

# Robust Predictions under Limited Depth of Reasoning

Kota Murayama\*

January, 2013

## Abstract

This paper elucidates how predictions in the depth of reasoning model, special cases of which include the Level-k and Cognitive Hierarchy models are robust to the common knowledge assumption of level-0 players' actions in two-player case. A prediction of the model is said to be  $p$ -dominant if level-0 players play a  $p$ -dominant action pair in the prediction. A sufficient condition is provided for a  $p$ -dominant prediction being robust to incomplete information *à la* Kajii and Morris (1997). Depending on assumed players' decision rules, even a  $p$ -dominant prediction with  $p \geq 1/2$  can be robust. A key mechanism behind this result is the effect of players' limited depth of reasoning on their strategic interaction through higher order beliefs, which is implied by Strzalecki (2010) in the case of Rubinstein's (1989) email game.

*Keywords:* Robustness, Limited Depth of Reasoning, Incomplete Information

---

\*Graduate School of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo, 186-8603, Japan. E-mail: murayama0209@gmail.com. I am most grateful to my advisor, Takashi Ui for his guidance and feedbacks that greatly improved this paper in many aspects. I would like to thank Hideshi Itoh, Atsushi Kajii, Takashi Kunimoto, and Takeshi Nishimura for time, advice, and encouragement. I also thank the seminar participants of Contract Theory Workshop East at Hitotsubashi University for helpful comments. Of course, all remaining errors are mine.

# 1 Introduction

Equilibrium analysis is often blamed for its low predictive power. A leading example is Rubinstein's (1989) email game. In the email game, players never reach cooperation, no matter how many confirmations they send, because of a grain of uncertainty in the information structure. This theoretical prediction is, however, not only intuitively unappealing, but has also rarely been supported by empirical evidence. For instance, Camerer (2003) reports that in an experimental setting less experienced players tend to choose a cooperative action when at most 6 or 7 emails are exchanged.

One of the most widely accepted non-equilibrium analyses is the depth of reasoning model, special cases of which include the Level- $k$  and Cognitive Hierarchy models (e.g., Nagel 1995; Stahl and Wilson 1995; Camerer et al. 2004).<sup>1</sup> In the depth of reasoning model, players follow the decision rule under which each player has a non-negative integer, interpreted as his reasoning level, and believes that other players have strictly lower reasoning levels than him. For instance, a level-1 player believes that his opponents are level-0 players with probability 1. A major distinction between the Level- $k$  and the Cognitive Hierarchy models is their way of specifying players' beliefs about other players' reasoning levels. In the Level- $k$  model, a player with level  $k$  always believes that his opponents have a reasoning level exactly one lower than him; that is,  $k - 1$ . On the other hand, the Cognitive Hierarchy model assumes that players commonly know the population distribution of reasoning levels (say Poisson distribution), and given his own level, each player calculates the conditional probability of opponents' levels induced by the distribution. Though players' beliefs are modeled differently, both share the following procedure to obtain a solution. First, specify level-0 players' actions and assume that they are common knowledge. Then their actions work as an anchor for players with higher reasoning levels, so that, by iterating best responses to that anchor, the actions of higher-level players are determined inductively. Thus a level- $k$  player knows what action his opponents will choose as long as their reasoning levels are strictly lower than  $k$ , and best responds to those actions under his belief as if he were the smartest among others.

While the depth of reasoning model obtains more consistent predictions with empirically observed data than the equilibrium analysis in a certain class of games (e.g., Keynes' Beauty Contest game; Hide and Seek game), the validity of assumptions made in the model has been challenged in recent literature.<sup>2</sup> In particular, how to identify the level-0 players' actions has been one of the central issues as the situations to which the depth of reasoning model is applied become diverse. Most of the existing literature assumes that level-0 players are naïve or non-strategic, and choose their actions randomly. Let us

---

<sup>1</sup>For the comprehensive survey, see Crawford et al. (2012)

<sup>2</sup>For instance, Agranov et al. (2012) show that individual choices crucially depend on his belief about other players' reasoning levels, which necessitates in considering each player's reasoning level as private information. This implies that experimental models should carefully specify players' beliefs about other players' reasoning levels.

consider a  $2/3$  guessing game, in which each player chooses one number from 0 to 100 and the player who has chosen the closest number to the  $2/3$  of the group average wins a prize. In this game, level-0 players are mostly supposed to choose uniformly randomly (i.e., choose 50 on average). While this choice seems natural at first, Burchardi and Penczynski (2011) identify that approximately one third of the participants were classified as level-0 players and the level-0 actions were not uniformly distributed in this guessing game.<sup>3</sup> In addition, as the complexity of games increases, the specification of level-0 players' actions for an outside observer becomes more subtle.<sup>4</sup> Crawford and Iriberry (2007) analyze the auction environment in which players follow the Level- $k$  decision rule, and they allow two possibilities for level-0 players' action: Either bid uniformly between the lowest and highest possible values or bid their value truthfully. Although these flexible specifications are effective to obtain robust predictions to the misspecification of level-0 players' actions, the need for such robustness inherently casts doubt on the common knowledge assumption of level-0 players' actions. That is, it is hard to believe that level-0 players' actions are common knowledge among players if the choice of level-0 players' actions is not so evident for the outside observer. In fact, Penczynski (2011) experimentally finds that players did not share the belief on the level-0 actions in the Hide and Seek game.

The purpose of this paper is to investigate how predictions in the depth of reasoning model depend on the common knowledge assumption of level-0 players' actions in two-player case. This question can be analogously understood to the classic question of game theory: "How sensitive the conclusions of game theory are to the common knowledge of payoff assumptions." Hence, we take an approach which examines how those predictions are robust to incomplete information *à la* Kajii and Morris (1997). Kajii and Morris (1997) consider the situation in which an outside analyst knows which game players will play with high probability but, with small probability, he does not know and the players can play a totally different game in terms of its payoff structure. Analogously, this paper considers the situation in which level-0 players' actions are given as a primitive of games, and the outside analyst knows which game players will play with high probability; however, with small probability, players can play a totally different game in terms of its payoff structure *and level-0 players' actions*. In this way, we answer the question: "How sensitive the predictions of depth of reasoning model are to the common knowledge assumption of level-0 players' actions."

Specifically, we formulate the depth of reasoning model as a complete information game with *depth of reasoning space*, in which each player is assigned a reasoning level and his belief about other players' reasoning levels satisfies the assumption we explained. Also, a solution of this game called *prediction* is the one we described above, and assume

---

<sup>3</sup>They find that the average number level-0 players chose was about 65 which is higher than 50. They explain this by pointing out the salience of 66 in this game.

<sup>4</sup>Arad and Rubinstein (2011) insist that the level-0 type specification should be intuitively appealing in games used for studying the Level- $k$  theory.

that level-0 players' actions are given as a primitive of this game.<sup>5</sup> We fix a prediction of this game, and say that it is *robust to incomplete information* if behavior of *all reasoning levels of each player* which is close to it constitutes a Bayesian prediction of every nearby incomplete information game with depth of reasoning space.<sup>6</sup> The word “nearby” means that the sets of players and actions are the same with original complete information game, and with high probability, each player knows that his payoffs *and level-0 players' actions* are the same. Thus, our robustness concept is different from the robustness of Kajii and Morris (1997) (henceforth, KM robustness) especially in the solution concept and the way of perturbation.

Among the sufficient conditions for KM robust equilibria, we focus on the sufficient condition that clarifies a relationship between players' higher order beliefs and  $p$ -dominance.<sup>7</sup> An action pair is said to be  $p$ -dominant if for both players, the action specified in it constitutes a best response whenever he believes that the other player also chooses the action specified in it with at least probability  $p$ . We say a  $p$ -dominant action pair induces a  $p$ -dominant prediction in the depth of reasoning model if level-0 players' actions are given by that  $p$ -dominant action pair. Kajii and Morris (1997) show that if  $p < 1/2$ , then a  $p$ -dominant action pair is KM robust. This paper examines how this sufficient condition would be changed in the depth of reasoning model. A theoretical motivation underlying this argument stems from the observation that players' limited depth of reasoning has a non-negligible effect on their strategic interaction through higher order beliefs. To the best of our knowledge, Strzalecki (2010) is the first who formulates a game theoretic model which includes the Level- $k$  and Cognitive Hierarchy models as special cases.<sup>8</sup> He introduces a “cognitive type space” into a complete information game in which each cognitive type assigns a reasoning level to each player and requires him to believe others have cognitive types which give them strictly lower reasoning levels than him.<sup>9</sup> He analyzes the email game with the cognitive type space, and shows that there exists a (cognitive) equilibrium in which both players choose cooperation after exchanging a certain number of emails, which cannot be attained without players' limited depth of reasoning.

To obtain a sufficient condition for robust predictions, this paper takes the following steps. Take any  $p$ -dominant prediction  $s^*$  and the  $p$ -dominant action pair  $a^*$  which induces

---

<sup>5</sup>In this solution, Level-0 players' actions are common knowledge and higher level players iterate his best response to that anchor under his belief about other players' reasoning levels.

<sup>6</sup>As we will define later, Bayesian prediction is a natural incomplete information extension of prediction in complete information setting.

<sup>7</sup>Kajii and Morris (1997) also show that if a complete information game has a unique correlated equilibrium, then that equilibrium is KM robust. Ui (2001), for instance, shows that the Nash equilibrium which maximizes the potential of the game is KM robust.

<sup>8</sup>In relation to our paper, Kneeland (2012) analyzes the global game under limited depth of reasoning, and shows that her model can explain the anomalies in experiments reported by Heinemann (2004), for instance.

<sup>9</sup>Strzalecki's cognitive type space is more general than our depth of reasoning space in the sense that his model allows players with the same reasoning level to have different beliefs about other players' levels.

$s^*$ . Construct events  $(E_{(i,k_i)})_{i=1,2}^{k_i}$  such that for player  $i$  with level  $k_i$ ,  $a_i^*$  is a best response at any states *not in*  $E_{(i,k_i)}$ . By using the known result in Oyama and Tercieux (2012), we can provide an upper bound for the ex ante probability of  $E_{(i,k_i)}$ , which is uniform to  $k_i$  for any  $i = 1, 2$ . Since this upper bound can be arbitrarily small as long as the degree of incomplete information perturbation is small, we can conclude that  $s^*$  is robust. From the fact that assigning a certain condition on players' decision rules slows down the increase of  $E_{(i,k_i)}$  in  $k_i$ , our main theorem shows that even a  $p$ -dominant prediction with  $p \geq 1/2$  can be robust. This result also implies that the Level- $k$  model has the smallest set of robust equilibria and the Cognitive Hierarchy model has the largest with respect to  $p$ -dominance, since the size of the set of robust predictions depends on which decision rule we choose. Thus, we now know that these two prominent models are quite different with respect to the set of robust equilibria.

The rest of the paper is organized as follows. In Section 2, we formulate the framework. Section 3 introduces two preliminary concepts,  $p$ -dominance and  $p$ -belief. Section 4 derives a sufficient condition for a  $p$ -dominant prediction being robust, and Section 5 concludes.

## 2 Framework

### 2.1 Complete Information Game with Depth of Reasoning Space

In this section, we formulate the depth of reasoning model as a complete information game with depth of reasoning space (henceforth, DR space), which is a special case of the cognitive type space introduced in Strzalecki (2010).<sup>10</sup> Let us introduce a complete information game given by  $(\mathcal{I}, (A_i)_{i \in \mathcal{I}}, (g_i)_{i \in \mathcal{I}})$ , where  $\mathcal{I}$  is the finite set of players, and for each player  $i \in \mathcal{I}$ ,  $A_i$  is the finite set of actions, and  $g_i : A \rightarrow \mathbb{R}$  is the (bounded) payoff function where  $A = \times_{i \in \mathcal{I}} A_i$ . Let  $\Delta(A)$  denote a collection of probability measure on  $A$ . We use similar notational conventions whenever they are clear from the context. Then our depth of reasoning model is defined as follows.

**Definition 1.** A complete information game with DR space denoted by  $\mathcal{G}$  is given by  $\mathcal{G} = (\mathcal{I}, (A_i)_{i \in \mathcal{I}}, (K_i)_{i \in \mathcal{I}}, (\mu_i)_{i \in \mathcal{I}}, (g_i)_{i \in \mathcal{I}})$ , where for each  $i \in \mathcal{I}$ ,  $K_i = \mathbb{Z}_+$  is the set of reasoning levels, and  $\mu_i : K_i \rightarrow \Delta(K_{-i})$  is his belief about other players' reasoning levels that satisfies  $\mu_i(k_i)(\{k_{-i} \in K_{-i} : k_j < k_i \text{ for each } j \in \mathcal{I} \text{ with } j \neq i\}) = 1$  whenever  $k_i > 0$ .

**Remark 1.** The restriction on  $(\mu_i(k_i))_{i \in \mathcal{I}}$  requires that player  $i$  with reasoning level  $k_i$  ( $>0$ ) believes that other players must have strictly lower reasoning levels than  $k_i$  with probability 1. This strictness of inequality is crucial since, if not, we need to deal with such fixed-point arguments as “I best respond to players who best respond to me.”

**Remark 2.** In this setting, at least one player has an incorrect perception of the real world; hence, the following analysis is often referred as a nonequilibrium analysis (Crawford et al. 2009, 2012; Crawford and Iriberry 2007).

<sup>10</sup>Cognitive type space with Property 1 in Strzalecki (2010) reduces to our DR space.

**Remark 3.** Note that we do not allow payoff functions to depend on reasoning levels. For this reason, we call  $k_i$  player  $i$ 's reasoning level to distinguish from the *type* in incomplete information games.

**Remark 4.** Our DR space considerably restricts the existence of common prior on reasoning levels. We say that  $\mu \in \Delta(K)$  is a common prior on  $K$  if  $\mu_i(k_i)(k_{-i}) = \mu(k_i)(k_{-i})$  for all  $\mathbf{k} \in K$ . If there exists a common prior  $\mu$ , the restriction on  $(\mu_i(k_i))_{i \in \mathcal{I}}$  implies that  $\mu(\mathbf{k}) = 0$  for all  $\mathbf{k} \in K$  whenever at least two coordinates of  $\mathbf{k}$  being strictly larger than 0.

## 2.2 Solution Concept

Let us denote player  $i$ 's (pure) strategy in  $\mathcal{G}$  by  $s_i : K_i \rightarrow A_i$  for each  $i \in \mathcal{I}$ . For our purpose, we assume that actions of level-0 players,  $(s_i(0))_{i \in \mathcal{I}}$ , are common knowledge. The game  $\mathcal{G}$  is now redefined as  $\mathcal{G} = (\mathcal{I}, (A_i)_{i \in \mathcal{I}}, (K_i)_{i \in \mathcal{I}}, (\mu_i)_{i \in \mathcal{I}}, (g_i)_{i \in \mathcal{I}}, (s_i(0))_{i \in \mathcal{I}})$  with slight abuse of notation.

We propose a solution concept in which level-0 players' actions are common knowledge so that their actions work as an anchor for the players with higher reasoning levels. In this solution, each player determines his action by the following inductive algorithm. Level-1 players best respond to level-0 players' actions which are given exogenously. Level-2 players best respond to the mixture of level-0 and level-1 players' actions under his belief on the likelihood of each level. Actions of players with higher reasoning levels are determined in the same way. Our solution concept only requires that, given his reasoning level, each player (except level-0 players) best responds to other players' strategies under his belief so that he has no incentive to change his behavior.<sup>11</sup>

**Definition 2.** A strategy profile  $s^*$  constitutes a *prediction* of  $\mathcal{G}$  if, for any  $i \in \mathcal{I}$ ,  $k_i \in \mathbb{N}$ , and  $a_i \in A_i$ ,

$$\sum_{k_{-i} \in K_{-i}} g_i(s^*(k_i, k_{-i})) \mu_i(k_i)(k_{-i}) \geq \sum_{k_{-i} \in K_{-i}} g_i(a_i, s^*_{-i}(k_{-i})) \mu_i(k_i)(k_{-i}).$$

Given  $(\mu_i)_{i=1,2}$ , at least one prediction must exist in  $\mathcal{G}$ , and the uniqueness is ensured generically in payoffs. This solution concept is the most frequently used one in experimental literature. Two typical depth of reasoning models, the Level- $k$  and the Cognitive Hierarchy, are translated into our framework as follows.

**Example 1.** (Level- $k$  Model)

In the Level- $k$  model, each player believes that other players' reasoning levels are exactly one lower than his reasoning level. That is, a player with reasoning level  $k_i$  believes that others have a reasoning level,  $k_i - 1$ , with probability 1. Formal representation is as follows: for each  $i \in \mathcal{I}$  and  $k_i \in \mathbb{N}$ ,

$$\mu_i(k_i)(k_i - 1, k_i - 1, \dots, k_i - 1) = 1.$$

<sup>11</sup>By construction, players must have mutually inconsistent beliefs in the solution.

The *Level-k prediction* is a prediction of  $\mathcal{G}$  in which all players have the Level-k type belief described above.

**Example 2.** (Cognitive Hierarchy Model)

The Cognitive Hierarchy model assumes that each player's reasoning level is identically and independently distributed with a common distribution, say,  $\lambda \in \Delta(\mathbb{Z}_+)$ . Camerer et al. (2004) assume that  $\lambda$  follows Poisson distribution. Given this  $\lambda$ , each player's belief is constructed as follows: for each  $i \in \mathcal{I}$  and  $k_i \in \mathbb{N}$ ,

$$\mu_i(k_i)(k_{-i}) = \frac{\prod_{j \neq i} \lambda(k_j)}{(\sum_{l=0}^{k_i-1} \lambda(l))^{|\mathcal{I}|-1}}$$

if  $k_j < k_i$  for each  $j \in \mathcal{I}$  with  $j \neq i$ , otherwise  $\mu_i(k_i)(k_{-i}) = 0$ .

The *Cognitive hierarchy prediction* is a prediction of  $\mathcal{G}$  in which all players have the Cognitive Hierarchy type belief described above.

### 2.3 Embedded Incomplete Information Game with DR Space

We introduce a private information into the complete information game with DR space,  $\mathcal{G}$ . An incomplete information game with DR space,  $\mathcal{U}$ , is given by  $\mathcal{U} = (\mathcal{I}, (A_i)_{i \in \mathcal{I}}, (K_i)_{i \in \mathcal{I}}, (\mu_i)_{i \in \mathcal{I}}, (u_i)_{i \in \mathcal{I}}, \Theta, (\Pi_i)_{i \in \mathcal{I}}, P, (s_i(0, \Theta))_{i \in \mathcal{I}})$ , where  $\Theta$  is a countable set of payoff states,  $P$  is a probability measure defined on the  $\sigma$ -field generated by  $\Theta$ , and for each player  $i \in \mathcal{I}$ ,  $\Pi_i$  is the set of information partitions of  $\Theta$ , and  $u_i : A \times \Theta \rightarrow \mathbb{R}$  is the payoff function. With slight abuse of notation, for each  $i \in \mathcal{I}$ , player  $i$ 's (pure) strategy in  $\mathcal{U}$  is given by  $s_i : K_i \times \Theta \rightarrow A_i$ . Here we also assume that the strategies of level-0 players,  $(s_i(0, \Theta))_{i \in \mathcal{I}}$ , are common knowledge in  $\mathcal{U}$ . Let us assume  $s_i$  and  $u_i$  are measurable with respect to  $\Pi_i$  for any  $i \in \mathcal{I}$ . We write  $P(\theta)$  for the probability of the singleton event  $\{\theta\}$  and  $\pi_i(\theta)$  for the element of  $\Pi_i$  containing  $\theta$ . Furthermore, assume that  $P(\pi_i(\theta)) > 0$  for any  $\theta \in \Theta$  and  $i \in \mathcal{I}$  to make the conditional probability well defined. If  $\mathcal{U}$  satisfies all the above properties, we say that  $\mathcal{U}$  embeds  $\mathcal{G}$ . We write  $E(\mathcal{G})$  for the set of incomplete information games with DR space which embed  $\mathcal{G}$ . A solution in  $\mathcal{U}$  is defined as an incomplete information extension of the prediction of  $\mathcal{G}$ .

**Definition 3.** A strategy profile  $s^*$  constitutes a *Bayesian prediction* of  $\mathcal{U}$  if, for any  $i \in \mathcal{I}$ ,  $k_i \in \mathbb{N}$ ,  $\theta \in \Theta$ , and  $a_i \in A_i$ ,

$$\begin{aligned} & \sum_{(k_{-i}, \theta') \in K_{-i} \times \pi_i(\theta)} u_i(s^*(\mathbf{k}, \theta'), \theta') P(\theta' | \pi_i(\theta)) \mu_i(k_i)(k_{-i}) \\ & \geq \sum_{(k_{-i}, \theta') \in K_{-i} \times \pi_i(\theta)} u_i(a_i, s_{-i}^*(k_{-i}, \theta'), \theta') P(\theta' | \pi_i(\theta)) \mu_i(k_i)(k_{-i}). \end{aligned}$$

### 2.4 Robustness

We say that a prediction of  $\mathcal{G}$  is robust to incomplete information if that prediction is played with high probability in some Bayesian prediction for any  $\mathcal{U} \in E(\mathcal{G})$  whenever  $\mathcal{U}$  is

sufficiently close to  $\mathcal{G}$ . To formally express this notion, let us firstly introduce a “distance” of two different strategies in  $\mathcal{U}$ .

**Definition 4.** Fix  $\mathbf{k} \in K$ . An action distribution for  $\mathbf{k}$  induced by a strategy profile in  $\mathcal{U}$ ,  $s(\mathbf{k}, \Theta)$ , is given by  $\alpha_{\mathbf{k}}(a) = \sum_{\theta \in \Theta} 1_{s(\mathbf{k}, \theta)}(a)P(\theta)$  for any  $a \in A$ , where  $1_{s(\mathbf{k}, \theta)}(a)$  is an indicator function which takes 1 if  $a$  is chosen given  $\theta$  under  $s(\mathbf{k}, \Theta)$ . In particular, we say an action distribution profile  $(\alpha_{\mathbf{k}})_{\mathbf{k} \in K}$  is a *prediction action distribution profile* of  $\mathcal{U}$  if there exists a Bayesian prediction  $s^*$  of  $\mathcal{U}$  such that  $\alpha_{\mathbf{k}}(a) = \sum_{\theta \in \Theta} 1_{s^*(\mathbf{k}, \theta)}(a)P(\theta)$  for any  $a \in A$  and  $\mathbf{k} \in K$ .

Then the following measure defines the “distance” between two action distribution profiles,  $\alpha$  and  $\beta$ :

$$\|\alpha - \beta\| = \sup_{\mathbf{k} \in K} \max_{a \in A} |\alpha_{\mathbf{k}}(a) - \beta_{\mathbf{k}}(a)|$$

Secondly, a complete information game with DR space  $\mathcal{G}$  is said to be close to an incomplete information game with DR space  $\mathcal{U}$ , if the strategies of level-0 players and payoff functions under  $\mathcal{U}$  are equal to those under  $\mathcal{G}$  with high probability and players know that. For each incomplete information game  $\mathcal{U} \in E(\mathcal{G})$ , write  $\Omega_{\mathcal{U}}$  for a collection of such payoff states:  $\Omega_{\mathcal{U}} \equiv \{\theta \in \Theta : s_i(0, \theta') = s_i(0)$  and  $u_i(a, \theta') = g_i(a)$  for all  $a \in A$ ,  $\theta' \in \pi_i(\theta)$  and  $i \in \mathcal{I}\}$ . An incomplete information game with DR space,  $\mathcal{U}$ , is an  $\varepsilon$ -elaboration of  $\mathcal{G}$  if  $\mathcal{U} \in E(\mathcal{G})$  and  $P(\Omega_{\mathcal{U}}) = 1 - \varepsilon$ . Let  $E(\mathcal{G}, \varepsilon)$  be the set of all  $\varepsilon$ -elaborations of  $\mathcal{G}$ . At last, we are ready to define the robustness of a prediction to incomplete information.

**Definition 5.** An action distribution profile  $\alpha$  is robust to incomplete information in  $\mathcal{G}$  if, given  $(\mu_i)_{i \in \mathcal{I}}$ , for every  $\delta > 0$ , there exists a  $\bar{\varepsilon} > 0$  such that every  $\mathcal{U} \in E(\mathcal{G})$  has a prediction action distribution profile  $\beta$  such that  $\|\alpha - \beta\| \leq \delta$  for all  $\varepsilon \leq \bar{\varepsilon}$ .

### 3 Preliminaries: $p$ -Dominance and $p$ -Belief

Following Monderer and Samet (1989) and Morris et al. (1995), this section introduces two concepts, the  $p$ -dominance and the  $p$ -belief operator. In the following analysis, we only consider two-player case, that is,  $\mathcal{I} = \{1, 2\}$ . Note that, if we use  $i$  and  $j$  at the same time, the  $j$  means “not  $i$ ”.

#### 3.1 $p$ -Dominance

We need the concept of the  $p$ -dominance to measure the “strength” of each action pair. Fix a complete information game with DR space  $\mathcal{G}$ . Let  $\lambda_i \in \Delta(A_j)$  for  $i = 1, 2$  and denote the probability assigned to  $a_j \in A_j$  under  $\lambda_i$  by  $\lambda_i(a_j)$ .

**Definition 6.** An action pair  $a^* \in A$  is said to be  $p$ -dominant in  $\mathcal{G}$  if, for any  $i = 1, 2$ ,



$a_i \in A_i$ , and  $\lambda_i \in \Delta(A_j)$  with  $\lambda_i(a_j^*) \geq p$ , we have

$$\sum_{a_j \in A_j} \lambda_i(a_j) g_i(a_i^*, a_j) \geq \sum_{a_j \in A_j} \lambda_i(a_j) g_i(a_i, a_j).^{12}$$

Thus,  $a_i^*$  becomes a best response for player  $i$  if he believes that  $a_j^*$  is played by player  $j$  with at least probability  $p$ . Let us say that a prediction,  $s^*$ , of  $\mathcal{G}$  is a *p-dominant prediction* if level-0 players' actions in  $s^*$  constitute a  $p$ -dominant action pair. Also, let us say that a  $p$ -dominant action pair  $a^*$  induces a  $p$ -dominant prediction  $s^*$ , if  $(s_1^*(0), s_2^*(0)) = (a_1^*, a_2^*)$ . Clearly,  $p$ -dominant action pair is played by any levels of both players in the induced  $p$ -dominant prediction.

### 3.2 $p$ -Belief Operator

Let  $\mathcal{F}_i$  denote the  $\sigma$ -algebra generated by  $\Pi_i$  for each  $i = 1, 2$ , and let  $\mathcal{F}_1 \oplus \mathcal{F}_2 \equiv \{E \subseteq \Theta : E = E_1 \cup E_2 \text{ for some } E_i \in \mathcal{F}_i \text{ for each } i = 1, 2\}$ . To characterize player's conditional belief at given payoff state, let us define the  $p$ -belief operator as in Monderer and Samet (1989).

**Definition 7.** For any  $E \in \mathcal{F}_1 \oplus \mathcal{F}_2$ , the  $p$ -belief operator for player  $i$  is given by  $B_i^p(E) \equiv \{\theta \in \Theta : P(E \mid \pi_i(\theta)) \geq p\}$ .

Thus,  $B_i^p(E)$  is the collection of states in which player  $i$  believes the event  $E$  with probability at least  $p$ . Following Morris et al. (1995), it is convenient for us to define  $C_i^p(E) = B_i^p(B_j^p(E)) \cup E$  as a contagion operator, and let us inductively define  $[C_i^p]^k(E) = C_i^p([C_i^p]^{k-1}(E))$  for  $k \geq 1$  with  $[C_i^p]^0(E) = E$ , and  $[C_i^p]^\infty(E) = \bigcup_{k=1}^\infty [C_i^p]^k(E)$ . Lemma A.4. of Oyama and Tercieux (2012), which is a key to our results, yields an upper bound of the ex-ante probability of  $[C_i^p]^k(E)$ .

**Lemma A.4. of Oyama and Tercieux (2012).** For any  $p \in (0, 1]$ , and any event  $E \in \mathcal{F}_1 \oplus \mathcal{F}_2$ , we have

$$P([C_i^p]^K(E)) \leq P(E) \sum_{k=0}^{2K} \left(\frac{1-p}{p}\right)^k \quad (1)$$

for all  $i = 1, 2$ .<sup>13</sup>

Note that, if  $p > 1/2$ , the right hand side of (1) converges to  $P(E) \cdot p/(2p-1)$  as  $K \rightarrow \infty$ .

<sup>12</sup>In a complete information game *without* DR space, each  $p$ -dominant action pair constitutes a Nash equilibrium.

<sup>13</sup>The original lemma gives an upper bound of  $P([H_*^p]^K(E))$ , where  $H_*^p(E) = B_1^p(E) \cup B_2^p(E)$  and  $[H_*^p]^k(E) = H_*^p([H_*^p]^{k-1}(E))$  for  $k \geq 1$  with  $[H_*^p]^0(E) = E$ . To derive the result here, we use the fact that  $[C_i^p]^K(E) \subseteq [H_*^p]^{2K}(E)$ . Also, the original lemma allows *non-common priors* but we do not.

## 4 Robust Predictions under Limited Depth of Reasoning

### 4.1 Motivating Example

To begin with, the following simple example gives us an intuition about how player's limited of depth of reasoning relates to the robustness of predictions. Consider a  $2 \times 2$  coordination game. There are two players, and each player chooses one of the two actions,  $L$  or  $R$ . The payoffs are given by

	$L$	$R$
$L$	$1 - p, 1 - p$	$0, 0$
$R$	$0, 0$	$p, p$

For the moment, let us assume  $p \in [1/2, 1)$ . Then playing  $L$  is a best response for each player if he believes that his opponent will also choose  $L$  with probability larger than  $p$ . On the other hand,  $R$  becomes a best response for each player if he believes that his opponent will also choose  $R$  with probability larger than  $1 - p$ . Hence, this game has a  $p$ -dominant action pair  $(L, L)$  and a  $(1 - p)$ -dominant action pair  $(R, R)$ . Let us consider the following incomplete information game that embeds this game. The information structure is given by the triple  $(\Theta, (\Pi_i)_{i=1,2}, P)$  where  $\Theta = \{1, 2, 3, \dots\}$  is the set of states,  $\Pi_1 = \{\{1\}, \{2, 3\}, \{3, 4\}, \dots\}$  and  $\Pi_2 = \{\{1, 2\}, \{3, 4\}, \{5, 6\}, \dots\}$  are partition structures for each player, and  $P$  is a common prior given by  $P(n) = \varepsilon(1 - \varepsilon)^{n-1}$  for any  $n \in \mathbb{N}$ . Thus, the posteriors are given by

$\Pi_1$	1	$\frac{1}{2-\varepsilon}$	$\frac{1-\varepsilon}{2-\varepsilon}$	$\frac{1}{2-\varepsilon}$	$\frac{1-\varepsilon}{2-\varepsilon}$	...
$\Theta$	1	2	3	4	5	...
$\Pi_2$	$\frac{1}{2-\varepsilon}$	$\frac{1-\varepsilon}{2-\varepsilon}$	$\frac{1}{2-\varepsilon}$	$\frac{1-\varepsilon}{2-\varepsilon}$	...	

Table 1: Partitions and posteriors.

Now suppose that  $R$  becomes both players' strict dominant action in the event  $E = \{1\}$ . Then  $R$  becomes player 2's unique best response in  $\{1, 2\}$  since he believes that player 1 plays  $R$  with at least (interim) probability  $1/(2 - \varepsilon) > 1 - p$ . But in turn,  $R$  becomes player 1's unique best response in  $\{2, 3\}$  since he believes that player 2 plays  $R$  with at least probability  $1/(2 - \varepsilon) > 1 - p$ . Continuing this argument yields that playing  $(R, R)$  everywhere is a unique Bayesian equilibrium in this game (Morris et al. 1995). That is, though  $(L, L)$  is a strict Nash equilibrium in the original game,  $(L, L)$  is never played in any Bayesian equilibrium of the embedding incomplete information game once  $(R, R)$  is played. In other words, this complete information game has a  $\varepsilon(2 - \varepsilon)$ -elaboration in which  $(L, L)$  is never played in any equilibria. Since  $\varepsilon$  is arbitrary, we can conclude that  $(L, L)$  is not KM robust.

Next, we introduce the DR space into the original complete information game, and suppose that level-0 players choose  $L$ . Then, any levels of both players choose  $L$  in any predictions of this game. On the other hand, let us introduce the information structure

as we defined above, and suppose that  $R$  is a strictly dominant action for any levels of both players and level-0 players choose  $R$  in  $E = \{1\}$ . Then  $R$  becomes a unique best response for player 2 in  $\{1, 2\}$  since he believes that any levels of player 1 play  $R$  with at least probability  $1/(2 - \varepsilon) > 1 - p$ .

$k_1 = 2$	$R$	$R$	$L$	$L$	$L$	...				
$k_1 = 1$	$R$	$R$	$L$	$L$	$L$	...				
$k_1 = 0$	$R$	$L$	$L$	$L$	$L$	...				
	1	2	3	4	5	6	7	8	9	...
$k_2 = 0$	$R$	$L$	$L$	$L$	$L$	$L$	$L$	$L$	$L$	...
$k_2 = 1$	$R$	$L$	$L$	$L$	$L$	$L$	$L$	$L$	$L$	...
$k_2 = 2$	$R$	$L$	$L$	$L$	$L$	$L$	$L$	$L$	$L$	...

Table 2: Level 0, 1, and 2 actions when all levels of both players play  $R$  in  $E$ .

Consider player 2 with reasoning level 2. In the event  $\{3, 4\}$ , his best response varies with his belief about player 1's reasoning level. Unlike the previous analysis, a simple calculation yields that, if he believes that player 1 will be a level-1 player with probability strictly smaller than  $(1 - p)(2 - \varepsilon)$ , choosing  $L$  becomes his unique best response. However, this condition is not enough to prevent the action  $R$  from spreading over the entire payoff states as players' reasoning levels get higher. This is because player 2 with reasoning level 3 can choose  $R$  in  $\{3, 4\}$  if he believes that player 1's reasoning level will be 1 or 2 with sufficiently high probability. For this particular type of information structure, Strzalecki (2010) shows that if both players' beliefs satisfy *Nondivergent beliefs property*, which says, for any strictly increasing sequence  $(k^m) \in \mathbb{N}^\infty$ ,  $\inf_m \mu_i(k^m)(k^{m-1} \leq k_j) < (1 - p)(2 - \varepsilon)$ , then there exists a number  $N$  such that any levels of both players choose  $L$  for any payoff state  $\theta \geq N$ .<sup>14</sup> This result indicates that, depending on the specification of player's beliefs about other player's reasoning level, a  $p$ -dominant prediction with  $p \geq 1/2$  may be robust to incomplete information.<sup>15</sup> The following analysis starts with formalizing our intuition here for any information structure and derives a sufficient condition for robust predictions in relation to  $p$ -dominance.

## 4.2 Robustness and $p$ -Dominance

Given a complete information game with DR space  $\mathcal{G}$  and beliefs  $(\mu_i)_{i=1,2}$ , we investigate the robustness of a  $p$ -dominant prediction  $s^*$ , where  $p \in [0, 1)$ . Let us denote by  $a^*$  a  $p$ -dominant action pair which induces  $s^*$ . Our proof proceeds as follows: Let us define inductively the events  $(E_{(i,k_i)})_{i=1,2}^{k_i}$  such that a  $p$ -dominant action  $a_i^*$  is a (interim) best response in  $\Theta \setminus E_{(i,k_i)}$  for player  $i$  with level  $k_i$ . By construction, it is easily shown that, for

<sup>14</sup>In Appendix A, we further investigate the effect of limited depth of reasoning on so called *contagion* and provide a generalized version of this result.

<sup>15</sup>Trivially, if any levels of both players surely believe his opponent is a level-0 player, a  $p$ -dominant prediction with  $p \geq 1/2$  is robust.

any  $\varepsilon$ -elaboration of the original game, there exists a Bayesian Prediction in which  $a_i^*$  is played in  $\Theta \setminus E_{(i,k_i)}$  by player  $i$  with level  $k_i$ . Then we find an upper bound of the ex ante probability of  $E_{(i,k_i)}$  using Lemma A.4. of Oyama and Tercieux (2012), which is uniform with respect to  $k_i$ . Since that upper bound can be arbitrarily small as long as the degree of perturbation  $\varepsilon$  is sufficiently small, we have obtained the robustness of  $s^*$  as desired.

Fix  $\mathcal{U} \in E(\mathcal{G}, \varepsilon)$  and let  $E_{(i,0)} = E = \Theta \setminus \Omega_{\mathcal{U}}$  for any  $i = 1, 2$ . We inductively define the event  $E_{(i,k_i)}$  as follows:

$$E_{(i,k_i)} = \left\{ \theta \in \Theta : \sum_{t=0}^{k_i-1} \mu_i(k_i)(t) P(E_{(j,t)} \mid \pi_i(\theta)) \geq 1 - p \right\} \cup E$$

for any  $k_i \in \mathbb{N}$  and  $i = 1, 2$ .

**Lemma 1.** *Suppose  $s^*$  is a  $p$ -dominant prediction of  $\mathcal{G}$  and  $a^*$  induces  $s^*$ . Then for any  $\mathcal{U} \in E(\mathcal{G}, \varepsilon)$ , there exists a Bayesian Prediction  $s'$  in which player  $i$  with level  $k_i$  plays  $a_i^*$  in  $\Theta \setminus E_{(i,k_i)}$  for any  $k_i \in K_i$  and  $i = 1, 2$ .*

*Proof.* Fix  $\mathcal{U} \in E(\mathcal{G}, \varepsilon)$ . Suppose player  $j$  follows  $s'_j$ , that is, player  $j$  with level  $k_j$  plays  $a_j^*$  in  $\Theta \setminus E_{(j,k_j)}$  for any  $k_j \in K_j$ . Take any  $k_i \in K_i$ . If  $k_i = 0$ , we must have  $s'_i(0, \theta) = a_i^*$  for any  $\theta \in \Omega_{\mathcal{U}} = \Theta \setminus E_{(i,0)}$  by construction. Suppose  $k_i \geq 1$ . Then since  $a^*$  is  $p$ -dominant,  $a_i^*$  becomes a best response at  $\theta \in \Theta \setminus E$  for player  $i$  with level  $k_i$  if we have  $\sum_{t=0}^{k_i-1} \mu_i(k_i)(t) P(\Theta \setminus E_{(j,t)} \mid \pi_i(\theta)) \geq p$ . But then

$$\begin{aligned} \Theta \setminus E_{(i,k_i)} &= \left\{ \theta \in \Theta : \sum_{t=0}^{k_i-1} \mu_i(k_i)(t) P(E_{(j,t)} \mid \pi_i(\theta)) < 1 - p \right\} \setminus E \\ &= \left\{ \theta \in \Theta : \sum_{t=0}^{k_i-1} \mu_i(k_i)(t) P(\Theta \setminus E_{(j,t)} \mid \pi_i(\theta)) > p \right\} \setminus E \\ &\subseteq \left\{ \theta \in \Theta : \sum_{t=0}^{k_i-1} \mu_i(k_i)(t) P(\Theta \setminus E_{(j,t)} \mid \pi_i(\theta)) \geq p \right\} \setminus E. \end{aligned}$$

Thus,  $a_i^*$  is a best response in  $\Theta \setminus E_{(i,k_i)}$  for player  $i$  with level  $k_i$ . Since  $k_i$  is arbitrary and players are symmetric, we have the result.  $\square$

Our construction of  $(E_{(i,k_i)})_{i=1,2}^{k_i}$  is tight in the sense that there exists a class of complete information games with DR space such that, for some  $\mathcal{U} \in E(\mathcal{G}, \varepsilon)$  and any Bayesian prediction  $s'$  of  $\mathcal{U}$ , we have  $s'_i(k_i, \theta) = a_i^*$  if and only if  $\theta \in \Theta \setminus E_{(i,k_i)}$  for any  $k_i \in K_i$  and  $i = 1, 2$ . In fact, our leading example with the Level- $k$  type belief falls into this class of games.

### 4.3 A Sufficient Condition for Robustness when $p < 1/2$

This section investigates an upper-bound of  $(E_{(i,k_i)})_{i=1,2}^{k_i}$  through looking at the case of both players having the Level- $k$  type belief, that is, for each  $i = 1, 2$  and  $k_i \in \mathbb{N}$ ,

$\mu_i(k_i)(k_j) = 1$  if  $k_j = k_i - 1$ , otherwise  $\mu_i(k_i)(k_j) = 0$ . It reveals that this specification of beliefs yields the “worst-case” of robustness in the sense that, if a  $p$ -dominant prediction  $s^*$  is robust when both players have the Level- $k$  type belief, then  $s^*$  is robust for any beliefs. Observe that when both players have the Level- $k$  type belief,  $(E_{(i,k_i)})_{i=1,2}^{k_i}$  can be written as

$$E_{(i,k_i)} = \{\theta \in \Theta : P(E_{(j,k_{i-1})} | \pi_i(\theta)) \geq 1 - p\} \cup E = B_i^{1-p}(E_{(j,k_{i-1})}) \cup E$$

for all  $k_i \in \mathbb{N}$  and  $i = 1, 2$ . Let us especially denote this by  $\widehat{E}_{(i,k_i)}$  for all  $k_i \in \mathbb{N}$  and let  $\widehat{E}_{(i,0)} = E_{(i,0)}$  for any  $i = 1, 2$ . The following Lemma 2 states that  $\widehat{E}_{(i,k_i)}$  is an upper bound of  $E_{(i,k_i)}$  in the sense of set inclusion.

**Lemma 2.** *For any  $i = 1, 2$  and  $k_i \in K_i$ ,  $E_{(i,k_i)} \subseteq \widehat{E}_{(i,k_i)}$ .*

*Proof.* See Appendix B. □

By Lemma 1 and Lemma 2, we know that for any  $(\mu_i)_{i=1,2}$  and  $\mathcal{U} \in E(\mathcal{G}, \varepsilon)$ , there exists a Bayesian prediction of  $\mathcal{U}$  such that the upper bound of the ex ante probability of player  $i$  with level  $k_i$  playing other actions than  $a_i^*$  is given by  $P(\widehat{E}_{(i,k_i)})$ . Hence, if  $p < 1/2$ , the robustness of a  $p$ -dominant prediction follows from Lemma A.4. of Oyama and Tercieux (2012).

**Proposition 1.** *Suppose  $s^*$  is a  $p$ -dominant prediction of  $\mathcal{G}$  with  $p < 1/2$ . Then, for any beliefs  $(\mu_i)_{i=1,2}$ ,  $s^*$  is robust to incomplete information.*

*Proof.* Let  $a^*$  denote a  $p$ -dominant action pair which induces  $s^*$ . Note that the prediction action distribution profile  $\alpha$  induced by  $s^*$  is given by  $\alpha_{\mathbf{k}}(a) = 1$  if  $a = a^*$  and  $\alpha_{\mathbf{k}}(a) = 0$  otherwise for any  $\mathbf{k} \in K$ . For notational simplicity, let  $q = 1 - p$ . Fix any  $\delta > 0$ , and let  $\varepsilon < \delta(2q - 1)/2q$ . Take any  $\mathcal{U} \in E(\mathcal{G}, \varepsilon)$ . By Lemma 2, for any  $i = 1, 2$  and  $k_i \in K_i$ ,  $E_{(i,k_i)} \subseteq \widehat{E}_{(i,k_i)} \subseteq \widehat{E}_{(i,2k_i)}$ . Straightforward calculation yields that for any  $k \in K_i$  and  $i = 1, 2$ ,  $\widehat{E}_{(i,k)} = [C_i^q]^{\frac{k}{2}}(E)$  if  $k$  is even and  $\widehat{E}_{(i,k)} = B_i^q[C_j^q]^{\frac{k-1}{2}}(E) \cup E$  if  $k$  is odd. But then by applying Lemma A.4. of Oyama and Tercieux (2012), for any  $i = 1, 2$  and  $k_i \in K_i$ ,  $P(E_{(i,k_i)}) \leq P(\widehat{E}_{(i,2k_i)}) = P([C_i^q]^{k_i}(E)) \leq \varepsilon \sum_{k=0}^{2k_i} (\frac{1-q}{q})^k$ . Since  $q > 1/2$  and  $\widehat{E}_{(i,k_i)}$  is increasing in  $k_i$ , by letting  $k_i \rightarrow \infty$ , we have  $P(E_{(i,k_i)}) \leq \varepsilon q / (2q - 1) < \delta/2$  for any  $k_i \in K_i$  and  $i = 1, 2$ . Then by Lemma 1, there exists a Bayesian prediction  $s'$  of  $\mathcal{U}$  such that  $P(\{\theta \in \Theta : (s'_1(k_1, \theta), s'_2(k_2, \theta)) \neq (a_1^*, a_2^*)\}) \leq P(E_{(1,k_1)}) + P(E_{(2,k_2)}) < \delta$  for any  $\mathbf{k} \in K$ . Thus, the prediction action distribution profile  $\beta$  induced by  $s'$  satisfies  $\beta_{\mathbf{k}}(a^*) > 1 - \delta$  for any  $\mathbf{k} \in K$ . Therefore,  $\|\alpha - \beta\| \leq \delta$  as desired. □

#### 4.4 A Sufficient Condition for Robustness

We now know the robustness of a  $p$ -dominant prediction whenever  $p < 1/2$ . But Proposition 1 does not tell us anything about the robustness of  $p$ -dominant prediction with

$p \geq 1/2$ , since the inequality in Lemma A.4. of Oyama and Tercieux (2012) becomes useless as  $k_i \rightarrow \infty$  if  $p \geq 1/2$ . The following Proposition 2 yields a negative result, that is, there exists a class of complete information games in which a  $p$ -dominant prediction with  $p \geq 1/2$  is not robust when both players have the Level- $k$  type belief. As we implied, the example in Section 4.1 falls into this class.

**Proposition 2.** *If players have the Level- $k$  type belief, then there exists a class of complete information games with DR space in which a  $p$ -dominant prediction cannot be robust if  $p \geq 1/2$ .*

*Proof.* See Appendix B. □

By Proposition 2, we can see that, in a certain class of games, the Level- $k$  type belief does not allow a  $p$ -dominant prediction to be robust if  $p \geq 1/2$ . This observation is natural since players in the Level- $k$  model increases the number of best response iteration as their reasoning levels get higher, and then a  $p$ -dominant action pair is contagiously played in a certain class of games (Morris et al. 1995). However, the following Theorem 1 shows that, under a certain condition on the specification of beliefs about his opponent's reasoning level,  $p$ -dominant prediction with  $p \geq 1/2$  can be robust to incomplete information. This occurs since each player believes that his opponent has sufficiently lower reasoning level than him and this decreases the size of  $(E_{(i,k_i)})_{i=1,2}^{k_i}$  as Lemma 2 suggests.

**Theorem 1.** *Suppose  $s^*$  is a  $p$ -dominant prediction of  $\mathcal{G}$  with  $p \in [0, 1)$ . Then, if there exists a  $n \in \mathbb{N}$  such that  $\sup_{k \geq 2n} \sum_{t=2n-1}^{k-1} \mu_i(k)(t) < 2(1-p)$  for any  $i = 1, 2$ , then  $s^*$  is robust to incomplete information.*

*Proof.* See Appendix B. □

Theorem 1 implies that the size of the set of robust predictions varies with how we specify the players' beliefs about other players reasoning levels. Since it is easy to verify the Cognitive Hierarchy type belief satisfies the assumption in Theorem 1 for any  $p \in [0, 1)$ , Corollary 1 immediately follows.

**Corollary 1.** *If players have the Cognitive Hierarchy type belief, then a  $p$ -dominant prediction is robust to incomplete information for any  $p \in [0, 1)$ .*

Combining this corollary with the result in Proposition 2, we now know that two prominent models, the Level- $k$  and the Cognitive Hierarchy, give two contrasting examples with respect to the size of robust predictions. Finally, let us consider a mixture of the Level- $k$  and Cognitive Hierarchy type beliefs.

**Example 3.** (A mixture of the Level- $k$  and Cognitive Hierarchy type beliefs)

Let  $\alpha \in [0, 1]$  and  $\lambda \in \Delta(\mathbb{Z}_+)$  be a fixed and full supported distribution. For any  $k_i \in K_i$ , define

$$\mu_i(k_i)(k_j) = \begin{cases} \alpha + (1 - \alpha)\lambda(k_j) & \text{if } k_j = k_i - 1 \\ (1 - \alpha)\lambda(k_j) & \text{if } k_j < k_i - 1. \end{cases}$$

As Strzalecki (2010) argues, it is easy to verify that this belief specification satisfies the assumption in Theorem 1 if  $\alpha < 2(1-p)$ . Figure 1 shows the set of  $p$ -dominant predictions which are robust to incomplete information whenever they exist, as  $\alpha$  varies.

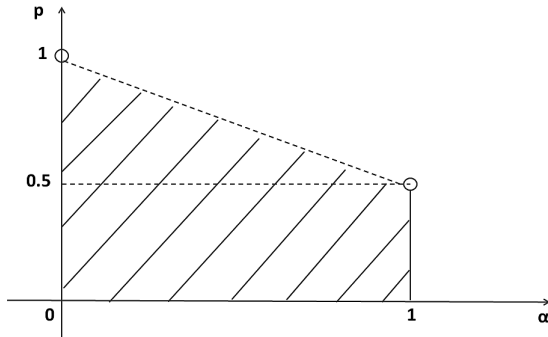


Figure 1. Set of robust predictions.

## 5 Conclusion

As a concluding remark, we discuss some future extensions. First, this paper restricts its attention only to DR space for simplicity. But we believe that the results here can be extended to the cognitive type space without much difficulty. Second, it is essential to extend our results to the many player case. Finally, recent papers such as Kets and Heifetz (2012) consider the type space in which players can do reasoning at most finitely many times and investigate a direct implication of such restriction on higher order beliefs. A relation to papers which take the higher order belief approach should be further investigated.

## Appendix A: Contagion under Limited Depth of Reasoning

This section investigates the effect of players' limited depth of reasoning on so called *contagion* by utilizing the techniques in the previous analysis. In the sequel, fix an incomplete information game with DR space  $\mathcal{U}$ , where  $|\Theta| = \infty$ , and  $|\pi_i| < \infty$  for any  $\pi_i \in \Pi_i$  and  $i = 1, 2$ . We say that an action pair  $a^*$  is *strict  $p$ -dominant* at  $\theta \in \Theta$  if, for any  $i = 1, 2$ ,  $a_i \in A_i$ , and  $\lambda_i \in \Delta(A_j)$  with  $\lambda_i(a_j^*) > p$ , we have  $\sum_{a_j \in A_j} \lambda_i(a_j) u_i(a_i^*, a_j, \theta) > \sum_{a_j \in A_j} \lambda_i(a_j) u_i(a_i, a_j, \theta)$ .<sup>16</sup> Thus,  $a_i^*$  becomes a (interim) *unique* best response for player  $i$  at  $\theta$  if he believes at  $\theta$  that  $a_j^*$  is played by player  $j$  with at least probability  $p$ . We say that an action pair  $a$  is played contagiously under  $(\mu_i)_{i=1,2}$  if, given  $(\mu_i)_{i=1,2}$ , once  $a$  is played by any levels of both players in some finite event  $E \in \mathcal{F}_1 \oplus \mathcal{F}_2$ , then for any Bayesian prediction  $s^*$  of  $\mathcal{U}$ , we have  $|\{\theta \in \Theta : (s_1^*(k_1, \theta), s_2^*(k_2, \theta)) = (a_1, a_2)\}| \rightarrow \infty$  as  $k_1, k_2 \rightarrow \infty$ . In incomplete information games *without* DR space, the *belief potential of event  $E$*  tells us whether such contagion occurs or not (Morris et al. 1995). The belief

<sup>16</sup>A strict  $p$ -dominant action pair is also a  $p$ -dominant action pair by continuity.

potential of event  $E$  is the largest probability  $p$  with which, at any states, such argument that player  $i$  believes that player  $j$  believes ... the event  $E$  holds for any  $i = 1, 2$ .

**Definition 8.** The belief potential of event  $E$  is given by  $\sigma(E) = \min_{i=1,2} \sigma_i(E)$ , where  $\sigma_i(E) = \sup\{p \in [0, 1] \mid [C_i^p]^\infty(E) = \Theta\}$ .

The main theorem of Morris et al. (1995) states that, suppose  $a^*$  is strict  $p$ -dominant at any states in  $\mathcal{U}$ . Then once  $a^*$  is played in some finite event  $E$  and if we have  $p < \sigma(E)$ , then  $a^*$  is played at any states in any equilibrium. In contrast, in our setting, we can show that even if  $a^*$  is played in  $E$  and  $p < \sigma(E)$ , under a certain condition on  $(\mu_i)_{i=1,2}$ ,  $a^*$  cannot be played contagiously. Before stating this result formally, let us introduce *the marginal belief potential of event* in order to characterize the process of contagion.

### A.1 The Marginal Belief Potential of Event

The belief potential of event is useful to know whether an action pair would be played contagiously once it is played in  $E$ ; however, it provides us with little information about the process of contagion. To see this, consider the following two information structures.

First, recall the information structure given in Section 4.1:  $\Theta = \{1, 2, 3, \dots\}$ ,  $\Pi_1 = \{\{1\}, \{2, 3\}, \{3, 4\}, \dots\}$  and  $\Pi_2 = \{\{1, 2\}, \{3, 4\}, \{5, 6\}, \dots\}$ , and  $P(n) = \varepsilon(1 - \varepsilon)^{n-1}$ . Let  $E = \{1\}$ . Then the belief potential of event  $E$  can be calculated as  $1/(2 - \varepsilon)$ . Next, consider a slightly different information structure in which all elements but  $P(3)$  and  $P(4)$  are defined same as before. Let  $P(3) = \varepsilon(1 - \varepsilon)^2/2$  and  $P(4) = \varepsilon(1 - \varepsilon)^3 + \varepsilon(1 - \varepsilon)^2/2$ . Table 3 shows the conditional probability  $P(\{n\} \mid \{n, n + 1\})$  for each  $n \in \mathbb{N}$ . Since  $P(3)/(P(3) +$

player 1	1	2	3	4	5	6	7	...
player 2								

Table 3: Partitions and posteriors of the second example

$P(4)) = 1/2(2 - \varepsilon)$ , the belief potential of event  $E$  is now given by  $1/2(2 - \varepsilon)$ . But the conditional probability  $P(\{n\} \mid \{n, n + 1\})$  is same for any  $n \geq 5$  and  $i = 1, 2$ . In this sense, the marginal process of contagion seems to be unchanged between two information structures for any  $n \geq 5$ . To be more precise, let us introduce the marginal belief potential defined as follows.

**Definition 9.** Given event  $E \in \mathcal{F}_1 \oplus \mathcal{F}_2$  and  $\sigma_i(E)$ , the marginal belief potential of event  $E$  for player  $i$  is defined by  $\xi_{(i,n)}^p(E) \equiv \sup\{q \in [0, 1] : B_i^q(B_j^p[C_i^p]^{n-1}(E)) \cap [C_i^p]^n(E) \neq \emptyset\}$  for any  $p \in [0, \sigma_i(E)]$ ,  $n \in \mathbb{N}$ , and  $i = 1, 2$ . For our later use, let  $\xi_i^p(E) = \sup_n \xi_{(i,n)}^p(E)$ .

**Remark 5.**  $\xi_{(i,n)}^p(E)$  is well defined since  $p \in \{q \in [0, 1] : B_i^q(B_j^p[C_i^p]^{n-1}(E)) \cap [C_i^p]^n(E) \neq \emptyset\}$  for any  $p \in [0, \sigma_i(E)]$ ,  $n \in \mathbb{N}$ , and  $i = 1, 2$ .



**Remark 6.**  $\xi_{(i,n)}^p(E) \geq \sigma_i(E)$  holds for any  $p \in [0, \sigma_i(E)]$  and  $i = 1, 2$ .

In the first example, we have  $\xi_{(i,n)}^p(E) = 1/(2 - \varepsilon)$  for any  $p \in [0, 1/(2 - \varepsilon)]$ ,  $n \in \mathbb{N}$ , and  $i = 1, 2$ . On the other hand,  $\xi_{(1,1)}^p(E) = 2/(3 - \varepsilon)$ ,  $\xi_{(1,2)}^p(E) = (3 - 2\varepsilon)/(2\varepsilon^2 - 6\varepsilon + 5)$ ,  $\xi_{(1,3)}^p(E) = 1/(2 - \varepsilon)$ , ... , and  $\xi_{(2,1)}^p(E) = 1/(2 - \varepsilon)$ ,  $\xi_{(2,2)}^p(E) = 1/2(2 - \varepsilon)$ , and  $\xi_{(2,3)}^p(E) = 1/(2 - \varepsilon)$ , ... for any  $p \in [0, 1/2(2 - \varepsilon)]$  in the second example. Thus, the marginal belief potential yields more detailed information about the generating process of the belief hierarchies.

## A.2 Contagion under Limited Depth of Reasoning

Consider an incomplete information game with DR space  $\mathcal{U}$  such that, at any  $\theta \in \Theta$ ,  $a^*$  is *strict*  $p$ -dominant and  $\bar{a}$  is *strict*  $(1 - p)$ -dominant. Suppose  $\bar{a}$  is played by any levels of both players in some finite event  $E \in \mathcal{F}_1 \oplus \mathcal{F}_2$ . The following Proposition 3 shows that, even if  $p < \sigma(E)$ ,  $a^*$  cannot be played contagiously provided that level-0 players play  $\bar{a}$  in  $\Theta \setminus E$  and player's belief about other player's reasoning level satisfies a certain condition. Our proof proceeds as follows: Let  $E_{(i,0)} = E$  for any  $i = 1, 2$  and define inductively  $E_{(i,k_i)} = \{\theta \in \Theta : \sum_{t=0}^{k_i-1} \mu_i(k_i)(t)P(E_{(j,t)} | \pi_i(\theta)) \geq p\} \cup E$ . Note that in this setting, for player  $i$  with level  $k_i$ ,  $\bar{a}_i$  is a *unique* (interim) best response in  $\Theta \setminus E_{(i,k_i)}$ . We show that under a certain condition on players' decision rules, the event  $E_{(i,k_i)}$  stops increasing at some  $k_i = k$  for any  $i = 1, 2$ . Then in any Bayesian Prediction of  $\mathcal{U}$ , other actions than  $\bar{a}_i$  is played at at most finitely many states.

**Proposition 3.** *Suppose  $p \in (0, \sigma(E)]$  and  $s_i(0, \theta) = \bar{a}_i$  for any  $\theta \in \Theta \setminus E$  and  $i = 1, 2$ . If, for any  $i = 1, 2$ , there exists a  $n \in \mathbb{N}$  such that  $\sum_{t=2n-1}^{k-1} \mu_i(k)(t) < p/\xi_i^p(E)$  for any  $k \geq 2n$ , then  $a^*$  cannot be played contagiously under  $(\mu_i)_{i=1,2}$ .*

*Proof.* See Section A.4. □

**Remark 7.** For a mixture of the Level- $k$  and Cognitive Hierarchy type beliefs, this assumption is satisfied as long as  $\alpha < p/\xi_i^p(E)$ .

## A.3 Application: Rubinstein's (1989) Email Game

Consider a version of email game as in Strzalecki (2010). Facing an enemy, each player must choose an action either *Attack* or *Not Attack*, and there are two possibilities: the enemy is strong or weak. For each case, a payoff matrix is given by

	<i>Attack</i>	<i>Not Attack</i>			<i>Attack</i>	<i>Not Attack</i>	
<i>Attack</i>	-2, -2	-2, 0			<i>Attack</i>	1, 1	-2, 0
<i>Not Attack</i>	0, -2	0, 0			<i>Not Attack</i>	0, -2	0, 0
	(a) When enemy is strong				(b) When enemy is weak		

Table 4: Payoffs in the email game.

The information structure is specified as follows: The enemy is strong or weak equally likely. Only player 1 knows the strength of his enemy and send an email to let player 2 know whether the enemy is strong or not. If each player gets an email, he/she will send a confirmation. But each email gets lost with small probability, say  $\varepsilon > 0$ , and he/she can not distinguish whether his/her email did not reach the other player or the other player's email has not delivered. Thus, their information partitions and posteriors are given by

$\Pi_1$	1	$\frac{1}{2-\varepsilon}$	$\frac{1-\varepsilon}{2-\varepsilon}$	$\frac{1}{2-\varepsilon}$	$\frac{1-\varepsilon}{2-\varepsilon}$	...
$\Theta$	(0, 0)	(1, 0)	(1, 1)	(2, 1)	(2, 2)	...
$\Pi_2$	$\frac{1}{1+\varepsilon}$	$\frac{\varepsilon}{1+\varepsilon}$	$\frac{1}{2-\varepsilon}$	$\frac{1-\varepsilon}{2-\varepsilon}$	...	

Table 5: Partitions and posteriors of the email game.

Note that choosing *Not Attack* is dominant for player 1 if the enemy is strong, and (*Not Attack*, *Not Attack*) is a 1/3-dominant action pair if the enemy is weak. Since the belief potential of event  $\{(0, 0)\}$  is  $1/(2 - \varepsilon)$ , *Not Attack* is played by both players no matter how many emails they exchange in any equilibrium of this game (Rubinstein 1989). Let us introduce DR space into this game, and assume that level-0 players choose *Attack* at any states other than  $(0, 0)$ .<sup>17</sup> We show that, under a certain condition on player's belief, there exists a number  $N$  such that after receiving  $N$  messages, any levels of both players start choosing *Attack*. Before stating this result, we need to provide the following lemma.

**Lemma 3.** *Take any constant  $M (> 0)$ . For any  $i = 1, 2$ , there exists a  $n \in \mathbb{N}$  such that  $\sum_{t=2n-1}^{k-1} \mu_i(k)(t) < M$  for any  $k \geq 2n$  if and only if  $\inf_{(m)} \mu_i(k^m)(\{k^j \geq k^{m-1}\}) < M$  for any strictly increasing sequence  $(k^m) \in \mathbb{N}^\infty$*

*Proof.* See Section A.4. □

**Corollary 2.** (Theorem 4 in Strzalecki 2010)

*If  $\inf_{(m)} \mu_i(k^m)(k^{m-1} \leq k_j) < (2-\varepsilon)/3$  for any strictly increasing sequence  $(k^m) \in \mathbb{N}^\infty$  and  $i = 1, 2$ , there exists a number of messages, say  $N$ , such that after receiving  $N$  messages all levels of both players choose *Attack*.*

*Proof.* Define the event  $E = \{(0, 0)\}$ . Then we have (1)  $\sigma(E) = 1/(2 - \varepsilon)$ ; (2) *Not Attack* is 1/3-dominant; (3)  $\xi_1^{1/3}(E) = \xi_2^{1/3}(E) = 1/(2 - \varepsilon)$ . Then by Lemma 3 and Proposition 3, there exists a  $T \in \mathbb{N}$  such that  $E_{(i, k_i)} \subseteq \widehat{E}_{(i, T)}$  for any  $k_i \in K_i$  and  $i = 1, 2$ . That is, there exists a number of messages, say  $N_i$ , such that after receiving  $N_i$  messages player  $i$  chooses *Attack* for any  $k_i \in K_i$ . Therefore, letting  $N = \max\{N_1, N_2\}$  yields that after receiving  $N$  messages any levels of both players choose *Attack*. □

---

<sup>17</sup>Same result follows if we assume level-0 players choose *Attack* after receiving a certain number of messages.

## A.4 Remaining Proofs of Appendix A

### Proof of Proposition 3

*Proof.* Let  $E_{(i,0)} = E$  for any  $i = 1, 2$ . Construct the events  $(E_{(i,k_i)})_{i=1,2}^{k_i}$  as we mentioned. Then, for any player  $i = 1, 2$  with level  $k_i$ , playing  $\bar{a}_i$  is his unique best response for any  $\theta \in \Theta \setminus E_{(i,k_i)}$ . Thus it suffices to show that there exists some finite event  $F$  such that  $E_{(i,k_i)} \subseteq F$  for any  $k_i \in K_i$  and  $i = 1, 2$ . Recall that  $\widehat{E}_{(i,2n-1)} = B_i^p([C_j^p]^{n-1}(E) \cup E$  and  $\widehat{E}_{(i,2n)} = [C_i^p]^n(E)$  for any  $n \in \mathbb{N}$  and  $i = 1, 2$ . Note that  $[C_i^p]^n(E)$  is finite for any  $n \in \mathbb{Z}_+$  and  $i = 1, 2$  by our assumption, so that  $\widehat{E}_{(i,k_i)} < \infty$  for any  $k_i \in K_i$  and  $i = 1, 2$ . Since  $p \leq \sigma(E)$ ,  $\widehat{E}_{(i,k_i)}$  is strictly increasing in  $k_i$  for all  $k_i \in K_i$  and  $i = 1, 2$ .

Claim 1: For any  $i = 1, 2$ , there exists a  $K_i \in \mathbb{N}$  such that, for all  $N \geq K_i$ , we have  $P(\widehat{E}_{(j,2n-2)} \mid \pi_i(\theta)) = 0$  for any  $\theta \in \widehat{E}_{(i,2N)} \setminus \widehat{E}_{(i,2N-1)}$ .

$\therefore$ ) Suppose not. Then  $\exists i \in \{1, 2\}, \forall K_i \in \mathbb{N}, \exists N \geq K_i$  such that  $P(\widehat{E}_{(j,2n-2)} \mid \pi_i(\theta_N)) > 0$  for some  $\theta_N \in \widehat{E}_{(i,2N)} \setminus \widehat{E}_{(i,2N-1)}$ . Define inductively  $N_1 = \min\{N \geq 1 : P(\widehat{E}_{(j,2n-2)} \mid \pi_i(\theta_N)) > 0 \text{ for some } \theta_N \in \widehat{E}_{(i,2N)} \setminus \widehat{E}_{(i,2N-1)}\}$ ,  $N_2 = \min\{N \geq N_1 + 1 : P(\widehat{E}_{(j,2n-2)} \mid \pi_i(\theta_N)) > 0 \text{ for some } \theta_N \in \widehat{E}_{(i,2N)} \setminus \widehat{E}_{(i,2N-1)}\}, \dots$ . This infinite sequence  $(N_m)$  is well defined by our assumption, and  $N_m \neq N_l$  if  $m \neq l$  by construction. This implies that  $\theta_{N_m} \neq \theta_{N_l}$  and  $\pi_i(\theta_{N_m}) \cap \pi_i(\theta_{N_l}) = \emptyset$  for any  $\theta_{N_m} \in \widehat{E}_{(i,2N_m)} \setminus \widehat{E}_{(i,2N_m-1)}$  and  $\theta_{N_l} \in \widehat{E}_{(i,2N_l)} \setminus \widehat{E}_{(i,2N_l-1)}$  with  $m \neq l$ . But then, since  $P(\widehat{E}_{(j,2n-2)} \mid \pi_i(\theta_{N_m})) > 0$  for all  $m \in \mathbb{N}$ , we must have  $|\widehat{E}_{(j,2n-2)}| = \infty$ . On the other hand, since  $|E| < \infty$  and  $|\pi_i| < \infty$  for all  $\pi_i \in \Pi_i$  and  $i = 1, 2$ , we have  $|\widehat{E}_{(j,2n-2)}| < \infty$ , contradicting.  $\square$

Define  $K = \max\{K_1, K_2\}$ . Then we have, for any  $i = 1, 2$  and  $N \geq K$ ,  $P(\widehat{E}_{(j,2n-2)} \mid \pi_i(\theta)) = 0$  for any  $\theta \in \widehat{E}_{(i,2N)} \setminus \widehat{E}_{(i,2N-1)}$ .

Claim 2:  $P(\widehat{E}_{(j,2K-1)} \mid \pi_i(\theta)) \leq \xi_i^p(E)$  for any  $\theta \in \widehat{E}_{(i,2K)} \setminus \widehat{E}_{(i,2K-1)}$ .

$\therefore$ ) Suppose not. Note that  $P(\widehat{E}_{(j,2K-1)} \mid \pi_i(\theta)) = P(B_j^p[C_i^p]^{K-1}(E) \mid \pi_i(\theta))$  by definition. Hence, if we have  $P(\widehat{E}_{(j,2K-1)} \mid \pi_i(\theta)) > \xi_i^p(E)$  for some  $\theta \in \widehat{E}_{(i,2K)} \setminus \widehat{E}_{(i,2K-1)}$ , then  $P(B_j^p[C_i^p]^{K-1}(E) \mid \pi_i(\theta)) > \xi_i^p(E)$ . Thus, there exists a  $\delta > 0$  such that  $P(B_j^p[C_i^p]^{K-1}(E) \mid \pi_i(\theta)) > \xi_i^p(E) + \delta$ . Hence, we have  $\theta \in B_i^{\xi_i^p(E) + \delta}(B_j^p[C_i^p]^{K-1}(E))$  and  $\theta \in \widehat{E}_{(i,2K)} = [C_j^p]^K(E)$ . That is,  $\xi_{(i,K)}^p(E) \geq \xi_i^p(E) + \delta$ , contradicting the definition of  $\xi_i^p(E)$ .  $\square$

Take any  $\theta \in \widehat{E}_{(i,2K)} \setminus \widehat{E}_{(i,2K-1)}$ . Remember that since  $\widehat{E}_{(i,k_i)}$  is increasing in  $k_i$  and constitutes an upper bound of  $E_{(i,k)}$ , we must have  $P(E_{(j,k)} \mid \pi_i(\theta)) = 0$  for any  $k \leq 2n - 2$  by Claim 1. Then, by Claim 1 and 2,

$$\begin{aligned} \sum_{t=0}^{2K-1} \mu_i(2K)(t)P(E_{(j,t)} \mid \pi_i(\theta)) &= \sum_{t=2n-1}^{2K-1} \mu_i(2K)(t)P(E_{(j,t)} \mid \pi_i(\theta)) \\ &\leq \sum_{t=2n-1}^{2K-1} \mu_i(2K)(t)P(\widehat{E}_{(j,2K-1)} \mid \pi_i(\theta)) \\ &< \frac{p}{\xi_i^p(E)} \cdot \xi_i^p(E) = p. \end{aligned}$$

Hence,  $\theta \notin E_{(i,2K)}$ , so that we can conclude  $E_{(i,2K)} \subseteq \widehat{E}_{(i,2K-1)}$  for  $i = 1, 2$ .

Claim 3:  $E_{(i,N)} \subseteq \widehat{E}_{(i,2K-1)}$  for any  $N \geq 2K$  and  $i = 1, 2$ .

$\therefore$ ) We show this by induction. We have just shown this for the case of  $N = 2K$ . Assume that  $E_{(i,N)} \subseteq \widehat{E}_{(i,2K-1)}$  for any  $2K \leq N \leq M$  and  $i = 1, 2$ . Suppose, in negation, that there exists a  $\bar{\theta} \in E_{(i,M+1)} \setminus \widehat{E}_{(i,2K-1)}$ . Observe that, by our induction hypothesis,  $\bar{\theta} \in E_{(i,M+1)} \subseteq B_i^q(\widehat{E}_{(j,2K-1)}) \cup E = \widehat{E}_{(i,2K)}$ . Hence,  $\bar{\theta} \in \widehat{E}_{(i,2K)} \setminus \widehat{E}_{(i,2K-1)}$ . But then by the previous argument, we have  $P(B_j^q[C_i^q]^{K-1}(E) \mid \pi_i(\bar{\theta})) \leq \xi_i^p(E)$  and  $P(E_{(j,k)} \mid \pi_i(\bar{\theta})) = 0$  for any  $k \leq 2n - 2$ . Hence,

$$\begin{aligned} \sum_{t=0}^M \mu_i(M+1)(t)P(E_{(j,t)} \mid \pi_i(\bar{\theta})) &= \sum_{t=2n-1}^M \mu_i(M+1)(t)P(E_{(j,t)} \mid \pi_i(\bar{\theta})) \\ &\leq \sum_{t=2n-1}^M \mu_i(M+1)(t)P(\widehat{E}_{(j,2K-1)} \mid \pi_i(\bar{\theta})) \\ &< \frac{p}{\xi_i^p(E)} \cdot \xi_i^p(E) = p. \end{aligned}$$

Thus, we have  $\bar{\theta} \notin E_{(i,M+1)}$ , a contradiction.  $\square$

By Claim 3, we have  $E_{(i,k_i)} \subseteq \widehat{E}_{(i,2K-1)}$  for any  $k_i \in K_i$  and  $i = 1, 2$ . Thus, letting  $F = \bigcup_{i=1,2} \widehat{E}_{(i,2K-1)}$  yields the desired result.  $\square$

### Proof of Lemma 3

*Proof.* (If part) To derive a contradiction, suppose,  $\forall n \in \mathbb{N}, \exists k \geq 2n, \sum_{t=2n-1}^{k-1} \mu_i(k)(t) \geq M$ . Inductively define a sequence  $(k^m) \in \mathbb{N}^\infty$  as follows:  $k^1 = \min\{k \geq 2 \mid \sum_{t=1}^{k-1} \mu_i(k)(t) \geq M\}$ ,  $k^2 = \min\{k \geq 2k^1 \mid \sum_{t=2k^1-1}^{k-1} \mu_i(k)(t) \geq M\}$ ,  $k^3 = \min\{k \geq 2k^2 \mid \sum_{t=2k^2-1}^{k-1} \mu_i(k)(t) \geq M\}$ , ... This  $(k^m)$  is well defined by our hypothesis and strictly increasing by its construction. But then  $\mu_i(k^m)(\{k^j \geq k^{m-1}\}) \geq \sum_{t=2k^{(m-1)-1}}^{k^m-1} \mu_i(k)(t) \geq M$  for any  $m \in \mathbb{N}$ , contradicting.

(Only if part) Suppose  $\exists n \in \mathbb{N}$  such that  $\sum_{t=2n-1}^{k-1} \mu_i(k)(t) < M$  for any  $k \geq 2n$ . Take any strictly increasing sequence  $(k^m) \in \mathbb{N}^\infty$ . By definition,  $\exists m'$  such that  $k^{m'-1} \geq 2n - 1$ . Hence,  $\sum_{t=2n-1}^{k^{m'}-1} \mu_i(k^{m'})(t) < M$ , so that  $\sum_{t=k^{m'}-1}^{k^{m'}-1} \mu_i(k^{m'})(t) < M$ . That is,  $\mu_i(k^{m'})(\{k^j \geq k^{m'-1}\}) < M$ .  $\square$

## Appendix B: Omitted Proofs

### Proof of Lemma 2

*Proof.* We show this by induction. First,  $E_{(i,0)} = \widehat{E}_{(i,0)}$  for any player  $i = 1, 2$  by definition. Next, suppose that for any  $i = 1, 2$ , we have  $E_{(i,k_i)} \subseteq \widehat{E}_{(i,k_i)}$  for all  $0 \leq k_i \leq k \in \mathbb{N}$ . Then

since  $E_{(j,k_j)} \subseteq \widehat{E}_{(j,k_j)}$  for any  $0 \leq k_j \leq k$ ,

$$\begin{aligned}
E_{(i,k+1)} &= \left\{ \theta \in \Theta : \sum_{t=0}^k \mu_i(k+1)(t) P(E_{(j,t)} \mid \pi_i(\theta)) \geq 1-p \right\} \cup E \\
&\subseteq \left\{ \theta \in \Theta : \sum_{t=0}^k \mu_i(k+1)(t) P(\widehat{E}_{(j,t)} \mid \pi_i(\theta)) \geq 1-p \right\} \cup E \\
&\subseteq \left\{ \theta \in \Theta : P(\widehat{E}_{(j,k)} \mid \pi_i(\theta)) \geq 1-p \right\} \cup E \\
&= \widehat{E}_{(i,k+1)}.
\end{aligned}$$

The third inequality follows since  $\widehat{E}_{(i,k_i)}$  is increasing in  $k_i$  for any  $i = 1, 2$ . By our induction hypothesis the result follows.  $\square$

## Proof of Proposition 2

*Proof.* Our proof is by construction. Consider the class of games in which there exists a  $(1-p)$ -dominant prediction with  $p \geq 1/2$ , and both players have the Level- $k$  type belief. For notational simplicity, let  $q = 1-p$ . Let us denote a  $p$ -dominant prediction by  $s^*$ , and a  $q$ -dominant prediction by  $\bar{s}$ . Also, let us denote the corresponding  $p$ -dominant action pair by  $a^*$ , and the corresponding  $q$ -dominant action pair by  $\bar{a}$  respectively. Take any  $\varepsilon'$ -elaboration  $\mathcal{U} \in E(\mathcal{G}, \varepsilon')$  as in the example in Section 4.1.<sup>18</sup> By our choice of  $\varepsilon'$ -elaboration, the conditional probability  $P(\{n\} \mid \{n, n+1\}) = 1/(2-\varepsilon) > 1/2$  for any  $n \geq 1$ . Hence, level  $k$  of player 1 plays  $\bar{a}_1$  in  $\{2n, 2n+1\}$  if level  $k-1$  of player 2 plays  $\bar{a}_2$  in  $\{2n-1, 2n\}$ . Thus, the set of states  $\bar{E}_{(1,k)}$  in which playing  $\bar{a}_1$  is a unique best response for level  $k$  of player 1 is given by  $\bar{E}_{(1,2k)} = \bar{E}_{(1,2k+1)} = \{1, 2, \dots, 2k+1\}$  for any  $k \in K_i$ . Hence,  $\bar{E}_{(1,k)} \uparrow \Theta$  as  $k \rightarrow \infty$ , so that every prediction action distribution profile  $\beta$  in  $\mathcal{U}$  satisfies,  $\inf_{\mathbf{k} \in K} \beta_{\mathbf{k}}(a^*) = 0$ . Since the prediction action distribution profile  $\alpha$  induced by  $s^*$  satisfies  $\alpha_{\mathbf{k}}(a^*) = 1$  for any  $\mathbf{k} \in K$ , we have  $\|\alpha - \beta\| = 1$ , so that  $s^*$  is not robust.  $\square$

## Proof of Theorem 1

*Proof.* By Proposition 1, it suffices to show the case where  $p \in [1/2, 1)$ . For notational convenience, let  $q = 1-p$ , then  $q \in (0, 1/2]$ . Suppose there exists a  $n \in \mathbb{N}$  such that  $\sup_{k \geq 2n} \sum_{t=2n-1}^{k-1} \mu_i(k)(t) < 2(1-p)$  for any  $i = 1, 2$ . Then there exists a  $\omega > 0$  such that  $\sum_{t=2n-1}^{k-1} \mu_i(k)(t) < (2-\omega)q$  for any  $k \geq 2n$ . Define  $\psi = \omega q / (4-\omega) \{1 - (2-\omega)q\}$ , and  $\sigma = 2/(4-\omega)$ . It is easy to check that  $0 < \psi < 1$  always holds, and we have  $\psi < \sigma$ . Take any  $\delta > 0$ , and let

$$\varepsilon = \frac{\delta \psi (2\sigma - 1)}{4\sigma(1+\psi) \sum_{k=0}^{2n-2} \left(\frac{1-q}{q}\right)^k}.$$

<sup>18</sup>That is,  $\Theta = \{1, 2, 3, \dots\}$ ,  $\Pi_1 = \{\{1\}, \{2, 3\}, \{3, 4\}, \dots\}$  and  $\Pi_2 = \{\{1, 2\}, \{3, 4\}, \{5, 6\}, \dots\}$ , and  $P(n) = \varepsilon(1-\varepsilon)^{n-1}$  for any  $n \in \mathbb{N}$ . For any player  $i = 1, 2$ ,  $\bar{a}_i$  is strictly dominant for any levels and level-0 of player  $i$  plays  $\bar{a}_i$  in  $E = \{1\}$ . Let  $\varepsilon' = \varepsilon(2-\varepsilon)$ .

Take any  $\mathcal{U} \in E(\mathcal{G}, \varepsilon)$ . We construct a collection of events which gives an upper bound for the ex-ante probability of  $(E_{(i,k_i)})_{i=1,2}^{k_i}$ . First, define  $H = \bigcup_{i=1,2} \{B_i^\psi(\widehat{E}_{(j,2n-2)}) \cup \widehat{E}_{(i,2n-2)}\}$ . Observe that by Lemma A.4. of Oyama and Tercieux (2012), we have

$$P(H) \leq 2\varepsilon \left(1 + \frac{1}{\psi}\right) \sum_{k=0}^{2n-2} \left(\frac{1-q}{q}\right)^k. \text{ }^{19}$$

Next, let us inductively define  $F_{(1,0)} = F_{(2,0)} = H$  and  $F_{(i,k)} = B_i^\sigma(F_{(j,k-1)}) \cup H$  for any  $k \in \mathbb{N}$  and  $i = 1, 2$ . Straight forward calculation yields that  $F_{(i,2n)} = [C_i^\sigma]^n(H)$  for any  $n \in \mathbb{Z}_+$ .

Claim:  $E_{(i,2n-2+k)} \subseteq F_{(i,k)}$  for any  $k \in \mathbb{Z}_+$  and  $i = 1, 2$ .

∴) We show this by induction. By definition of  $H$ , we have  $E_{(i,2n-2)} \subseteq F_{(i,0)}$  for any  $i = 1, 2$ . Suppose  $E_{(i,2n-2+k)} \subseteq F_{(i,k)}$  for any  $0 \leq k \leq m$  and  $i = 1, 2$ . We want to show  $E_{(i,2n-1+m)} \subseteq F_{(i,m+1)}$  for any  $i = 1, 2$ . Suppose, in negation, that there exists a  $\theta \in E_{(i,2n-1+m)} \setminus F_{(i,m+1)}$  for some  $i = 1, 2$ . Then we have

$$\begin{aligned} & \sum_{t=0}^{2n-2+m} \mu_i(2n-1+m)(t) P(E_{(j,t)} \mid \pi_i(\theta)) \\ = & \sum_{t=0}^{2n-2} \mu_i(2n-1+m)(t) P(E_{(j,t)} \mid \pi_i(\theta)) + \sum_{t=2n-1}^{2n-2+m} \mu_i(2n-1+m)(t) P(E_{(j,t)} \mid \pi_i(\theta)) \\ \leq & \sum_{t=0}^{2n-2} \mu_i(2n-1+m)(t) P(\widehat{E}_{(j,2n-2)} \mid \pi_i(\theta)) + \sum_{t=2n-1}^{2n-2+m} \mu_i(2n-1+m)(t) P(F_{(j,m)} \mid \pi_i(\theta)) \\ < & [1 - (2-\omega)q] \cdot \psi + (2-\omega)q \cdot \sigma = q. \end{aligned}$$

The third inequality follows since  $\theta \notin B_i^\psi(\widehat{E}_{(j,2n-2)})$ ,  $\theta \notin B_i^\sigma(F_{(j,m)}) \subseteq F_{(i,m+1)}$ , and  $(F_{(j,k)})$  is increasing in  $k$ . Hence,  $\theta \notin E_{(i,2n-1+m)}$ , a contradiction.  $\square$

Since  $F_{(i,k)}$  is increasing in  $k$  and  $\sigma > 1/2$ , the above claim and Lemma A.4. of Oyama and Tercieux (2012) imply  $P(E_{(i,2n-2+k)}) \leq P(F_{(i,k)}) \leq P(H) \cdot \sigma / (2\sigma - 1)$  for any  $k \in \mathbb{Z}_+$  and  $i = 1, 2$ . But then

$$P(E_{(i,k_i)}) \leq P(F_{(i,k_i)}) < \frac{2\sigma\varepsilon}{2\sigma-1} \left(1 + \frac{1}{\psi}\right) \sum_{k=0}^{2n-2} \left(\frac{1-q}{q}\right)^k = \frac{\delta}{2}$$

for any  $k_i \in K_i$  and  $i = 1, 2$ . By Lemma 1, there exists a Bayesian prediction  $s'$  of  $\mathcal{U}$  that satisfies  $P(\{\theta \in \Theta \mid (s'_1(k_1, \theta), s'_2(k_2, \theta)) \neq (a_1^*, a_2^*)\}) < \delta$  for any  $\mathbf{k} \in \mathbf{K}$ . Therefore, we are done.  $\square$

## References

- [1] Agranov, M., Potamites, E., Schotter, A., and Tergiman, C. (2012). “Beliefs and endogenous cognitive levels: An experimental study.” *Games and Economic Behavior*, 75, 449-463.

<sup>19</sup>Here we use the fact that  $P(B_i^p)(E) \leq P(E)/p$  for any  $0 < p < 1$  and  $i = 1, 2$ .

- [2] Arad, A., and Rubinstein, A. (2010). “The 11-20 money request game: Evaluating the upper bound of k-level reasoning.” *The American Economic Review*, forthcoming.
- [3] Burchardi, K., and Penczynski, S. (2010). “Out of your mind: Eliciting individual reasoning in one shot games.” mimeo.
- [4] Camerer, C. F. (2003). “*Behavioral Game Theory: Experiments in Strategic Interaction*.” Princeton University Press, Princeton, 2002.
- [5] Camerer, C. F., T.-H. Ho, and J.-K. Chong. (2004). “A Cognitive Hierarchy Model of Games.” *The Quarterly Journal of Economics*, 119, 861-898.
- [6] Crawford, V. P., Costa-Gomes, M. A., and Iriberri, N. (2012). “Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications.” *Journal of Economic Literature*, forthcoming.
- [7] Crawford, V. P., and Iriberri, N. (2007). “Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner’s Curse and Overbidding in Private Value Auctions?” *Econometrica*, 75, 1721-1770.
- [8] Crawford, V. P., Kugler, T., Neeman, Z., and Pauzner, A. (2009). “Behaviorally optimal auction design: Examples and observations.” *Journal of the European Economic Association*, 7, 377-387.
- [9] Heinemann, F., Nagel, R., and Ockenfels, P. (2004). “The theory of global games on test: Experimental analysis of coordination games with public and private information.” *Econometrica*, 72, 1583-1599.
- [10] Kajii, A., Morris, S. (1997). “The Robustness of Equilibria to Incomplete Information.” *Econometrica*, 65, 1283-1309.
- [11] Kets, W., and Heifetz, A. (2012). “All types naive and canny.” mimeo.
- [12] Kneeland, T. (2012). “Coordination under Limited Depth of Reasoning.” mimeo.
- [13] Monderer, D., Samet, D. (1989). “Approximating Common Knowledge with Common Beliefs.” *Games and Economic Behavior*, 1, 170-190.
- [14] Morris, S., Rob, R., and Shin, H.S. (1995). “p-Dominance and belief potential.” *Econometrica*, 63, 145-157.
- [15] Nagel, R. (1995). “Unraveling in guessing games: An experimental study.” *The American Economic Review*, 85, 1313-1326.
- [16] Oyama, D., Tercieux, O. (2012). “On the strategic impact of an event under non-common priors.” *Games and Economic Behavior*, 74, 321-331.

- [17] Penczynski, S. P. (2011). "Strategic Thinking: The Influence of the Game." Working Paper.
- [18] Rubinstein, A. (1989). "The Electronic Mail Game: Strategic Behavior Under "Almost Common Knowledge." *The American Economic Review*, 79, 385-391.
- [19] Stahl, D. O., and Wilson, P. W. (1995). "On Players' Models of Other Players: Theory and Experimental Evidence." *Games and Economic Behavior*, 10, 218-254.
- [20] Strzalecki, T. (2010). "Depth of Reasoning and Higher Order Belief." mimeo.
- [21] Ui, T. (2001). "Robust equilibria of potential games." *Econometrica*, 69, 1373-1380.