On the possibility of information transmission*

Shintaro Miura† Takuro Yamashita‡

July 12, 2014

Abstract

We study a cheap-talk game *à la* Crawford and Sobel (1982), where it is common knowledge that the bias parameter is close to zero with a high probability. We show that "assuming full revelation when the bias is commonly believed to be zero" yields a drastically different prediction for the equilibrium behavior of those who have the bias arbitrarily close to (but possibly different from) zero. We interpret the result as casting a question for a common practice of selecting the "most informative" equilibrium in (applied) cheap-talk settings.

1 Introduction

A class of games of strategic information transmission, called *cheap-talk* games, provides an important theoretical framework to study which kind of information may be transmitted between asymmetrically informed players in various contexts. In the literature, it is often assumed that the preference structure of the game is common knowledge among players. Although it may

*We are grateful to Attila Ambrus, Sidartha Gordon, Navin Kartik, Kohei Kawamura, Frederic Koessler, Christoph Kuzmics, Michel Le Breton, Ming Li, Joel Sobel, and Satoru Takahashi for their helpful comments.

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

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

seem to be a reasonable "simplifying assumption" in many situations, in view of an outside observer who attempts to predict possible outcome of this environment of strategic information transmission, it is not always possible to precisely know the players' preference and belief structures. Even if the observer could be sure that a specific preference structure is "close to" common knowledge in a certain sense (maybe based on some data), there may remain a possibility that the observer's model is slightly misspecified. Therefore, it would be an important exercise to examine how the results could change if we relax some of common knowledge assumptions.

More specifically, we consider a situation where it is commonly known that the parameter values of interest are "consistently estimated", i.e., the parameter estimates are close to the true values with a high probability (or in other words, they are close in the topology of convergence in probability). For example, there may exist some publicly available data, both to the players and the observer, which yield such consistent estimators. In this sense, the players and outside observer nearly agree to each other about the parameter values, but not completely. In particular, the observer cannot exclude any such (high-order) beliefs of the players that do not contradict this situation that everyone nearly agrees to each other.

To illustrate the idea concretely, in Section 2 and 3, we consider a cheap-talk environment à la Crawford and Sobel (1982) with quadratic-loss utilities, where the sender's preference may be biased than the receiver's. As opposed to the standard approach where this bias parameter is assumed to be common knowledge, we consider a situation where the bias parameter is close to but not necessarily common knowledge. To be specific, we assume that the bias parameter is close to zero (i.e., the players' preferences are close to being aligned) in the topology of convergence in probability.[1]

Our goal is to study what this observer can predict about the players' be-

[1] Although we focus on a simple cheap-talk environment to illustrate our main idea, similar insights are applicable to some other environments as well. For example, a companion paper of ours, Miura and Yamashita (2014), obtains a qualitatively similar result in a costly-signaling model à la Spence (1973).

haviors, in particular about information transmission in equilibria. A common practice in (applied) cheap-talk settings is that, when the game has multiple equilibria, then the maximum-partition equilibrium is selected for various reasons. In case the bias is zero, this corresponds to the fully-revealing equilibrium.[2] Although it may sound natural to believe that full revelation occurs with no bias, we show that this assumption can have a rather drastic implication on the equilibrium behavior of the players for whom the bias is close to (but not necessarily exactly) zero. More precisely, we show that there exists a type space of Harsanyi (1967-68) such that (i) it is commonly known that the bias is close to zero in the topology of convergence in probability, and (ii) for any perfect Bayesian equilibrium such that full revelation occurs in any belief-closed subspace where the bias is commonly believed to be zero, the set of equilibrium actions that may be played by the receiver does not vary with the realized state, and furthermore, coincides with the entire action space. Therefore, the outside observer, who knows that it is commonly known that the bias is close to zero in our sense, cannot predict anything about the outcomes in those equilibria. In Section 4, we further show that this no-prediction result holds beyond the quadratic-loss model.

We interpret this result as casting a question for a common practice of selecting the most informative equilibrium for various reasons in the (applied) cheap-talk literature.[3] Even when such a selection induces a sharp prediction for the model where no bias is common knowledge, if that common-knowledge model is a slightly misspecified one, then the price of such misspecification could be significant.

[2]Spector (2000) and Agastya, Bag, and Chakraborty (2014) show that the maximum-partition equilibrium converges to the fully-revealing equilibrium as the bias goes to 0. This result holds even in a general environment beyond the uniform-quadratic model.

[3]In some environments, a less informative equilibrium may rather be regarded as a reasonable outcome. For example, see Gordon (2011).

1.1 Some related papers

In cheap-talk games, there often exist multiple equilibria, and selecting reasonable equilibria has been studied in the literature. Given that the maximum-partition equilibrium (which is also called the "most-informative" equilibrium) is the one that is ex ante Pareto dominant in the Crawford-Sobel setting, many applied papers focus on such equilibria. Recently, Chen, Kartik, and Sobel (2008) find a selection criterion called the NITS ("no incentive to separate") condition, which selects the maximum-partition equilibrium under certain regularity conditions.[4] Our aim is to point out that such a selection may have a drastic implication in a slightly misspecified environment.

The idea of uncertain bias in one-shot interactions is studied in the literature.[5] The bias parameter is either zero or a constant positive in Morgan and Stocken (2003), and is either a constant positive or a constant negative in Li and Madarász (2008). Dimitrakas and Sarafidis (2006) and Diehl and Kuzmics (2014) consider situations where the sender's preference is parametrized by a continuous random variable. Gordon (2010, 2011) analyze a framework including the scenario where the bias parameter could be state dependent. However, 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.[6] We are rather interested in implications of (possibly infinite) high-order uncertainty in terms of the bias parameter, even though it is commonly known that it is close to

[4] Under Condition (M) of Crawford and Sobel (1982), the most informative equilibrium is the unique equilibrium satisfying the NITS condition. Recently, Gordon (2011) suggests another criterion called "iterative stability", which can select a unique equilibrium in the environment where NITS condition fails to provide a unique prediction.

[5] See also Sobel (1985), Benabou and Laroque (1992) and Morris (2001) that study the impact of uncertain bias in repeated interactions.

[6] Diehl and Kuzmics (2014) also consider a situation with third-order uncertainty in the state parameter (but without any uncertainty in the sender's preference) and show that their conclusion would be the same as in their main model, i.e., the one with the receiver's first-order uncertainty for the sender's preference and the state parameter.

zero.

The robustness of fully revealing equilibrium is discussed by Lu (2013) in the context of multiple-sender models. Lu (2013) characterizes the equilibria that are robust to small noises into the senders' private information, and obtains a negative implication for the validity of the fully revealing equilibrium.[7] Although our motivation and perturbation are related to those of Lu (2013), these papers investigate different issues. Specifically, Lu (2013) studies ε-equilibria in the perturbed games to select a reasonable equilibrium in the base game, while we study exact equilibria of the perturbed games to discuss the impact of an equilibrium in the base game to the perturbed games.[8]

In terms of methodology, this paper is related to the literature of robust prediction in games with respect to players' high-order beliefs. In this literature, one often considers certain perturbation of players' high-order beliefs from the base model, and see the implication of such perturbation. The following two types of perturbations are frequently considered in the literature. We discuss the connections to the literature in more detail at the end of Section 3.

Some papers such as Rubinstein (1989), Weinstein and Yildiz (2007), Penta (2013), and Chen, Takahashi, and Xiong (2014) consider situations where the observer assumes that the players mutually know the base model up to an (arbitrarily high) finite level, but not up to an infinite level (and hence the model is not *common* knowledge). More specifically, each player believes that the base model is the true model, each player believes that each player believes that the base model is the true model, and so on, up to an arbitrary finite level, but at a sufficiently high level, this mutual belief breaks down: some types of a player may believe that the true model is very

[7] In contrast, Ambrus and Lu (2014) restrict the class of perturbation, and show that almost fully revealing equilibria are robust to the restricted perturbation.

[8] There are several other differences between the setups of these papers. For example, Lu (2013) studies a perturbation to the senders' observations in a multiple-sender model, while we consider a perturbation to the bias parameter in a single-sender model.

different from the base model, and non-existence of such types cannot be guaranteed in a sufficiently high level. In fact, in these papers, the predictions in games are crucially affected by those types of a player whose beliefs are significantly different from those presumed in the base model. For example, Weinstein and Yildiz (2007) obtain interim correlated rationalizability as the observer's prediction under such a perturbation. Penta (2013), and more recently Chen, Takahashi, and Xiong (2014), generalize the framework developed by Weinstein and Yildiz (2007) to environments where each type's belief is more restricted than Weinstein and Yildiz (2007) (for example, based on some common knowledge assumptions imposed by the observer).

Another set of papers, such as Monderer and Samet (1989), Kajii and Morris (1997), and Morris and Ui (2005), considers another approach of perturbing players' beliefs such that the base model is common p-belief with p close to (but not) one.[9] This notion of perturbation is generalized in Oury and Tercieux (2007)[10] to allow for small misspecification of the model possibly with a high probability, referred to as $(1-p, \eta)$-*elaboration*. Namely, it is common p-belief that the actual game played is within η ($\in \mathbb{R}$) of the base model (in terms of the players' preferences). More specifically, each player assigns a probability at least p to the event that the true model is within η of the base model, each player assigns a probability at least p to the event that each player assigns a probability at least p to the event that the true model is within η of the base model, and so on, up to an infinite level.[11] These studies investigate whether the predicted behavior in the base model is "robust" in the sense that a similar behavior is an equilibrium even if the model is not common knowledge but almost common knowledge in this sense.[12]

[9] There are papers that combine both of those perturbations, such as Oyama and Tercieux (2010).

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

[11] For example, the perturbation adopted by Kajii and Morris (1997) is a $(1-p, 0)$-elaboration (Oury and Tercieux (2007)).

[12] Morris and Ui (2005) investigate the set of behaviors in the base model that are robustly predicted in the sense that any strategy that may played in perturbed models is

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 ($i = 2$).[13] The sender knows the true state of the world, $\theta \in \Theta$, while the receiver does not. The receiver takes an action $a \in A$. We consider a message game where the sender sends $m \in M$ in the first stage, and then the receiver takes $a \in A$ in the second stage after observing m. In the following, we assume $\Theta = A = M = \mathbb{R}$ to simplify the argument. However, the same result applies in a more standard case of bounded Θ and A.[14] The sender's utility is $u = -(a - \theta - d)^2$, while the receiver's utility is $v = -(a - \theta)^2$, where $d \in D = \mathbb{R}$ represents the difference in their preferences, called the *bias* parameter.

In the standard model, we assume that d is known, and θ follows a specific (common-knowledge) distribution. Instead, we allow for the possibility that the model is slightly misspecified from the point of view of the modeler.

To illustrate the idea concretely, we assume that the modeler believes that the bias parameter is close to zero (so their preferences are almost aligned), but the players may have a slightly different view. To make the departure from the standard model small, we still assume that it is common knowledge between the players that "d is close to zero with a high probability".[15] More specifically, there exists $\varepsilon > 0$ such that it is common knowledge that

$$\Pr(|d| \leq \varepsilon) > 1 - \varepsilon.$$

On the other hand, with respect to θ, we assume a (full-support) common-knowledge distribution, which we denote by μ, as in the standard assumption.

in that set.

[13]Following the convention, we treat the sender as male and the receiver as female throughout this paper.

[14]The result is available from the authors upon the request.

[15]Hence, ε is close to zero in the sense of the convergence in probability. The assumption may be reasonable in a setting where the modeler has a consistent estimator (perhaps based on data) of the bias parameter.

Even with these assumptions, the players' beliefs, represented by Harsanyi's 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 $b_i : T_i \to \Delta(T_{-i})$, i.e., given his type t_i, his belief over T_{-i} is given by $b_i(t_i) \in \Delta(T_{-i})$. 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 $b_2^D(t_2) \in \Delta(D)$, i.e., for each measurable set $E \subseteq D$,

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

and his marginal belief about θ is denoted by $b_2^\Theta \in \Delta(\Theta)$, i.e., for each measurable set $E' \subseteq \Theta$,

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

Let $D_\varepsilon = [-\varepsilon, \varepsilon]$. To represent the common knowledge assumptions introduced above, for every t_2, we assume that (i) $b_2^D(D_\varepsilon|t_2) > 1 - \varepsilon$, and (ii) $b_2^\Theta(\cdot|t_2) = \mu(\cdot)$.

Let $\mathbb{T}$ represent the class of the type spaces satisfying those conditions. We consider a situation where the modeler does not know which $\mathcal{T} \in \mathbb{T}$ is the true type space.

In the message game, given $\mathcal{T}$, let $\sigma_1 : T_1 \to M$ denote the sender's (pure) strategy, and $\sigma_2 : T_2 \times M \to A$ denote the receiver's (pure) strategy.[16] Let $\sigma^* = (\sigma_1^*, \sigma_2^*)$ denote a perfect Bayesian equilibrium in the message game. We are interested in the set of actions that may be played in each θ, if indeed the bias is small, i.e., $d \in D_\varepsilon$. Let $A^*(\theta)$ denote the set of equilibrium actions

[16] We focus on pure strategies to save the notation, but allowing for mixed strategies does not change the result.

of the receiver when the true state is θ and the bias is in D_ε, i.e.,

$$A^*(\theta) = \left\{\sigma_2^*\big(t_2|\sigma_1^*(t_1)\big) \;\middle|\; (t_1,t_2) \in T \text{ s.t. } d(t_1) \in D_\varepsilon,\ \theta(t_1) = \theta\right\}.$$

Our goal is to study what kind of prediction the modeler could expect for $A^*(\theta)$ in a situation where the modeler does not know which $\mathcal{T} \in \mathbb{T}$ is a true type space.

3 An implication of "assuming full revelation when $d = 0$ is common knowledge"

A common practice in (applied) cheap-talk settings is that, when the game has multiple equilibria, then the most informative (and hence the ex-ante welfare maximizing) equilibrium is selected for various reasons. In case the bias is zero, this corresponds to the fully-revealing equilibrium. Although it may sound natural to believe that full revelation occurs with no bias, we show that this assumption can have a rather drastic implication on the equilibrium behavior of the players for whom $d = 0$ is not common knowledge (but d being close to zero with a high probability is).

To formalize the idea, we introduce the following additional notation. Given a type space $\mathcal{T} = (T_1, T_2, b_1, b_2) \in \mathbb{T}$, consider a subset of types $\tilde{T}_i \subseteq T_i$ for each i such that, for each i and $t_i \in \tilde{T}_i$, $b_i(\tilde{T}_{-i}|t_i) = 1$.[17] If, in addition, $d(t_1) = 0$ for each $t_1 \in \tilde{T}_1$, we say that *no bias is commonly believed in* $\tilde{T} = \tilde{T}_1 \times \tilde{T}_2$.

Definition 1. A perfect Bayesian equilibrium σ^* given $\mathcal{T}$ has *Property FR0* if, for any belief-closed subset $\tilde{T}$ such that no bias is commonly believed, for every $\theta \in \Theta$, $t_1 \in \tilde{T}_1$ such that $\theta(t_1) = \theta$, and $t_2 \in \tilde{T}_2$, we have $\sigma_2^*(t_2|\sigma_1^*(t_1)) = \theta$.

That is, whenever $d = 0$ is commonly believed among the players, then full revelation of information occurs.

[17]Hence $\tilde{T}$ is a *belief-closed* subset.

Theorem 1. There exists $\mathcal{T} \in \mathbb{T}$ such that, for any perfect Bayesian equilibrium σ^* with Property FR0, we have $A^*(\theta) = A$ for every θ.

That is, if we assume that full revelation occurs among the players for those who commonly believe $d = 0$, then there exists a Harsanyi's type space $\mathcal{T}$ in which no prediction becomes possible for those who do not commonly believe $d = 0$ (but do commonly believe d being close to zero with a high probability).

The rest of the section is devoted to the proof of the theorem.

Proof. We first construct $\mathcal{T} = (T_1, T_2, b_1, b_2)$ as follows. First, define T_i^0 for each i as follows. 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 state θ, and (iii) believes that the receiver's type is in T_2^0, i.e.,

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

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., $b_2(T_1^0 | t_2^0) = 1$).[18]

Note that $d = 0$ is commonly believed among them. Therefore, we will assume that full revelation occurs among them, and see how this assumption implies for the other types' equilibrium behaviors.

Next, for each $d \in D_\varepsilon$, 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 state θ, and (iii) believes that the receiver's type is t_2^0 for certain, i.e.,

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

Let $T_1^1 = \bigcup_{d \in D_\varepsilon} T_1^1(d)$.

[18] By assumption, t_2^0 believes that θ follows μ, and the sender's type is $t_1^0(\theta)$ for certain when the state is θ.

Let $T_2^1 = \{t_2^1(d) | d \in D_\varepsilon\}$ be a set of "level-1" types of the receiver, where for each $d \in D_\varepsilon$, $t_2^1(d)$ believes that the sender's type is in $T_1^1(d)$ (i.e., $b_2(T_1^1(d)|t_2^1(d)) = 1$).[19]

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 $d \in D_\varepsilon$ and $t_2 \in T_2^k$, let $T_1^{k+1}(d, t_2)$ be such that $T_1^{k+1}(d, t_2) = \{t_1^{k+1}(d, \theta, t_2) | \theta \in \Theta\}$, 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 state θ, and (iii) believes that the receiver's type is t_2 for certain, i.e.,

$$d(t_1^{k+1}(d, \theta, t_2)) = d, \ \theta(t_1^{k+1}(d, \theta, t_2)) = \theta, \text{ and } b_1(T_2^k | t_1^{k+1}(\theta)) = 1.$$

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

Similarly, let T_2^{k+1} be another subset of the receiver's types ("level-$(k+1)$" types) as follows. We let $T_2^{k+1} = \{t_2^{k+1}(d, t_2) | d \in D_\varepsilon, t_2 \in T_2^k\}$, where, for each $d \in D_\varepsilon$ 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., $b_2(T_1^{k+1}(d, t_2)|t_2^{k+1}(d, t_2)) = 1$).[20]

We complete the description of the type space by defining $T_i = \bigcup_{k=0}^{\infty} T_i^k$ for each i.[21] One interpretation may be that type 0 is the "naive" type who believes that there is no conflict in their preferences. A type of the sender in T_1^k tries to best respond to a type of the receiver in T_2^{k-1}, and a type of the receiver in T_2^k tries to best respond to a type of the sender in T_1^k.[22]

[19] By assumption, $t_2^1(d)$ believes that θ follows μ, and the sender's type is $t_1^1(d, \theta)$ for certain when the state is θ. Also, because $t_2^1(d)$ believes that the bias is $d (\in D_\varepsilon)$ with probability one, our assumption that $\Pr(|d| \in D_\varepsilon) > 1 - \varepsilon$ is satisfied.

[20] By assumption, $t_2^{k+1}(d, t_2))$ believes that θ follows μ, and the sender's type is $t_1^{k+1}(d, \theta, t_2)$ for certain when the state is θ. Also, because $t_2^{k+1}(d, t_2)$ believes that the bias is $d (\in D_\varepsilon)$ with probability one, our assumption that $\Pr(|d| \in D_\varepsilon) > 1 - \varepsilon$ is satisfied.

[21] As is clear from the construction, in this type space, not only $\Pr(d \in D_\varepsilon) > 1 - \varepsilon$ but also $\Pr(d \in D_\varepsilon) = 1$ is common knowledge among the players (regardless of their types). In this sense, our result could be stated in a slightly stronger way. Nevertheless, we prefer the current approach because of its natural relation with the topology of the convergence in probability.

[22] Although our construction of the type space is analogous to that often used in the

Now we consider any perfect Bayesian equilibrium given this type space, σ^*, with Property FR0. 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 $m \in M$.

The following lemma completes the proof of the theorem.

Lemma 1. $A^*(\theta) = A$ for any θ.

Proof. (of the lemma)

As in the statement, we assume that $t_1^0(\theta)$ reports θ truthfully, and t_2^0 plays $\sigma_2^*(t_2^0|m) = m$ if she receives m.

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 $d \in D_\varepsilon$. Because she believes the sender's type is one of those in $T_1^1(d)$, her unique best response is $\sigma_2^*(t_2^1|m) = m - d$. Let

$$\begin{aligned} A^1(\theta) &= \{\sigma_2^*(t_2^1|\theta)|t_2^1 \in T_2^1\} \\ &= [\theta - \varepsilon, \theta + \varepsilon], \end{aligned}$$

where $A^1(\theta)$ denotes the set of the actions that the receiver in T_2^1 can play in the equilibrium, if she receives message θ (for example, the sender with type $t_1^0(\theta)$ sends message 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 $d \in D_\varepsilon$ 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

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, we interpret the level as a measure of distance from the base model where $d = 0$ is common knowledge. Strzalecki (2010) adopts the level-k theory for representing each player's higher-order beliefs about the opponents' depth of reasoning.

receiver with type $t_2^{k+1}(d', t_2) \in T_2^{k+1}$ where $d' \in D_\varepsilon$. Because she believes the sender's type is one of those in $T_1^{k+1}(d', t_2)$, her unique best response is $\sigma_2^*(t_2^{k+1}(d', t_2)|m) = m - d' - \delta_k$. Let

$$\begin{aligned} A^{k+1}(\theta) &= \{\sigma_2^*(t_2^{k+1}|\theta) | t_2^{k+1} \in T_2^{k+1}\} \\ &= [\theta - (k+1)\varepsilon, \theta + (k+1)\varepsilon], \end{aligned}$$

where $A^{k+1}(\theta)$ denotes the set of the actions that the receiver in T_2^{k+1} can play in the equilibrium, if she receives message θ (for example, the sender with type $t_1^0(\theta)$ sends message θ).

Therefore, $A^*(\theta) = \bigcup_k A^k(\theta) = \mathbb{R}(= A)$ for every θ. □

□

The intuition behind the proof is roughly as follows. As we assume, if the sender has no bias (type 0), and reports θ truthfully, then the type-0 receiver takes $a = \theta$. Now, if the sender is actually biased, then he may report untruthfully. In particular, if this biased sender believes that the receiver has type 0, then he would report θ plus the level of his bias. Given this, now the receiver who believes that the sender is of such a type would try to adjust the action according to the level of the bias, i.e., her action would be the reported message minus the level of the bias. But then, the sender who believes such type of the receiver would adjust his message even more so that the message he sends is the true state plus *twice* of the level of his bias. By induction, it is shown that "type-k" receiver can take any action that is $k\varepsilon$ away from the reported message.

The main message of the result is the following. If we make a "seemingly natural" assumption that full revelation occurs for those who commonly believe $d = 0$, then we need to allow for any action choice at any θ for those who commonly believe that d is close to zero with a high probability.[23] Note that,

[23] Note that if we consider ε-equilibria instead of the exact equilibria like Lu (2013), then the no-prediction result as in Theorem 1 does not hold. More precisely, for the sender with any $d \in D_\varepsilon$, truth-telling is an ε-best response (though not exactly a best response) if the

on the equilibrium path, the receiver's belief (on Θ) and action vary with respect to received messages. In this sense, communication plays a non-trivial role in σ^*, especially *given* the type of the receiver being fixed. However, because the observer does not know which type of the receiver realizes, any action choice is reasonable at any θ (i.e., $A^*(\theta) = A$ for any θ). Therefore, in view of the observer, any meaningful prediction about the equilibrium outcome is impossible.[24]

It may be worth noting that, in this sense that communication plays a non-trivial role, σ^* is fundamentally different from a "babbling equilibrium" (i.e., an equilibrium where the sender sends the same message regardless of θ, and the receiver plays $a = E(\theta)$ for any received message). As we mentioned above, in σ^*, the receiver's belief (on Θ) and action vary with respect to received messages, while in the babbling equilibrium, the receiver's belief and action are invariant given whatever messages are observed. On the other hand, the observer can precisely predict the equilibrium outcome in the babbling equilibrium, while in σ^*, the observer cannot make any meaningful prediction. In other words, we say that the impossibility of meaningful predictions is fundamentally different from the well-known impossibility of information transmission in the babbling equilibrium.

Remark. Our approach in perturbing the base model in the sense of the convergence in probability is related to other well-known approaches in the literature mentioned in the introduction. In this remark, we briefly discuss the main differences between our approach and those in the literature. Recall that there are two key features in our construction. First, *every* type of the players in a possible type space believes that the true model is "close" to the base model, and second, this itself is common knowledge among the players.

receiver follows the sender's recommendation, and such a behavior of the receiver is a best response if the sender reports truthfully. As a result, the fully-revealing strategy profile is an ε-equilibrium.

[24]As the convention in the literature, a mapping from state of nature to induced action in that equilibrium is called *equilibrium outcome*.

In the approach where the perturbation is based on (arbitrarily high) finite-order mutual knowledge, such as in Rubinstein (1989), Weinstein and Yildiz (2007), and Penta (2013), there exist some types of players ("commitment types") who believe that the true model is very different from the base model in that they have dominant (or uniquely rationalizable) actions. Moreover, it cannot be *common* knowledge among the "normal" or "non-commitment" types of the players that such commitment types do not exist. The non-commitment types' actions are hence significantly affected by those of commitment types, and such a contagious structure is crucial for their results.[25] This is in contrast with our approach where there does not exist such a commitment type whose belief about the true model is very different from the other types.

In another approach where perturbation is based on common p-belief, such as in Monderer and Samet (1989), Kajii and Morris (1997), Morris and Ui (2005), and Oury and Tercieux (2007), in order to obtain sufficient conditions for their robustness, they (implicitly) consider possibilities that those types who do not exhibit common p-belief may play arbitrary strategies in the game. Again, this is in contrast with our approach in that we do not assume that certain types play arbitrarily. In this sense that we do not assume that certain types play arbitrarily, we consider a weaker robustness test than Oury and Tercieux (2007). In fact, our construction of the type spaces can be interpreted as a special case of $(\varepsilon, \varepsilon)$-elaboration of Oury and Tercieux (2007).[26]

Which kinds of perturbation is relevant would depend on the specificity of the uncertainties faced by the modeler or the observer. We believe that

[25] Recently, Chen, Takahashi, and Xiong (2014) generalizes the framework to situations where such commitment types who have dominant actions do not necessarily exist, but their study has a similar feature in that some types may have very different beliefs than other types, and such significant belief difference is crucial for their robust prediction results.

[26] Because other papers such as Kajii and Morris (1997) consider a $(1-p, 0)$-elaboration, it is not directly comparable with our approach.

our notion of perturbation is relevant if there is publicly available data both for the players and the observer to consistently estimate parameters of the model, but they do not necessarily completely agree on the actual parameter values.

4 More General Preferences

While we have focused on the quadratic-loss preferences in the previous sections, our no-prediction result 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 $D_\varepsilon = [-\varepsilon, \varepsilon]$. We impose the following assumptions on u and v.

Assumption 1. 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 exist $a^1(\theta, d) = \arg\max_{a \in A} u(a, \theta, d)$ and $a^2(\theta) = \arg\max_{a \in A} v(a, \theta)$ for any $\theta \in \Theta$ and $d \in D_\varepsilon$.

Assumption 1 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 $d \in D_\varepsilon$, (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 the no-prediction result holds beyond the quadratic-loss environment.

Assumption 2. $a^1(\cdot, d) : \Theta \to A$ is bijective for any $d \in D_\varepsilon$.[27]

As in Section 3, we consider any perfect Bayesian equilibrium such that full revelation occurs if $d = 0$ is common knowledge. As in the quadratic-loss case, we say that such a perfect Bayesian equilibrium satisfies *Property FR0*.

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

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 2. Suppose that Assumptions 1 and 2 hold. Then, there exists $\mathcal{T} \in \mathbb{T}$ such that, for any perfect Bayesian equilibrium σ^* with Property FR0, we have $A^*(\theta) = A$ for any $\theta \in \Theta$.

The proof is essentially parallel to (though more complicated than) the proof of Theorem 1. Therefore, we prove Theorem 2 in the appendix.

There are three remarks about the assumptions. First, Assumption 1 and its implications are standard in the literature, for example, as in Crawford and Sobel (1982). Second, we need Assumption 2 in order to apply the argument used in the previous section to this general environment, although it is not standard in the literature. Assumption 2 means that any available action can be supported as the sender's ideal action in some state whatever the bias parameter d is. It implies a one-to-one relationship between observed messages and states given any fixed bias parameter. Without Assumption 2, $A^*(\theta)$ may be bounded, and hence Theorem 2 does not hold. In this sense, Assumption 2 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.

5 Concluding remarks

We studied a certain class of cheap-talk games *à la* Crawford and Sobel (1982) (which includes a quadratic-loss case as a special case), where it is common knowledge that the bias parameter is close to zero with a high probability. We showed that there exists a type space that satisfies the following. For any perfect Bayesian equilibrium of the game given this type space where full revelation occurs among the types of the players who commonly believe the bias to be zero, *any* action may be played by some type of the receiver in *any* state of the world. For the observer who cannot exclude the possibility that this type space is the true type space that describes the players' (high-order) beliefs, essentially no prediction is possible for the equilibrium outcome.

The result casts a question for a common practice of selecting the most informative equilibrium for various reasons in the (applied) cheap-talk literature. Even when such a selection induces a sharp prediction for the model where $d = 0$ is common knowledge, if that common-knowledge model is a slightly misspecified one, then the price of such misspecification could be significant.

There are three concluding remarks. First, the no-prediction result of this paper was shown in a somewhat special setting. For example, the sets of the states and actions are unbounded, and $d = 0$ (rather than some non-zero value) is fixed as a "benchmark" case. It is left open for future research how much our result generalizes to alternative specifications in these senses.

Also, we focused on the fully revealing equilibrium among many other equilibria when $d = 0$ is common knowledge. It would be interesting to study if similar no-prediction results hold even if we assume other (non-fully-revealing) equilibrium behaviors in the base model. If the conclusion turns out to vary across different assumptions on the equilibrium behaviors, the methodology offered in this paper could be interpreted as providing a new dimension of comparing equilibria in the base model, based on behaviors in nearby environments. This might be useful for arguing equilibrium selection and ranking of equilibria.

A related question is whether similar no-prediction results can occur in other environments. This paper focused on a cheap-talk environment to illustrate in a simple model our approach of the model misspecification and its potential consequence. However, the basic idea of this paper applies to some other environments as well. For example, a companion paper of ours, Miura and Yamashita (2014), shows a similar no-prediction result as in this paper 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 state of the world in nearby environments in the topology of convergence in probability.

Finally, despite our seemingly negative result, our "common sense" tells us that some information transmission should be possible in a number of situations even if some parameters are not necessarily common knowledge. A fruitful future direction might be a more extensive study of the conditions on the players' information and strategic environments under which (at least some amount of) information can "robsutly" be transmitted. As one of such robustness tests, we believe that the perturbation in the topology of convergence in probability could be useful.

Appendix: Proof of Theorem 2

First, we introduce additional notation. For each d and θ, we define $\nu_d(a^1(\theta, d)) = \theta$. We interpret $\nu_d(a) \in \Theta$ as the 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 1 and 2. Moreover, $\nu_0(a^2(\theta)) = \theta$ by Assumption 1 (ii).

Lemma 2. Under Assumptions 1 and 2, $\nu_d(a)$ is (i) strictly increasing in a given any $d \in D_\varepsilon$, and (ii) strictly decreasing in d given any $a \in A$.

Proof. (i) Fix arbitrary $d \in D_\varepsilon$ 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 $d, d' \in D_\varepsilon$ 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 $A(\theta) = A$ for any $\theta \in \Theta$.

First, we consider the level-0 types of each player, i.e., 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 $\sigma_1^*(t_1^0(\theta)) = \theta$ and $\sigma_2^*(t_2^0|m) = a^2(m)$, respectively. Hence,

$$A^0(\theta) = \{\sigma_2^*(t_2^0|\theta)|t_2^0 \in T_2^0\} = \{a^2(\theta)\}.$$

Next, we consider the level-1 types of each player, i.e., T_i^1 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 $t_2^1(d)$ who receives message m believes that the state is $\nu_d(a^2(m))$. Thus, her best response is $\sigma_2^*(t_2^1(d)|m) = a^2(\nu_d(a^2(m))) = a^2(\nu_d(\sigma_2^*(t_2^0|m)))$. Because ν is continuous and strictly decreasing in d by Lemma 2, we have

$$d \in D_\varepsilon \iff \nu_d(a^2(\theta)) \in [\nu_\varepsilon(a^2(\theta)), \nu_{-\varepsilon}(a^2(\theta))],$$

and hence,

$$\begin{aligned} A^1(\theta) &= \{\sigma_2^*(t_2^1|\theta)|t_2^1 \in T_2^1\} = [a^2(\nu_\varepsilon(a^2(\theta))), a^2(\nu_{-\varepsilon}(a^2(\theta)))] \\ &= [a^2(\nu_\varepsilon(\sigma_2^*(t_2^0|\theta))), a^2(\nu_{-\varepsilon}(\sigma_2^*(t_2^0|\theta)))]. \end{aligned}$$

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

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

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

where $\alpha_-^{k-1}(\theta) = \min A^{k-1}(\theta)$ and $\alpha_+^{k-1}(\theta) = \max A^{k-1}(\theta)$. 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 2, for each $\theta \in \Theta$ and $a \in A^k(\theta)$,

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

Also, because ν is continuous and strictly increasing in a by Lemma 2, for each $\theta \in \Theta$ and $d \in D_\varepsilon$,

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

where $\alpha_-^k(\theta) = \min A^k(\theta)$ and $\alpha_+^k(\theta) = \max A^k(\theta)$. Hence,

$$A^{k+1}(\theta) = \{\sigma_2^*(t_2^{k+1}|\theta)|t_2^{k+1} \in T_2^{k+1}\} = [a^2(\nu_\varepsilon(\alpha_-^k(\theta))), a^2(\nu_{-\varepsilon}(\alpha_+^k(\theta)))].$$

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

Finally, we show that $A^*(\theta) = \bigcup_k A^k(\theta) = A$ for any $\theta \in \Theta$. Suppose contrarily that there exists $\theta \in \Theta$ such that $A^*(\theta) \neq A$. That is, either $\inf A^*(\theta) > -\infty$ or $\sup A^*(\theta) < +\infty$. Without loss of generality, assume that $\inf A^*(\theta) = \alpha_-^*(\theta) > -\infty$. Let $\mathcal{A} = [\alpha_-^*(\theta), a^2(\theta)]$, 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 $\hat{a} \in \mathcal{A}$ such that $\Delta(a) \geq \Delta(\hat{a}) = \hat{\delta} > 0$ for all $a \in \mathcal{A}$.

Lemma 3. For any k, $\nu_\varepsilon(\alpha_-^k(\theta)) \leq \theta - (k+1)\hat{\delta}$.

Proof. (of the lemma)

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

Because $\alpha_-^*(\theta)$ is finite, $\nu_\varepsilon(\alpha_-^*(\theta))$ is also finite. However, Lemma 3 says that, $\nu_\varepsilon(\alpha_-^k(\theta)) < \nu_\varepsilon(\alpha_-^*(\theta))$ holds for sufficiently large k. This implies

$\alpha^k_-(\theta) < \alpha^*_-(\theta)$, which contradicts $\alpha^k_-(\theta) = \inf A^k(\theta) \geq \inf A^*(\theta) = \alpha^*_-(\theta)$. Therefore, $\inf A^*(\theta) = -\infty$ for any $\theta \in \Theta$, and likewise, $\sup A^*(\theta) = +\infty$ for any $\theta \in \Theta$. We thus conclude that $A^*(\theta) = A$ for any $\theta \in \Theta$. □

References

Agastya, M., P. K. Bag, and I. Chakraborty (2014): "Proximate Preferences and Almost Full Revelation in the Crawford-Sobel Game," *Economic Theory Bulletin*, forthcoming.

Ambrus, A., and S. E. Lu (2014): "Almost Fully Revealing Cheap Talk with Imperfectly Informed Senders," mimeo, Duke University and Simon Fraser University.

Benabou, R., and G. Laroque (1992): "Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility," *Quarterly Journal of Economics*, 107(3), 921–58.

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), 1431–51.

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.

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

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

Harsanyi, J. C. (1967-68): "Games with Incomplete Information Played by Bayesian Players," *Management Science*, 14, 159–182,320–334,486–502.

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

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

Lu, S. E. (2013): "Coordination-Free Equilibria in Cheap Talk Games," mimeo, Simon Fraser University.

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), 1313–26.

Oury, M., and O. Tercieux (2007): “Contagion in Games with Strategic Complementarities,” mimeo, HEC and Paris School of Economics.

Oyama, D., and O. Tercieux (2010): “Robust Equilibria under Non-Common Priors,” *Journal of Economic Theory*, 145(2), 752–784.

Penta, A. (2013): “On the Structure of Rationalizability for Arbitrary Spaces of Uncertainty,” *Theoretical Economics*, 8(2).

Rubinstein, A. (1989): “The Electronic Mail Game: Strategic Behavior under “Almost Common Knowledge”,” *American Economic Review*, 79(3), 385–91.

Sobel, J. (1985): “A Theory of Credibility,” *Review of Economic Studies*, 52(4), 557–73.

Spector, D. (2000): “Pure Communication between Agents with Close Preferences,” *Economics Letters*, 66(2), 171–178.

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. (2010): "Depth of Reasoning and Higher Order Beliefs," mimeo, Harvard University.

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