon])(m^*(x^+) - m^*(x^-)) \qquad (46)\\
&\quad + \int_{\inf \tilde{M}}^{\sup \tilde{M}} 1\{m \notin (m^*(x^-), m^*(x^+)),\ \forall x \in X_0\}\mu([m-2\varepsilon, m+2\varepsilon]) dm \\
&\leq \alpha\varepsilon + \sum_{x\in X_0} \beta \int_{m^*(x^-)}^{m^*(x^+)} \mu([m-(1+\alpha)\varepsilon, m+(1+\alpha)\varepsilon]) dm \\
&\quad + \int_{\inf \tilde{M}}^{\sup \tilde{M}} 1\{m \notin (m^*(x^-), m^*(x^+)),\ \forall x \in X_0\}\mu([m-2\varepsilon, m+2\varepsilon]) dm \\
&\leq \alpha\varepsilon + \beta \int_{\inf \tilde{M}}^{\sup \tilde{M}} \mu([m-\gamma\varepsilon, m+\gamma\varepsilon]) dm \\
&\leq (\alpha + 2\beta\gamma)\varepsilon,
\end{aligned}$$

where the first equality is by Lemma 5-(ii) and definition of Lebesgue integral (i.e., integrating the probability measure that corresponds to each level of the vertical axis); the first inequality is by Lemmas 4 and 5-(i); the second inequality is by $\gamma = \max\{2, 1+\alpha\}$; and the last inequality is by Fubini's theorem. Therefore:

$$\frac{1}{\mu([c_1, c_2])} \int_{c_1}^{c_2} \sup A(\theta|\sigma) - \inf A(\theta|\sigma) d\mu \leq \left[\frac{\alpha + 2\beta\gamma}{\mu([c_1, c_2])}\right] \varepsilon. \tag{47}$$

Therefore, it is sufficient to redefine ε so that the right hand side of (47) is smaller than δ. $\square$

A.5 Proof of Claim 1

Suppose that there exists common prior $\phi \in \Delta(\Theta \times D)$ over $\Theta \times D$, and fix an equilibrium σ and on-path message $m \in \tilde{M}$, arbitrarily. Because of the common prior assumption, $b_2(\{t_1 \in T_1 \mid \theta(t_1) = \theta \text{ and } d(t_1) = d\}|t_2) = \phi(\theta, d)$ holds for any $t_2 \in T_2$, $\theta \in \Theta$ and $d \in D$. Hence, given message m, for any $t_2 \in T_2$, conditional belief $b_2^{\Theta\times D}(t_2, m)$ is

$$\begin{aligned} b_2^{\Theta\times D}(\theta, d|t_2, m) &= \frac{1\{\sigma_1(t_1) = m\} db_2(t_1|t_2)}{\int_{T_1} 1\{\sigma(t_1') = m\} db_2(t_1'|t_2)}, \\ &= \frac{1\{\sigma_1(\theta, d)\} d\phi(\theta, d)}{\int_{\Theta\times D} 1\{\sigma_1(\theta', d') = m\} d\phi(\theta', d')}. \end{aligned} \tag{48}$$

Thus, it is obvious that $b_2^{\Theta\times D}(t_2, m) = b_2^{\Theta\times D}(t_2', m)$ for any t_2, $t_2' \in T_2$ and $m \in \tilde{M}$, which means that equilibrium σ does not exhibits divergent interpretation.

Because of the strict concavity of utility function v in action a and infinite action set A, no divergent interpretation implies that $\sigma_2(t_2, m) = \sigma_2(t_2', m) = a$ holds for any t_2, $t_2' \in T_2$. Therefore, for measurable $E \subset A$ and $t_1 \in T_1$,

$$\begin{aligned} b_1^A(E|t_1, m) &= \int_{T_2} 1\{\sigma_2(t_2, m) \in E\} db_1(t_2|t_1) \\ &= 1\{a \in E\}. \end{aligned} \tag{49}$$

It is obvious that $b_1^A(t_1, m) = b_1^A(t_1', m)$ for any t_1, $t_1' \in T_1$ and $m \in \tilde{M}$, which means that equilibrium σ does not exhibit divergent prediction. Because equilibrium σ is arbitrarily, the statement holds. $\square$

Appendix B: More general preferences

While we have focused on the quadratic-loss preferences in the main body of the paper, our first result, that is, maximal miscommunication, holds with more general preferences. Let $u : A \times \Theta \times D \to \mathbb{R}$ and $v : A \times \Theta \to \mathbb{R}$ be the sender and the receiver's utility functions, where we continue to assume $A = \Theta = \mathbb{R}$ and $D = [-\varepsilon, \varepsilon]$. We impose the following assumptions on u and v.

Assumption 2. The players' utility functions u and v satisfy that: (i) $v(a, \theta) = u(a, \theta, 0)$ for any $a \in A$ and $\theta \in \Theta$; (ii) u is twice continuously differentiable in each argument; (iii) $u_{11} < 0 < u_{12}$ and $u_{13} > 0$ denoting partial derivatives by subscripts; and (iv) there exists $a^1(\theta, d) = \arg\max_{a \in A} u(a, \theta, d)$ and $a^2(\theta) = \arg\max_{a \in A} v(a, \theta)$ for any $\theta \in \Theta$ and $d \in D_\varepsilon$.

Assumption 2 implies that: (i) $a^2(\theta) = a^1(\theta, 0)$ for any $\theta \in \Theta$; (ii) $a^1(\theta, d)$ is unique for any $\theta \in \Theta$ and $d \in D$; (iii) a^1 is differentiable in each argument; and (iv) $a^1_1 > 0$ and $a^1_2 > 0$ denoting partial derivatives by subscripts. In addition to those properties, we impose the following assumptions on the ideal-action mapping a^1. These two assumptions assure that our argument holds beyond the quadratic-loss environment.

Assumption 3. $a^1(\cdot, d) : \Theta \to A$ is bijective for any $d \in D$.[37]

As in Section 3, we consider any PBE such that full revelation occurs if $d = 0$ is common knowledge. As in the quadratic-loss case, we say that such a PBE satisfies *Property FR0*. The only difference from the quadratic-loss case is that, given that each θ is truthfully revealed, the receiver plays $a^2(\theta)$ (instead of θ).

Theorem 4. Suppose that Assumptions 2 and 3 hold. For each $\varepsilon > 0$, there exists $\mathcal{T} \in \mathbb{T}^\varepsilon$ such that: (i) a PBE with Property FR0 exists; and (ii) for any PBE σ^* with Property FR0, we have $A(\theta|\sigma^*) = A$ for any $\theta \in \Theta$.

There are three remarks about the assumptions. First, Assumption 2 and its implications are standard in the literature, for example, as in Crawford

[37]It is equivalent to assume that function $a^1(\cdot, d)$ is surjective because Assumption 2-(ii) already guarantees that $a^1(\cdot, d)$ is injective.

and Sobel (1982). Second, we need Assumption 3 in order to apply the argument used in the previous section to this general environment, although it is not standard in the literature. Assumption 3 means that any available action can be supported as the sender's ideal action in some state, whatever the bias parameter $d(\in D = [-\varepsilon, \varepsilon])$ is. It implies a one-to-one relationship between observed messages and states given any fixed bias parameter. Without Assumption 3, $A(\theta|\sigma^*)$ may be bounded, and hence Theorem 4 does not hold. In this sense, Assumption 3 is essential to our argument. Finally, it is worth noting that the quadratic-loss model is a special case of the environment satisfying these assumptions.

It is also obvious that equilibrium σ^* constructed in Theorem 4 exhibits both divergent interpretation and divergent prediction, and their complementary nature implies unbounded $A(\theta|\sigma^*)$ as in the quadratic-loss case. However, the extension of Theorems 2 and 3 appears non-trivial. Because our arguments fully exploit the structure of quadratic-loss preferences, they cannot be directly applied to the general environment. Hence, this exercise is left for future research.

B.1 Proof of Theorem 4

First, we introduce additional notation. For each d and θ, let $\nu_d(a^1(\theta, d)) = \theta$. We interpret $\nu_d(a) \in \Theta$ as the payoff-state in which a is the ideal action of the sender with bias d. Note that $\nu_d(a)$ is well-defined, and is continuous both in a and in d by Assumptions 2 and 3. Moreover, $\nu_0(a^2(\theta)) = \theta$ by

Assumption 2-(ii).

Lemma 6. Under Assumptions 2 and 3, $\nu_d(a)$ is: (i) strictly increasing in a given any $d \in D$; and (ii) strictly decreasing in d given any $a \in A$.

Proof. (i) Fix arbitrary $d \in D$ and $a, a' \in A$ with $a > a'$. Let $\nu_d(a) = \theta$ and $\nu_d(a') = \theta'$. By definition, $a^1(\theta, d) = a > a' = a^1(\theta', d)$. Because $a^1_1 > 0$, we have $\theta > \theta'$, or equivalently, $\nu_d(a) > \nu_d(a')$.

(ii) Fix arbitrary $a \in A$ and $d, d' \in D$ with $d > d'$. Let $\nu_d(a) = \theta$ and $\nu_{d'}(a) = \theta'$. By definition, $a^1(\theta, d) = a^1(\theta', d') = a$. Because $a^1_1 > 0$, $a^1_2 > 0$ and $d > d'$, we have $\theta < \theta'$, or equivalently, $\nu_d(a) < \nu_{d'}(a)$. □

We consider the same Harsanyi's type space $\mathcal{T}$ as in the proof of Theorem 1, and hence we omit its description (see the proof of Theorem 1). We now show that $A(\theta|\sigma^*) = A$ for any $\theta \in \Theta$.

First, we consider the level-0 types of each player, that is, T^0_i for $i = 1, 2$. Because we assume that a fully-revealing equilibrium is played among level-0 types, their equilibrium strategies are $\sigma^*_1(t^0_1(\theta)) = \theta$ and $\sigma^*_2(t^0_2, m) = a^2(m)$, respectively. Hence, $A^0(\theta|\sigma^*) = \{a^2(\theta)\}$.

Next, we consider the level-1 types of each player, that is, T^1_i for $i = 1, 2$. Because the sender with type $t^1_1(d, \theta) \in T^1_1$ believes that each message m induces action $a^2(m)$, his best response $\sigma^*_1(t^1_1(d, \theta))$ must be such that $a^2(\sigma^*_1(t^1_1(d, \theta))) = a^1(\theta, d)$, or equivalently, $\sigma^*_1(t^1_1(d, \theta)) = \nu_0(a^1(\theta, d))$.

Then, the receiver with type $t^1_2(d)$ who receives message m believes that the state is $\nu_d(a^2(m))$. Thus, her best response is $\sigma^*_2(t^1_2(d), m) = a^2(\nu_d(a^2(m))) =$

$a^2(\nu_d(\sigma_2^*(t_2^0, m)))$. Because ν is continuous and strictly decreasing in d by Lemma 6, we have:

$$d \in D \Longleftrightarrow \nu_d(a^2(\theta)) \in [\nu_\varepsilon(a^2(\theta)), \nu_{-\varepsilon}(a^2(\theta))], \tag{50}$$

and hence:

$$\begin{aligned} A^1(\theta|\sigma^*) &= [a^2(\nu_\varepsilon(a^2(\theta))), a^2(\nu_{-\varepsilon}(a^2(\theta)))] \\ &= [a^2(\nu_\varepsilon(\sigma_2^*(t_2^0, \theta))), a^2(\nu_{-\varepsilon}(\sigma_2^*(t_2^0, \theta)))]. \end{aligned} \tag{51}$$

Note that $A^0(\theta|\sigma^*) \subsetneq A^1(\theta|\sigma^*)$ for any $\theta \in \Theta$.

By induction, we consider level-$(k+1)$ types of each player, that is, T_i^{k+1} for each $i = 1, 2$. As an induction hypothesis, we assume that for any $\theta \in \Theta$:

$$A^k(\theta|\sigma^*) = [a^2(\nu_\varepsilon(\alpha_-^{k-1}(\theta))), a^2(\nu_{-\varepsilon}(\alpha_+^{k-1}(\theta)))] \supsetneq A^{k-1}(\theta|\sigma^*) = [\alpha_-^{k-1}(\theta), \alpha_+^{k-1}(\theta)],$$

where $\alpha_-^{k-1}(\theta) = \min A^{k-1}(\theta|\sigma^*)$ and $\alpha_+^{k-1}(\theta) = \max A^{k-1}(\theta|\sigma^*)$. Let $t_2 \in T_2^k$. Because the sender with type $t_1^{k+1}(d, \theta, t_2)$ believes that each message m induces action $\sigma_2^*(t_2, m)$, his best response $\sigma_1^*(t_1^{k+1}(d, \theta, t_2))$ must be such that $\sigma_2^*(t_2, \sigma_1^*(t_1^{k+1}(d, \theta, t_2))) = a^1(\theta, d)$.

Then, the receiver with type $t_2^{k+1}(d', t_2) \in T_2^{k+1}$ who receives message m believes that the state is $\nu_{d'}(\sigma_2^*(t_2, m))$. Thus, her best response is $\sigma_2^*(t_2^{k+1}(d', t_2), m) = a^2(\nu_{d'}(\sigma_2^*(t_2, m)))$. Because ν is continuous and strictly decreasing in d by

Lemma 6, for each $\theta \in \Theta$ and $a \in A^k(\theta|\sigma^*)$:

$$d \in D \Longleftrightarrow \nu_d(a) \in [\nu_\varepsilon(a), \nu_{-\varepsilon}(a)].$$

Also, because ν is continuous and strictly increasing in a by Lemma 6, for each $\theta \in \Theta$ and $d \in D$:

$$a \in A^k(\theta|\sigma^*) \Longleftrightarrow \nu_d(a) \in [\nu_d(\alpha_-^k(\theta)), \nu_d(\alpha_+^k(\theta))],$$

where $\alpha_-^k(\theta) = \min A^k(\theta|\sigma^*)$ and $\alpha_+^k(\theta) = \max A^k(\theta|\sigma^*)$. Hence:

$$A^{k+1}(\theta|\sigma^*) = [a^2(\nu_\varepsilon(\alpha_-^k(\theta))), a^2(\nu_{-\varepsilon}(\alpha_+^k(\theta)))]. \tag{52}$$

Because $a_1^2 > 0$, $\nu_\varepsilon(\alpha_-^k(\theta)) < \alpha_-^k(\theta)$ and $\alpha_+^k(\theta) < \nu_{-\varepsilon}(\alpha_+^k(\theta))$, we have $A^{k+1}(\theta|\sigma^*) \supsetneq A^k(\theta|\sigma^*)$ for any θ.

Finally, we show that $\tilde{A}(\theta|\sigma^*) = \bigcup_k A^k(\theta|\sigma^*) = A$ for any $\theta \in \Theta$. Suppose contrarily that there exists $\theta \in \Theta$ such that $\tilde{A}(\theta|\sigma^*) \neq A$; that is, either $\inf \tilde{A}(\theta|\sigma^*) > -\infty$ or $\sup \tilde{A}(\theta|\sigma^*) < +\infty$. Without loss of generality, assume that $\inf \tilde{A}(\theta|\sigma^*) = \tilde{\alpha}_-(\theta) > -\infty$. Let $\mathcal{A} = [\tilde{\alpha}_-(\theta), a^2(\theta)]$, and define a function $\Delta : \mathcal{A} \to \mathbb{R}$ so that $\Delta(a) = \nu_0(a) - \nu_\varepsilon(a)$ for $a \in \mathcal{A}$. Because Δ is continuous on a compact set $\mathcal{A}$ and $\Delta(a) > 0$ for all $a \in \mathcal{A}$, there exists $\tilde{a} \in \mathcal{A}$ such that $\Delta(a) \geq \Delta(\tilde{a}) = \tilde{\delta} > 0$ for all $a \in \mathcal{A}$.

Lemma 7. For any k, $\nu_\varepsilon(\alpha_-^k(\theta)) \leq \theta - (k+1)\tilde{\delta}$.

Proof. We prove the statement by induction on k. For $k = 0$, $\alpha_-^0(\theta) =$

$a^2(\theta)$. Hence, $\Delta(\alpha_-^0(\theta)) = \nu_0(\alpha_-^0(\theta)) - \nu_\varepsilon(\alpha_-^0(\theta)) = \theta - \nu_\varepsilon(\alpha_-^0(\theta)) \geq \tilde{\delta}$, or $\nu_\varepsilon(\alpha_-^0) \leq \theta - \tilde{\delta}$. Suppose that this inequality holds up to k. For $k+1$, we have $\Delta(\alpha_-^{k+1}(\theta)) = \nu_0(a^2(\nu_\varepsilon(\alpha_-^k(\theta)))) - \nu_\varepsilon(\alpha_-^{k+1}(\theta)) = \nu_\varepsilon(\alpha_-^k(\theta)) - \nu_\varepsilon(\alpha_-^{k+1}(\theta)) \geq \tilde{\delta}$, and hence, $\nu_\varepsilon(\alpha_-^{k+1}(\theta)) \leq \nu_\varepsilon(\alpha_-^k(\theta)) - \tilde{\delta} \leq \theta - (k+1)\tilde{\delta} - \tilde{\delta} = \theta - (k+2)\tilde{\delta}$. Thus, we obtain $\nu_\varepsilon(\alpha_-^k(\theta)) \leq \theta - (k+1)\tilde{\delta}$ for all k. $\square$

Because $\tilde{\alpha}_-(\theta)$ is finite, $\nu_\varepsilon(\tilde{\alpha}_-(\theta))$ is also finite. However, Lemma 7 says that $\nu_\varepsilon(\alpha_-^k(\theta)) < \nu_\varepsilon(\tilde{\alpha}_-(\theta))$ holds for sufficiently large k. This implies $\alpha_-^k(\theta) < \tilde{\alpha}_-(\theta)$, which contradicts $\alpha_-^k(\theta) = \inf A^k(\theta) \geq \inf \tilde{A}(\theta|\sigma^*) = \tilde{\alpha}_-(\theta)$. Therefore, $\inf \tilde{A}(\theta|\sigma^*) = -\infty$ for any $\theta \in \Theta$, and likewise, $\sup \tilde{A}(\theta|\sigma^*) = +\infty$ for any $\theta \in \Theta$. We thus conclude that $\tilde{A}(\theta|\sigma^*) = A$ for any $\theta \in \Theta$, which implies that $A(\theta|\sigma^*) = A$ for any $\theta \in \Theta$ because $\tilde{A}(\theta|\sigma^*) \subseteq A(\theta|\sigma^*)$. $\square$

References

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

Blume, A. (2018): "Failure of Common Knowledge of Language in Common-Interest Communication Games," *Games and Economic Behavior*, 109, 132–155.

Blume, A., and O. J. Board (2013): "Language Barriers," *Econometrica*, 81(2), 781–812.

——— (2014): "Intentional Vagueness," *Erkenntnis*, 79(4), 855–899.

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

Carlsson, H., and E. van Damme (1993): "Global Games and Equilibrium Selection," *Econometrica*, 61(5), 989–1018.

Chassang, S., and S. Takahashi (2011): "Robustness to Incomplete Information in Repeated Games," *Theoretical Economics*, 6(1), 49–93.

Chen, Y., N. Kartik, and J. Sobel (2008): "Selecting Cheap-Talk Equilibria," *Econometrica*, 76(1), 117–136.

Chen, Y.-C., S. Takahashi, and S. Xiong (2014): "The Weinstein-Yildiz Critique and Robust Predictions with Arbitrary Payoff Uncertainty," mimeo, National University of Singapore and University of Bristol.

Crawford, V. P. (2003): "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions," *American Economic Review*, 93(1), 133–149.

Crawford, V. P., and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50(6), 1431–1451.

Diehl, C., and C. Kuzmics (2014): "The (Non-)Robustness of Influential Cheap Talk Equilibria When the Sender's Preferences are

State-Independent," mimeo, Bielefeld University and Université Paris 1 Panthéon–Sorbonne.

Dimitrakas, V., and Y. Sarafidis (2006): "Cheap Talk from a Sender with Unknown Motives," mimeo, BlackRock and Charles River Associates.

Giovannoni, F., and S. Xiong (2017): "Communication under Language Barriers," mimeo, University of Bristol.

Gordon, S. (2010): "On Infinite Cheap Talk Equilibria," mimeo, SciencesPo.

——— (2011): "Iteratively Stable Cheap Talk," mimeo, SciencesPo.

Kajii, A., and S. Morris (1997): "The Robustness of Equilibria to Incomplete Information," *Econometrica*, 65(6), 1283–1310.

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

Kartik, N., M. Ottaviani, and F. Squintani (2007): "Credulity, Lies, and Costly Talk," *Journal of Economic Theory*, 134(1), 93–116.

Lam, H., and C. Mottet (2016): "Tail Analysis without Tail Information: A Worst-case Perspective," mimeo, University of Michigan and Boston University.

Li, M., and K. Madarász (2008): "When Mandatory Disclosure Hurts: Expert Advice and Conflicting Interests," *Journal of Economic Theory*, 139(1), 47–74.

Lightle, J. P. (2014): "The Paternalistic Bias of Expert Advice," *Journal of Economics & Management Strategy*, 23(4), 876–898.

Lu, S. E. (2017): "Coordination-Free Equilibria in Cheap Talk Games," *Journal of Economic Theory*, 168, 177–208.

Meyer-ter-Vehn, M., and S. Morris (2011): "The Robustness of Robust Implementation," *Journal of Economic Theory*, 146(5), 2093 – 2104.

Miura, S., and T. Yamashita (2014): "On the Possibility of Information Transmission: A (Costly) Signaling Case," mimeo, Kanagawa University and Toulouse School of Economics.

Monderer, D., and D. Samet (1989): "Approximating Common Knowledge with Common Beliefs," *Games and Economic Behavior*, 1(2), 170–190.

Morgan, J., and P. C. Stocken (2003): "An Analysis of Stock Recommendations," *RAND Journal of Economics*, 34(1), 183–203.

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

MORRIS, S., AND T. UI (2005): "Generalized potentials and robust sets of equilibria," *Journal of Economic Theory*, 124(1), 45–78.

NAGEL, R. (1995): "Unraveling in Guessing Games: An Experimental Study," *American Economic Review*, 85(5), 1313–1426.

OURY, M., AND O. TERCIEUX (2007): "Contagion in Games with Strategic Complementarities," mimeo, HEC and Paris School of Economics.

PEI, D. (2015): "Uncertainty about Uncertainty in Communication," mimeo, Massachusetts Institute of Technology.

RUBINSTEIN, A. (1989): "The Electronic Mail Game: Strategic Behavior under "Almost Common Knowledge"," *American Economic Review*, 79(3), 385–391.

SOBEL, J. (1985): "A Theory of Credibility," *Review of Economic Studies*, 52(4), 557–573.

SPENCE, M. (1973): "Job Market Signaling," *Quarterly Journal of Economics*, 87(3), 355–374.

STAHL, D. O., AND P. W. WILSON (1994): "Experimental evidence on players' models of other players," *Journal of Economic Behavior & Organization*, 25(3), 309–327.

——— (1995): "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10(1), 218–254.

Strzalecki, T. (2014): “Depth of Reasoning and Higher Order Beliefs,” *Journal of Economic Behavior & Organization*, 108, 108–122.

Weinstein, J., and M. Yildiz (2007): “A Structure Theorem for Rationalizability with Application to Robust Predictions of Refinements,” *Econometrica*, 75(2), 365–400.