Robust Predictions under Finite Depth of Reasoning

Kota Murayama*

April, 2013

Abstract

This paper elucidates how predictions in finite depth of reasoning model, special cases of which include level-k and cognitive-hierarchy models are robust to the common knowledge assumption of level-0 types' actions. We examine whether an outside observer can ignore a slight violation of the common knowledge assumption when she knows which game players will play with high probability but, with small probability, she does not know and they will play different games with respect to payoff structure and actions of level-0 types. A sufficient condition is provided for a **p**-dominant cognitive equilibrium, which is a natural analogue of **p**-dominant action profile in the finite depth of reasoning model, being robust to incomplete information *á la* Kajii and Morris (1997). Under a certain condition on players' beliefs, we show that even a **p**-dominant cognitive equilibrium with $\sum_{i\in\mathcal{I}} p_i \geq 1$ can be robust. As a by-product, this result implies that the level-k model has the smallest set of robust equilibria, and the cognitive-hierarchy model has the largest regarding **p**-dominance.

Keywords: Robustness, finite depth of reasoning, level-k model, cognitive-hierarchy model, equilibrium refinement.

*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, Takeshi Nishimura, Daisuke Oyama, and Satoru Takahashi for time, advice, and encouragement. I also thank the seminar participants of Contract Theory Workshop East and 10th Game Theory Workshop at Hitotsubashi University. Of course, all remaining errors are mine.

1 Introduction

Standard game theory is often blamed for its low descriptive power. A leading example is Rubinstein's (1989) email game. In the email game, players can 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-standard game theoretical models is *finite depth of reasoning model*, special cases of which include level-k and cognitive-hierarchy models (e.g., Nagel 1995; Stahl and Wilson 1995; Camerer et al. 2004).[1] In the finite depth of reasoning model, players follow a 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, level-1 types believe that his opponents are level-0 types for sure.[2] The model's prediction called *cognitive equilibirum* is obtained through the following procedure: First, specify level-0 types' actions and assume that they are common knowledge. Then those actions work as an anchor for higher-level types, that is, level-1 type best responds to the actions of level-0 types, level-2 type best responds to the best responses of level-1 types and/or to the actions of level-0 types under his belief, and so on.

As it is clear from the definition, prediction in the finite depth of reasoning model crucially depends on what level-0 types will choose and how other higher-level types model level-0 types. In general, level-0 types are assumed to be naïve or non-strategic, and hence choose his action randomly or choose a salient action. For instance, Nagel (1995) uses the random level-0 type in a 2/3 guessing game. In hide-and-seek game where each hider (seeker) chooses one box to hide (seek) a treasure, Crawford and Iriberri (2007b) assume that level-0 types respond to labels (A or B) or locations (middle or end) of the boxes. Unfortunately, these reasonable specifications may not be so true as a recent experiment by Burchardi and Penczynski (2012) shows: level-0 types respond to the salient number (66) in the 2/3 guessing game, and their average number is significantly higher than 50. Moreover, as the complexity of games increases, it becomes more and more difficult to find a candidate for level-0 actions.[3] In the auction environment, Crawford and Iriberri (2007a) propose two possibilities for actions of level-0 types: either bid uniformly between

[1] For the comprehensive survey, see Crawford et al. (2012)

[2] A major distinction between the level-k and cognitive-hierarchy models is their way of specifying players' beliefs about other players' reasoning levels. In the level-k model, level-k_i type of player i surely believes that his opponents have a reasoning level exactly one lower than him; that is, $k_i - 1$. On the other hand, the cognitive-hierarchy model assumes that given his own level, players calculate the conditional probability of opponents' levels induced by some distribution (e.g., Poisson).

[3] Arad and Rubinstein (2012) emphasize the importance of simplicity of games for allowing players to have a clear candidate for level-0 actions.

the lowest and highest possible values or truthfully bid his own value. But then allowing several possibilities for level-0 actions inherently casts doubt on the common knowledge assumption, since the number of combinations between level-0 actions and (degenerate) beliefs about how other types predict level-0 actions increases exponentially.[4]

The purpose of this paper is to investigate how predictions in the finite depth of reasoning model would be affected if players *do not* commonly know level-0 actions. A pioneering paper by Strzalecki (2010) solves the paradoxical result of the email game by introducing finite depth of reasoning. He shows that if each player's belief about the other player's reasoning level satisfies a certain condition, then there exists a cognitive equilibrium in which both players choose cooperation after exchanging a certain number of emails. In other words, cognitive equilibrium is somewhat robust to the equilibrium selection through higher-order uncertainty *á la* Rubinstein (1989). Theoretically speaking, this paper explores a possibility if cognitive equilibrium is *generally robust* to such higher-order uncertainty without restricting our attention to the email game by using techniques proposed in Kajii and Morris (1997). However, by introducing higher-order uncertainty in level-0 actions, together with uncertainty in payoffs, this approach enables us to answer the question: "*how sensitive the predictions of finite depth of reasoning model are to the common knowledge assumption of level-0 actions*." That is, we consider the situation in which an outside analyst knows which game players will play including level-0 actions with high probability; however, with small probability, players can play a totally different game in terms of its payoff structure *and level-0 actions*. With small uncertainty in level-0 actions, we examine whether we can use a solution for the simplified game (i.e., ignore the unobserved uncertainty) as a prediction for the *real* game.

To be more specific, we formulate the finite depth of reasoning model as an incomplete information game with *cognitive type* due to Strzalecki (2010). In our model, each cognitive type assigns a reasoning level to each player, and decision rules such as level-k and cognitive-hierarchy are interpreted as beliefs about other players' cognitive types with satisfying that each player believes others have strictly lower cognitive types with probability 1.[5] Given level-0 actions, *cognitive equilibrium* is obtained through inductively calculating a best response for each cognitive type.[6] A cognitive equilibrium is said to be *robust to incomplete information* if the behavior of all cognitive types of each player which is close to the cognitive equilibrium constitutes a (Bayesian) cognitive equilibrium of every nearby incomplete information game.[7] The word "nearby" means that the sets of players and

[4]In fact, Penczynski (2011) experimentally finds that players do not share the belief on actions of level-0 types in the hide-and-seek game.

[5]Strzalecki's cognitive type space is more general than ours in the sense that his model allows two cognitive types with the same reasoning level to have different beliefs about other players' cognitive types.

[6]In this solution, players must have inconsistent beliefs, so that finite depth of reasoning model is often referred as a nonequilibrium analysis (Crawford et al. 2009, 2012; Crawford and Iriberri 2007). Here we follow the terminology in the theoretical literature.

[7]*Bayesian cognitive equilibrium* is a natural extension of cognitive equilibrium when we have extra uncertainty Θ with respect to payoffs. Formal definition will be given in Section 2.

actions are the same with original incomplete information game with cognitive type, and each player knows with high probability that his payoffs *and level-0 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 perturbed objects.

Among the sufficient conditions for KM robust equilibria, we focus on the sufficient condition that clarifies a relationship between players' higher order beliefs and $\mathbf{p}$-dominance.[8] Cognitive equilibrium is said to be $\boldsymbol{p}$*-dominant* if each player i's equilibrium strategy becomes a best response whenever he (perceptually) believes that the other players would follow the equilibrium strategy with probability at least p_i. Kajii and Morris (1997) show that if $\sum_{i\in\mathcal{I}} p_i < 1$, then a $\boldsymbol{p}$*-dominant action profile* is KM robust. In contrast, our main result states that if there exists a level n such that any higher-level type of players than n put sufficiently high probability on his opponents being lower-level types than n, then a $\boldsymbol{p}$*-dominant cognitive equilibrium* is robust even if $\sum_{i\in\mathcal{I}} p_i \geq 1$. Intuitively, this sufficient condition is satisfied if all the "thoughtful" people think that normal human beings can think at most n depths and that less fraction of smart people can think deeper than n. According to this result, two prominent finite depth of reasoning models show a great difference in terms of the set of robust equilibria: When players have level-k belief, a $\mathbf{p}$-dominant cognitive equilibrium becomes robust if $\sum_{i\in\mathcal{I}} p_i < 1$. In contrast, under cognitive-hierarchy belief, any $\mathbf{p}$-dominant cognitive equilibrium with $\mathbf{p} \in [0,1)^{\mathcal{I}}$ is robust.

The rest of the paper is organized as follows. Section 2 proposes our framework of the finite depth of reasoning model. Section 3 introduces two preliminary concepts, $\mathbf{p}$-dominance and common $\mathbf{p}$-belief. Section 4 derives a sufficient condition for a $\mathbf{p}$-dominant cognitive equilibrium being robust, and Section 5 discusses the other possible notions of robustness and of type spaces.

2 Framework

2.1 Incomplete Information Game with Cognitive Type

In this section, we formulate finite depth of reasoning model as an incomplete information game with cognitive type. A player in this game faces an uncertainty about other players' cognitive types (or reasoning levels interchangeably), and forms a belief on his opponents' cognitive types given his own type. Importantly, we assume that each player must believe with certain that the other players have strictly lower reasoning levels than him own and this is common knowledge among players. In the sequel, we follow the conventional notations such as $A = \times_i A_i$ and $A_{-i} = \times_{j\neq i} A_j$, and let $\Delta(A)$ denotes a collection of Borel probability measures on A. The finite depth of reasoning model is formally defined as follows: An incomplete information game with cognitive type denoted by $\mathcal{G}$ is given by

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

$\mathcal{G} = (\mathcal{I}, \{A_i, K_i, \mu_i, g_i\}_{i\in\mathcal{I}})$, where $\mathcal{I} = \{1, 2, ..., I\}$ is the finite set of players, and for each player $i \in \mathcal{I}$, A_i is the finite set of actions, $g_i : A \to \mathbb{R}$ is the (bounded) payoff function, $K_i = \mathbb{Z}_+$ is the set of cognitive types, and $\mu_i : K_i \to \Delta(K_{-i})$ is his (interim) belief about other players' cognitive types that satisfies $\mu_i(k_i)(\{k_{-i} \in K_{-i} : k_j < k_i \text{ for each } j \in \mathcal{I} \text{ with } j \neq i\}) = 1$ for any $k_i \in \mathbb{N}$.

Remark 2.1. We do not put any restriction on level-0 types' beliefs. Since their actions are determined outside of the model, their beliefs have no practical effect on our analysis.

Remark 2.2. By definition of our cognitive type space, we consider the situation where beliefs about other players' reasoning levels are common knowledge among players. Since the level-k and cognitive-hierarchy models also assume this, we lose little generality for our purpose. However, this assumption can be relaxed by considering Strzalecki's cognitive type space: see Section 5.

Remark 2.3. The restriction on μ_i requires that player i with level k_i (>0) certainly believes that other players have strictly lower levels than k_i. This strictness of inequality allows us to avoid fixed-point arguments like "I best respond to players who best respond to me", and to recursively derive a best response of each cognitive type given actions of level-0 types.

2.2 Solution Concept

We use a solution concept called *cognitive equilibrium* in which, given level-0 types' actions, each cognitive type chooses his best response under his belief about other players' cognitive types.[9] Let us denote player i's (pure) strategy in $\mathcal{G}$ by $s_i : K_i \to A_i$, and let S_i denote the set of player i's strategy for each $i \in \mathcal{I}$.[10]

Definition 2.1. A strategy profile s^* constitutes a *cognitive equilibrium* of $\mathcal{G}$ if, for any $i \in \mathcal{I}$, $k_i \in \mathbb{N}$, and $a_i \in A_i$,

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

Since we can freely specify level-0 actions, we have at least as many cognitive equilibria as the cardinality of A. In the experimental literature, this multiplicity problem is often solved by assigning a specific action to level-0 type and assuming players commonly know it. Then, generically in payoffs, we obtain a unique cognitive equilibrium.

Two typical finite depth of reasoning models, level-k and cognitive-hierarchy, are translated into our framework as follows.

[9] By construction, players must have mutually inconsistent beliefs in this equilibrium. Thus the word "equilibrium" here has a different meaning from that of Bayesian equilibrium.

[10] Just for notational simplicity, we do not allow randomization in strategies. But mixed strategies can be incorporated into our model in a standard way.

Example 2.1. (Level-k Model)
In the level-k model, each player believes that other players' levels are exactly one lower than his own level. That is, a player with level k_i believes that others have a level $k_i - 1$ with probability 1. Formal representation is as follows: for each $i \in \mathcal{I}$ and $k_i \in \mathbb{N}$,

$$\mu_i^{lk}(k_i)(k_{-i}) = \begin{cases} 1 & \text{if } k_j = k_i - 1, \forall j \neq i \\ 0 & \text{if otherwise} \end{cases}$$

Thus the level-k model is an incomplete information game with cognitive type denoted by $\mathcal{G}^{lk}$ in which players have this level-k belief, and its prediction is a cognitive equilibrium of $\mathcal{G}^{lk}$. We especially call the cognitive equilibrium of $\mathcal{G}^{lk}$ *level-k equilibrium.*

Example 2.2. (Cognitive-Hierarchy Model)
The cognitive-hierarchy model (CH model) assumes that given his own level, each player calculates conditional probabilities about other players' levels, which are induced by some distribution $\lambda \in \Delta(\mathbb{Z}_+)$. For instance, Camerer et al. (2004) assume λ follows Poisson distribution. Given this λ, each player's belief is constructed as follows: for each $i \in \mathcal{I}$ and $k_i \in \mathbb{N}$,

$$\mu_i^{ch}(k_i)(k_{-i}) = \begin{cases} \frac{\Pi_{j \neq i} \lambda(k_j)}{(\sum_{l=0}^{k_i - 1} \lambda(l))^{I-1}} & \text{if } k_j < k_i, \forall j \neq i \\ 0 & \text{if otherwise} \end{cases}$$

As in Example 2.1, the CH model is denoted by $\mathcal{G}^{ch}$, and its cognitive equilibrium is called *cognitive-hierarchy equilibrium* (CH equilibrium).

2.3 Embedding Incomplete Information Game with Cognitive Type

An embedding incomplete information game with cognitive type, $\mathcal{U}$, is given by $\mathcal{U} = (\mathcal{I}, \Theta, P, \{A_i, K_i, \mu_i, u_i, \Pi_i\}_{i \in \mathcal{I}})$, where Θ is a countable set of payoff states, P is a Borel probability measure on Θ, and for each player $i \in \mathcal{I}$, Π_i is the set of information partitions of Θ, and $u_i : A \times \Theta \to \mathbb{R}$ is the payoff function. Let us assume u_i is measurable with respect to Π_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 Π_i containing θ. Furthermore, we 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 cognitive type that embed $\mathcal{G}$.

A solution in $\mathcal{U}$ is defined as a Bayesian extension of the cognitive equilibrium in $\mathcal{G}$. For each $i \in \mathcal{I}$, let us denote player i's (pure) strategy in $\mathcal{U}$ by $t_i : K_i \times \Theta \to A_i$, and assume that t_i is Π_i-measurable.

Definition 2.2. A strategy profile t^* constitutes a *Bayesian cognitive equilibrium* of $\mathcal{U}$ if, for any $i \in \mathcal{I}$, $k_i \in \mathbb{N}$, $\theta \in \Theta$, and $a_i \in A_i$,

$$\sum_{(k_{-i},\ \theta') \ \in \ K_{-i}\times\pi_i(\theta)} u_i(t^*(\mathbf{k},\ \theta'),\theta')P(\theta' \mid \pi_i(\theta))\mu_i(k_i)(k_{-i})$$
$$\geq \sum_{(k_{-i},\ \theta') \ \in \ K_{-i}\times\pi_i(\theta)} u_i(a_i, t^*_{-i}(k_{-i},\ \theta'),\theta')P(\theta' \mid \pi_i(\theta))\mu_i(k_i)(k_{-i}).$$

2.4 Robustness

We say that a cognitive equilibrium s^* is robust to incomplete information, if that equilibrium is played with high probability in some Bayesian cognitive equilibrium of any $\mathcal{U} \in E(\mathcal{G})$ whenever $\mathcal{U}$ is sufficiently "close" to $\mathcal{G}$ provided that level-0 types' actions are commonly known in $\mathcal{G}$. To formally express this notion, let us firstly introduce a "distance" of two different strategies in $\mathcal{U}$.

Definition 2.3. For each $\mathbf{k} \in K$, an action distribution induced by a strategy profile t of $\mathcal{U}$ is given by $\alpha_{\mathbf{k}}(a) = \sum_{\theta\in\Theta} 1_{t(\mathbf{k},\theta)}(a)P(\theta)$ for any $a \in A$, where $1_{t(\mathbf{k},\theta)}(a)$ is an indicator function which takes 1 if a is chosen given $\mathbf{k}$ and θ under t.

In particular, we say an action distribution profile $(\alpha_{\mathbf{k}})_{\mathbf{k}\in K}$ is an *equilibrium action distribution profile* of $\mathcal{U}$, if there exists a Bayesian cognitive equilibrium t^* such that $\alpha_{\mathbf{k}}(a) = \sum_{\theta\in\Theta} 1_{t^*(\mathbf{k},\theta)}(a)P(\theta)$ for any $a \in A$ and $\mathbf{k} \in K$. The following measure defines the "distance" between two action distribution profiles, α and β:

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

Secondly, let us consider the situation where actions of level-0 types are common knowledge. Let $s^*(0)$ denote this action profile, and let us redefine an incomplete information game with cognitive type $\mathcal{G}^* = (\mathcal{I}, A, K, \mu, g, s^*(0))$. Then $\mathcal{G}^*$ is said to be close to an embedding game $\mathcal{U}$, if the actions of level-0 types and payoff functions under $\mathcal{U}$ are equal to those under $\mathcal{G}^*$ with high probability and players know that.[11] For each embedding game $\mathcal{U} \in E(\mathcal{G}^*)$, write $\Omega_{\mathcal{U}}$ for a collection of such payoff states: $\Omega_{\mathcal{U}} \equiv \{\theta \in \Theta : t_i(0, \theta') = s^*_i(0)$ and $u_i(a, \theta') = g_i(a)$ for all $a \in A$, $\theta' \in \pi_i(\theta)$ and $i \in \mathcal{I}\}$. An embedding game, $\mathcal{U}$, is an *ε-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 ε-elaborations of $\mathcal{G}^*$. At last, we are ready to define the robustness of cognitive equilibria.

Definition 2.4. An action distribution α induced by a cognitive equilibrium s^* of $\mathcal{G}^*$ is robust to incomplete information if, for every $\delta > 0$, there exists a $\bar{\varepsilon} > 0$ such that every $\mathcal{U} \in E(\mathcal{G}^*, \varepsilon)$ has a Bayesian cognitive equilibrium t^* that induces the equilibrium action distribution β with $\parallel \alpha - \beta \parallel \leq \delta$ for all $\varepsilon \leq \bar{\varepsilon}$.

[11] Our result is unchanged if we allow either one of level-0 actions or payoffs are different from the embedded game.

Two comments are to be added on the definition of robustness. First, by definition of ε-elaboration, we assume that payoff uncertainty does not affect players' beliefs on cognitive types; that is, $\mathcal{G}^*$ and $\mathcal{U}$ share the same μ, and $\mathbf{k}$ and θ are independent. Second, we implicitly suppose the outside observer does not know anything about players' cognitive types by using a sup-metric with respect to players' reasoning levels. But it is often assumed that players have an upper-bound, say $\bar{k}$, about their cognitive levels. Section 5 formally expresses these alternative notions of robustness, and shows resulting differences.

3 Preliminaries: p-Dominance and Common p-Belief

Following Monderer and Samet (1989) and Morris et al. (1995), this section introduces two concepts, the **p**-dominance and the common **p**-belief. To connect these two concepts with our robustness, the well-known lemma in Kajii and Morris (1997) so called "*critical path result*" is also introduced.

3.1 p-Dominance

We need the concept of the **p**-dominance to measure the "strength" of each action profile and cognitive equilibrium. Fix an incomplete information game with cognitive type $\mathcal{G}$. Let us denote $\mathbf{p} = (p_1, p_2, ..., p_i) \in [0,1]^I$. Let $\phi_i \in \Delta(A_{-i})$ for each $i \in \mathcal{I}$ and denote the probability assigned to $a_{-i} \in A_{-i}$ under ϕ_i by $\phi_i(a_{-i})$.

Definition 3.1. An action profile $a^* \in A$ is said to be **p**-*dominant* if, for any $i \in \mathcal{I}$, $a_i \in A_i$, and $\phi_i \in \Delta(A_{-i})$ with $\phi_i(a^*_{-i}) \geq p_i$, we have

$$\sum_{a_{-i} \in A_{-i}} \phi_i(a_{-i}) g_i(a^*_i, a_{-i}) \geq \sum_{a_{-i} \in A_{-i}} \phi_i(a_{-i}) g_i(a_i, a_{-i}).^{12}$$

Thus a^*_i becomes a best response for player i if he believes that other players will play a^*_{-i} with probability at least p_i. We also say that a^* is *strict* **p**-*dominant* if, for any $i \in \mathcal{I}$, $a_i \in A_i \setminus \{a^*_i\}$, and $\phi_i \in \Delta(A_{-i})$ with $\phi_i(a^*_{-i}) > p_i$, we have $\sum_{a_{-i} \in A_{-i}} \phi_i(a_{-i}) g_i(a^*_i, a_{-i}) > \sum_{a_{-i} \in A_{-i}} \phi_i(a_{-i}) g_i(a_i, a_{-i})$.

Next, we propose a natural analogue of **p**-dominant action profile in games with cognitive type. For any $s_i \in S_i$ and $k_i \in K_i$, let $s_i|_{k_i} : \{0, 1, ..., k_i - 1\} \to A_i$ denote a restricted strategy of player i such that $s_i|_{k_i}(k) = s_i(k)$ for any $0 \leq k < k_i$. $S_i|_{k_i}$ is the set of such restricted strategies. Let $\lambda_i(k_i) \in \Delta(S_{-i}|_{k_i})$ for each $k_i \in \mathbb{N}$ and $i \in \mathcal{I}$.[13] Thus $\lambda_i(k_i)$ can be interpreted as a limited conjecture of level-k_i type of player i about what the other players will choose only when their reasoning levels are strictly lower than k_i. Let $\lambda_i(k_i)(s_{-i}|_{k_i})$ denote the probability assigned to a restricted strategy profile $s_{-i}|_{k_i}$ under $\lambda_i(k_i)$.

[12] In standard complete information games, **p**-dominant action profile constitutes Nash equilibrium.
[13] We write $S_{-i}|_{k_i} = \Pi_{j \neq i} S_j|_{k_i}$, and $s_{-i}|_{k_i} = \Pi_{j \neq i} s_j|_{k_i}$.

Definition 3.2. A strategy profile s^* is said to be $\mathbf{p}$-*dominant* if, for any $i \in \mathcal{I}$, $k_i \in \mathbb{N}$, $a_i \in A_i$, and $\lambda_i(k_i)$ with $\lambda_i(k_i)(s^*_{-i}|_{k_i}) \geq p_i$, we have

$$\sum_{s_{-i}|_{k_i} \in S_{-i}|_{k_i}} \lambda_i(k_i)(s_{-i}|_{k_i}) \sum_{0 \leq k_j < k_i, \forall j \neq i} g_i(s_i^*(k_i), s_{-i}|_{k_i}(k_{-i}))\mu_i(k_i)(k_{-i})$$
$$\geq \sum_{s_{-i}|_{k_i} \in S_{-i}|_{k_i}} \lambda_i(k_i)(s_{-i}|_{k_i}) \sum_{0 \leq k_j < k_i, \forall j \neq i} g_i(a_i, s_{-i}|_{k_i}(k_{-i}))\mu_i(k_i)(k_{-i}).$$

This definition says that it is optimal for level-k_i type of player i to follow $s_i^*(k_i)$ if he (perceptually) believes other players would follow s^*_{-i} with probability at least p_i under his belief $\mu_i(k_i)$. Strict $\mathbf{p}$-dominance is defined in a similar manner as in strict $\mathbf{p}$-dominant action profile. This notion of $\mathbf{p}$-dominance satisfies the following desirable properties:

(1) s^* is a cognitive equilibrium if and only if it is 1-dominant;

(2) If s^* is $\mathbf{p}$-dominant, then it is $\mathbf{q}$-dominant for any $\mathbf{p} \leq \mathbf{q}$;

(3) For any *strict* cognitive equilibrium s^*, there exists some $\mathbf{p} \in [0,1)^{\mathcal{I}}$ such that s^* is $\mathbf{p}$-dominant;[14]

(4) For any cognitive equilibrium s^*, if $s^*(0)$ is a $\mathbf{p}$-dominant action profile, then s^* is $\mathbf{p}$-dominant with $s_i^*(k_i) = s_i^*(0)$ for any $k_i \in K_i$ and $i \in \mathcal{I}$.

Example 3.1. Consider the following two-player coordination game. There are two actions L and R, and payoffs are given in Table 1.

	L	R
L	2, 2	0, 0
R	0, 0	1, 1

Table 1: Payoff matrix of coordination game.

This game has two pure strategy Nash equilibria: (L, L) is 1/3-dominant and (R, R) is 2/3-dominant.[15] Now, if level-0 types play L (resp. R), then all higher-level types of any players will play L (resp. R) in any cognitive equilibrium. By the property (4), playing L (resp. R) by all higher-level types of any players is a 1/3-dominant (resp. 2/3-dominant) cognitive equilibrium. Next, suppose level-0 type of player 1 chooses L and level-0 type of player 2 chooses R, and both players have level-k belief. Then induced cognitive equilibrium will be:

$$s_1^*(k_1) = \begin{cases} R & \text{if } k_1 \text{ is odd} \\ L & \text{if } k_1 \text{ is even} \end{cases} \qquad s_2^*(k_2) = \begin{cases} L & \text{if } k_2 \text{ is odd} \\ R & \text{if } k_2 \text{ is even} \end{cases}$$

This level-k equilibrium is 2/3-dominant. Finally, suppose level-0 types choose L or R equally likely. Then since (L, L) is risk dominant, all levels of any players choose L in a unique cognitive equilibrium. This cognitive equilibrium is 2/3-dominant.

[14] Strict cognitive equilibrium is defined by replacing the inequality in Definition 1 into strict one.

[15] Whenever $p_i = p$ for any $i \in \mathcal{I}$, we just write p-dominance instead of $\mathbf{p}$-dominance.

However, the set of actions given in the **p**-dominant action profile does not necessarily coincide with that of **p**-dominant cognitive equilibrium.

Example 3.2. "The 11-20 Money Requesting Game" (Arad and Rubinstein 2012) There are two players and each player requests an amount of money between 11 and 20 shekels. Each player will receive the amount of money he requests, and if he calls exactly one shekel less than the other player, he will obtain additional 20 sheckels.[16] Thus, player i's payoff function is given by:

$$u_i(n_i, n_j) = \begin{cases} n_i + 20 & \text{if } n_i = n_j - 1 \\ n_i & \text{if otherwise} \end{cases}$$

This simple game has no (pure strategy) Nash equilibrium, so that there exists no **p**-dominant action pair for any $\mathbf{p} \in [0,1]^2$. But the following strategy profile constitutes a 19/20-dominant level-k equilibrium. As in Arad and Rubinstein (2012), suppose "requesting 20" is the action for level-0 types and both players have level-k belief. Then a strategy profile $s_i(k_i) = 20 - t$ where $t \equiv k_i \pmod{10}$ with $0 \leq t \leq 9$ for each $k_i \in K_i$ and $i = 1, 2$ is a 19/20-dominant level-k equilibrium.[17]

3.2 Common p-Belief

This subsection is intended to introduce the idea of common p-belief, and showing that if the ex ante probability of an event E is high then the ex ante probability of the event in which E is common **p**-belief is also high by the critical path result. Fix an information structure of embedding game $\mathcal{U}$, $(\mathcal{I}, \Theta, P, \{\Pi_i\}_{i\in\mathcal{I}})$. Let $\mathcal{F}_i$ denote a σ-algebra generated by Π_i for each $i \in \mathcal{I}$. To characterize player's conditional belief at given payoff state, let us define the p-belief operator as in Monderer and Samet (1989). For any $E \in \mathcal{F}_i$, 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 a collection of states in which player i believes the event E with probability at least p. For any $\mathbf{p} \in [0,1]^I$, define $B_*^{\mathbf{p}}(E) = \bigcap_{i\in\mathcal{I}} B_i^{p_i}(E)$. Then $B_*^{\mathbf{p}}(E)$ is a set of states in which event E is believed by each player i with probability at least p_i. We say that an event is *common* ***p**-belief* if it is believed by each player with probability at least p_i that it is believed by each player with probability at least p_i that ... , and so on, *ad infinitum*.

Definition 3.3. An event E is *common* ***p**-belief* at θ if $\theta \in C^{\mathbf{p}}(E) \equiv \bigcap_{n\geq 1}[B_*^{\mathbf{p}}]^n(E)$.

We say an event E is *simple* if $E = \bigcap_{i\in\mathcal{I}} E_i$ for each $E_i \in \mathcal{F}_i$. Kajii and Morris (1997) shows the following:

[16] There exists a cyclic version of the 11-20 game in which each player gets 20 shekels as a bonus if he requests 20 and the other player requests 11. This game has a 29/40-dominant level-k equilibrium.

[17] In this equilibrium, level-10 types get 20 by requesting 20 since level-9 types would request 11. Consider the worst case scenario where he believes level-9 type requests 11 with probability p_i and requests 20 with probability $1 - p_i$. Then if he chooses 19, he would receive $39 - 20p_i$ instead of 20 in the equilibirum. Thus we need $20 \geq 39 - 20p_i$ for each $i = 1, 2$ to make the equilibrium **p**-dominant.

Lemma 3.1. (Proposition 4.2 of Kajii and Morris (1997))
If $\sum_{i\in\mathcal{I}} p_i < 1$, any simple event E satisfies:

$$P[C^{\mathbf{p}}(E)] \geq 1 - (1 - P(E))\left(\frac{1 - \min_{i\in\mathcal{I}}(p_i)}{1 - \sum_{i\in\mathcal{I}} p_i}\right).$$

The following result gives a loose upper-bound for the ex ante probability of event $[B_*^{\mathbf{p}}]^K(E)$.

Lemma 3.2. (Lemma B of Kajii and Morris (1997))
For any $\mathbf{p} \in [0,1)^I$ and any measurable event E, we have

$$P[[B_*^{\mathbf{p}}]^K(E)] \geq 1 - \left(1 + \sum_{i\in\mathcal{I}} \frac{p_i}{1-p_i}\right)^K (1 - P(E)).$$[18]

4 Robust Predictions under Finite Depth of Reasoning

4.1 Motivating Example - Coordinated Attack Game

The following simple example from Strzalecki (2010) gives an intuition that how player's finite depth of reasoning affects the robustness of equilibria. There are two players, and each player chooses either one of the two actions, "*Attack*" (A) or "*Not Attack*" (NA). Payoffs are shown in the left-hand payoff matrix of Table 2.

	Attack	*Not Attack*
Attack	1, 1	-2, 0
Not Attack	0, -2	0, 0

	Attack	*Not Attack*
Attack	-2, -2	-2, 0
Not Attack	0, -2	0, 0

* Only when the enemy is strong.

Table 2: Payoff matrices.

Observe that, for each player, playing A becomes a best response whenever he believes that the other player would choose A with probability at least 2/3. On the other hand, NA is a best response whenever the other player would choose NA with probability at least 1/3. Thus (NA, NA) is a *risk dominant* action pair, and (A, A) is *risk dominated.*

Case 1: Standard Complete Information Game

First, we consider a standard complete information game and explain how the test of KM robustness works. In this game, it follows from Proposition 5.3 of Kajii and Morris (1997) that (NA, NA) is KM robust but (A, A) is not. To see why the latter is *not* KM robust, let us think about the following e-mail game type elaboration[19]: The information structure is given by the triple $(\Theta, (\Pi_i)_{i=1,2}, P)$, where $\Theta = \{1, 2, 3, ...\}$ is the set of states, $\Pi_1 = \{\{1\}, \{2, 3\}, \{4, 5\}, ...\}$ and $\Pi_2 = \{\{1, 2\}, \{3, 4\}, \{5, 6\}, ...\}$ are partition structures for each player, and P is a common prior given by $P(n) = \varepsilon(1 - \varepsilon)^{n-1}$ for any $n \in \mathbb{N}$. Posteriors are shown in Table 3.

[18] This bound is loose in the sense that the right-hand side goes zero as $K \to \infty$.
[19] For the interpretation, see Section A.3 of Appendix A.

Π_1	1	$\frac{1}{2-\varepsilon}$	$\frac{1-\varepsilon}{2-\varepsilon}$	$\frac{1}{2-\varepsilon}$	$\frac{1-\varepsilon}{2-\varepsilon}$	...
Θ	1	2	3	4	5	...
Π_2	$\frac{1}{2-\varepsilon}$	$\frac{1-\varepsilon}{2-\varepsilon}$	$\frac{1}{2-\varepsilon}$	$\frac{1-\varepsilon}{2-\varepsilon}$	...	

Table 3: Partitions and posteriors.

Suppose players (except an outside analyst) know their enemy is strong and have a different payoff structure at state $\theta = 1$: see the right-hand payoff matrix of Table 2. Let $\widehat{\varepsilon} = \varepsilon(2-\varepsilon)$, then this incomplete information game is a $\widehat{\varepsilon}$-elaboration (in the framework of Kajii and Morris (1997)) of the original coordinated attack game. We argue that, in this $\widehat{\varepsilon}$-elaboration, (*A, A*) is a unique Bayesian Nash equilibrium, so that (*NA, NA*) is not KM robust. Now since player 1 knows at $\theta = 1$ that the enemy is strong and hence *NA* is chosen, player 2 believes that player 1 would choose *NA* with probability at least $1/(2-\varepsilon)$ in the event $\{1, 2\}$. Then since *NA* is risk dominant, player 2's (interim) unique best response in $\{1, 2\}$ is *NA*. Given this, in the event $\{2, 3\}$, player 1 believes that player 2 would choose *NA* with probability at least $1/(2-\varepsilon)$, so that player 1's (interim) unique best response in $\{2, 3\}$ is *NA*. Continuing similar arguments, we can conclude that (*NA, NA*) is played everywhere in *any* Bayesian Nash equilibrium (Rubinstein 1989).

Case 2: Level-k Model

Next, let us introduce cognitive types in this game, and suppose players have level-k belief. Consider a level-k equilibrium in which level-0 types choose A. We show that this 2/3-dominant level-k equilibrium is *not* robust. Consider the same email-game type elaboration, and let player 1's level-0 type choose *NA* at $\theta = 1$. Then *NA* becomes a unique (interim) best response for any cognitive type of player 2 in $\{1, 2\}$ since he believes any cognitive type of player 1 would play *NA* with probability at least $1/(2-\varepsilon)$. But in turn *NA* would be chosen in $\{2, 3\}$ by any level (except level-0) of player 1: see Table 4. Continuing similar arguments, the ex ante probability of (*NA, NA*) being played converges to 1 as players' reasoning levels get higher in *any* Bayesian level-k equilibrium.[20] Hence, the original level-k equilibrium is not robust.

$k_1 = 2$	*NA*	*NA*		*A*		*A*		*A*		...
$k_1 = 1$	*NA*	*NA*		*A*		*A*		*A*		...
$k_1 = 0$	*NA*	*A*		*A*		*A*		*A*		...
	1	2	3	4	5	6	7	8	9	...
$k_2 = 0$	*NA*		*A*		*A*		*A*			...
$k_2 = 1$	*NA*		*A*		*A*		*A*			...
$k_2 = 2$	*NA*		*NA*		*A*		*A*			...

Table 4: Level 0, 1, and 2 types' actions.

[20]For the formal treatment, see Proposition A.3 in Appendix A.

Case 3: Cognitive-hierarchy Model

Finally, suppose players have cognitive-hierarchy belief instead of level-k belief, and especially we assume that players' reasoning levels are uniformly distributed.[21] The first difference arises for the action of player 2's level-2 type in $\{3,4\}$. Since he believes player 1 is level-0 or level-1 types equally likely and believes A would be chosen with probability $1-1/2(2-\varepsilon)$, his unique best response in $\{3,4\}$ is A if ε is sufficiently small: see Table 5.

	1	2	3	4	5	6	7	8	9	...
$k_1=2$	NA	NA		A		A		A		...
$k_1=1$	NA	NA		A		A		A		...
$k_1=0$	NA	A		A		A		A		...
$k_2=0$	NA		A		A		A			...
$k_2=1$	NA		A		A		A			...
$k_2=2$	NA		$\boxed{A}$		A		A			...

Table 5: Level 0, 1, and 2 types' actions.

Actually, we can show that ex ante probability of event in which (NA, NA) is played by some cognitive type can be arbitrarily small. In other words, *(A, A) is robust to the e-mail game type elaboration*, which is not attained when players have level-k belief.

For this particular information structure, Strzalecki (2010) shows that the ex ante probability of (NA, NA) being played by some cognitive type can be arbitrary small if both players' beliefs satisfy *Nondivergent beliefs property*: for any strictly increasing sequence $(k^m)\in\mathbb{N}^\infty$, $\inf_m \mu_i(k^m)(k^{m-1}\le k_j)<(1-p)(2-\varepsilon)$.[22] Thus Strzalecki's result tells us when a cognitive equilibrium is robust to *the email-game type elaboration*. Our objective is to derive a sufficient condition for a cognitive equilibrium being robust to *any* ε*-elaboration* including the email-game type elaboration. Main result of this paper implies that, for two-player cases, p-dominant level-k equilibrium is robust to any ε-elaboration if $p<1/2$. In contrast, p-dominant CH equilibrium is robust for any $p\in[0,1)$. The following analysis starts with formalizing our intuition here by using tools such as **p**-dominance and common **p**-belief.

4.2 Robustness and p-Dominance

Fix any $\mathcal{G}$ and action profile of level-0 types $s^*(0)$. For any ε-elaboration of $\mathcal{G}^*$, we construct an event in which $s_i^*(k_i)$ becomes a best response for level-k_i type of player i. Take any $\mathcal{U}\in E(\mathcal{G}^*,\varepsilon)$ and let $E_{(i,0)}=E_0=\Omega_{\mathcal{U}}$ for any $i\in\mathcal{I}$. We inductively define the event $E_{(i,k_i)}$ as follows:

$$E_{(i,k_i)}=\Big\{\theta\in\Theta:\sum_{0\le k_j<k_i,\forall j\ne i}\mu_i(k_i)(k_{-i})P(E_{k_{-i}}\mid\pi_i(\theta))\ge p_i\Big\}\cap E_0$$

[21] Uniform distribution is assumed just for simplicity. The following argument holds for any distribution.

[22] Appendix A yields a generalization of this result in the spirit of Morris et al. (1995).

for any $k_i \in \mathbb{N}$ and $i \in \mathcal{I}$, where $E_{k_{-i}} = \bigcap_{j \neq i} E_{(j,k_j)}$. The following Lemma 4.1 ensures that $E_{(i,k_i)}$ is the event we are looking for.

Lemma 4.1. *Suppose s^* is a $\mathbf{p}$-dominant cognitive equilibrium of $\mathcal{G}^*$. Then for any $\mathcal{U} \in E(\mathcal{G}^*, \varepsilon)$, there exists a Bayesian cognitive equilibrium t^* in which level-k_i type of player i plays $s_i^*(k_i)$ in $E_{(i,k_i)}$ for any $k_i \in K_i$ and $i \in \mathcal{I}$.*

Proof. Our proof proceeds parallel with Lemma 5.2 of Kajii and Morris (1997). Fix $\mathcal{U} \in E(\mathcal{G}^*, \varepsilon)$. Since $E_{(i,0)} = \Omega_{\mathcal{U}}$ for any $i \in \mathcal{I}$, we have $t_i(0, \theta) = s_i^*(0)$ for any $\theta \in E_{(i,0)}$ and $i \in \mathcal{I}$. Consider the modified embedding game $\mathcal{U}'$ where each player's strategy must satisfy $t_i(k_i, \theta) = s^*(k_i)$ for any $\theta \in E_{(i,k_i)}, k_i \in \mathbb{N}$, and $i \in \mathcal{I}$. There exists a Bayesian cognitive equilibrium t^* of the modified game. By definition, t_i^* is a best response to t_{-i}^* at any $\theta \notin E_{(i,k_i)}$. Let $\theta \in E_{(i,k_i)}$ and consider level-k_i type of player i. Then since we have $\sum_{0 \leq k_j < k_i, \forall j \neq i} \mu_i(k_i)(k_{-i}) P(E_{k_{-i}} \mid \pi_i(\theta)) \geq p_i$ by definition, he believes other players would follow t_{-i}^* with probability at least p_i and we have $u_i(a, \theta) = g_i(a)$ for any $a \in A$. Therefore, $s_i^*(k_i)$ is a best response for level-k_i type of player i. Thus t^* is a Bayesian cognitive equilibrium of $\mathcal{U}$ which satisfies the desired property. Since our choice of $\mathcal{U}$ is arbitrary, we are done. □

Our construction of $(E_{(i,k_i)})_i^{k_i}$ is tight in the sense that there exists a class of incomplete information games with cognitive type such that, for some embedding game $\mathcal{U} \in E(\mathcal{G}^*, \varepsilon)$, any Bayesian cognitive equilibrium t^* of $\mathcal{U}$ satisfies $t_i^*(k_i, \theta) = s_i^*(k_i)$ if and only if $\theta \in E_{(i,k_i)}$ for any $k_i \in K_i$ and $i \in \mathcal{I}$. In fact, Case 2 of Section 4.1 falls into this class.

4.3 A Sufficient Condition for Robustness when $\sum_{i \in \mathcal{I}} p_i < 1$

This subsection shows that level-k belief yields the most severe test for robustness in the sense that, if a $\mathbf{p}$-dominant cognitive equilibrium s^* is robust when players have level-k belief, then s^* is robust for any other belief specifications. Observe that under μ^{lk}, $(E_{(i,k_i)})_i^{k_i}$ can be written as: $E_{(i,k_i)} = \{\theta \in \Theta : P(E_{(\mathbf{k_i - 1})} \mid \pi_i(\theta)) \geq p_i\} \cap E_0 = B_i^{p_i}(E_{(\mathbf{k_i-1})}) \cap E_0$ for all $k_i \in \mathbb{N}$ and $i \in \mathcal{I}$. Let us especially denote this $E_{(i,k_i)}$ by $\widehat{E}_{(i,k_i)}$, and $E_{k_{-i}}$ by $\widehat{E}_{k_{-i}}$. The following Lemma 4.2 states that $\widehat{E}_{(i,k_i)}$ is a lower bound of $E_{(i,k_i)}$ in the sense of set inclusion.

Lemma 4.2. *For any $i \in \mathcal{I}$ and $k_i \in K_i$, $\widehat{E}_{(i,k_i)} \subseteq E_{(i,k_i)}$.*

Proof. See Appendix B. □

By Lemma 4.1 and Lemma 4.2, we know that for any $\mathcal{U} \in E(\mathcal{G}^*, \varepsilon)$, there always exists a Bayesian cognitive equilibrium of $\mathcal{U}$ in which the original cognitive equilibirum of $\mathcal{G}^*$ is played by level-k_i type of player i at any $\theta \in \widehat{E}_{(i,k_i)}$. Hence, if $\sum_{i \in \mathcal{I}} p_i < 1$, the robustness of $\mathbf{p}$-dominant cognitive equilibrium follows from Lemma 3.1.

Proposition 4.1. *Suppose s^* is a $\mathbf{p}$-dominant cognitive equilibrium of $\mathcal{G}^*$ with $\sum_{i \in \mathcal{I}} p_i < 1$. Then s^* is robust to incomplete information.*

Proof. Let s^* denote a $\mathbf{p}$-dominant cognitive equilibrium of $\mathcal{G}^*$ with $\sum_{i\in\mathcal{I}} p_i < 1$. Let α denote the equilibrium action distribution profile induced by s^*, which is given by $\alpha_{\mathbf{k}}(s^*(\mathbf{k})) = 1$ for any $\mathbf{k} \in K$. Fix any $\delta > 0$, and let $\varepsilon < \delta(1 - \sum_{i\in\mathcal{I}} p_i)/(1 - \min_{i\in\mathcal{I}}(p_i))$. Take any $\mathcal{U} \in E(\mathcal{G}^*, \varepsilon)$. Let us denote $\bar{k} = \max_i k_i$ for any $\mathbf{k} \in K$. By Lemma 4.2 $\widehat{E}_{(i,k_i)} \subseteq E_{(i,k_i)}$ for any $i \in \mathcal{I}$ and $k_i \in K_i$, and by construction $[B_*^{\mathbf{p}}]^{\bar{k}}(E) \subseteq \bigcap_{i\in\mathcal{I}} \widehat{E}_{(i,k_i)}$ for all $\mathbf{k} \in K$. But then by Lemma 3.1, for any $i \in \mathcal{I}$ and $\mathbf{k} \in K$, $P(E_{(i,k_i)}) \geq P(\widehat{E}_{(i,k_i)}) \geq P([B_*^{\mathbf{p}}]^{\bar{k}}(E_0)) \geq P([C^{\mathbf{p}}](E_0) \geq 1 - \varepsilon(1 - \min_{i\in\mathcal{I}}(p_i)/(1 - \sum_{i\in\mathcal{I}} p_i) > 1 - \delta$. By Lemma 4.1, there exists a Bayesian cognitive equilibrium t^* of $\mathcal{U}$ such that $P\big(\{\theta \in \Theta : (t^*(\mathbf{k}, \theta) = s^*(\mathbf{k})\}\big) \geq P(\bigcap_{i\in\mathcal{I}} \widehat{E}_{(i,k_i)}) > 1 - \delta$ for any $\mathbf{k} \in K$. Thus the equilibrium action distribution profile β induced by t^* satisfies $\beta_{\mathbf{k}}(s^*(\mathbf{k})) > 1 - \delta$ for any $\mathbf{k} \in K$. Therefore, $\parallel \alpha - \beta \parallel \leq \delta$ as desired. □

4.4 A Sufficient Condition for Robustness

We now know the robustness of a $\mathbf{p}$-dominant cognitive equilibrium whenever $\sum_{i\in\mathcal{I}} p_i < 1$. But Proposition 4.1 does not tell us anything about the robustness of $\mathbf{p}$-dominant cognitive equilibrium with $\sum_{i\in\mathcal{I}} p_i \geq 1$, since Lemma 3.1 cannot be applied directly in this case. To see how the condition $\sum_{i\in\mathcal{I}} p_i < 1$ is important, consider the email-game type elaboration given in Section 4.1, i.e., $\Theta = \{1, 2, 3, ...\}$, $\Pi_1 = \{\{1\}, \{2, 3\}, \{3, 4\}, ...\}$ and $\Pi_2 = \{\{1, 2\}, \{3, 4\}, \{5, 6\}, ...\}$, and $P(n) = \varepsilon(1 - \varepsilon)^{n-1}$ for any $n \in \mathbb{N}$. Suppose both players have level-k belief and let $E^c = \{1\}$. Then since we have $P(\{n\} \mid \{n, n + 1\}) = 1/(2 - \varepsilon) > 1/2$ for any $n \geq 1$, once a strict $(1 - p)$-dominant action pair with $p \geq 1/2$ is played in E^c, that action pair will be played contagiously in the limit for any cognitive equilibrium. Hence, if there exist two action pairs which are strict p-dominant and strict $(1 - p)$-dominant respectively with $p \geq 1/2$, the induced strict p-dominant cognitive equilibrium cannot be robust in the level-k model.

This observation is natural since, under level-k belief, actions of crazy types are contagiously played step by step as reasoning levels get higher. Hence, by applying the p-belief operator repeatedly, the ex ante probability of event $E_{(i,k_i)}$ can be arbitrarily small if $\sum_{i\in\mathcal{I}} p_i \geq 1$ through the contagion effect. However, the following Theorem 4.1 shows that, under a certain condition on belief profile μ, $\mathbf{p}$-dominant cognitive equilibrium with $\sum_{i\in\mathcal{I}} p_i \geq 1$ can be robust to incomplete information. This occurs since, if each player believes that his opponents have sufficiently lower reasoning levels than him own, $E_{(i,k_i)}$ gets smaller compared to the case of level-k belief as Lemma 4.2 suggests. So, in this case, we can find a tighter lower-bound for $(E_{(i,k_i)})_i^{k_i}$.

Theorem 4.1. *Suppose s^* is a $\boldsymbol{p}$-dominant cognitive equilibrium of $\mathcal{G}^*$ with $\boldsymbol{p} \in [0, 1)^I$. If there exists a $n \in \mathbb{N}$ such that*

$$\inf_{k_i \geq n} \sum_{0 \leq k_j < n, \forall j \neq i} \mu_i(k_i)(k_{-i}) > \frac{I \cdot p_i - 1}{I - 1}$$

for any $i \in \mathcal{I}$, then s^* is robust to incomplete information.

Proof. See Appendix B. ☐

Theorem 4.1 implies that the size of the set of robust equilibria varies with how we specify the players' beliefs about other players reasoning levels. Since it is easy to verify that cognitive-hierarchy belief satisfies the assumption in Theorem 4.1 for any $p_i \in [0, 1)$, Corollary 4.1 immediately follows.

Corollary 4.1. *Generically in payoffs, any cognitive-hierarchy equilibrium is robust to incomplete information.*

We now know that two prominent models, level-k and cognitive-hierarchy, give two contrasting examples with respect to the size of robust predictions. Finally, let us consider a mixture of level-k and cognitive-hierarchy beliefs.

Example 4.1. (A mixture of level-k and cognitive hierarchy beliefs)
Suppose there are two players and $p_1 = p_2 = p$. Let $\alpha \in [0, 1]$ and $\lambda \in \Delta(\mathbb{Z}_+)$. For any $k_i \in \mathbb{N}$, define

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

By the previous argument, it follows that this belief satisfies the assumption in Theorem 4.1 if $\alpha < 2(1-p)$. Figure 1 shows the set of p-dominant cognitive equilibria, which are robust to incomplete information whenever they exist, as α and p varies.

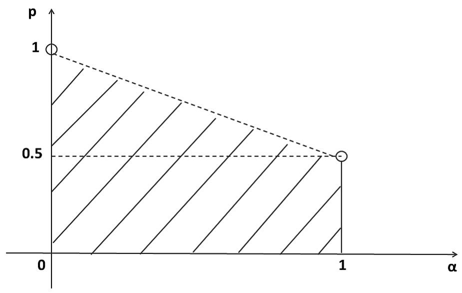

Figure 1. Set of robust equilibria.

5 Discussions

5.1 More General Type Spaces

Instead of defining incomplete information game with cognitive type, we can start the setup from incomplete information game with more general type spaces. An incomplete information game $\mathcal{U}$ is given by: $\mathcal{U} = (\mathcal{I}, \{A_i, T_i, \kappa_i, \nu_i, u_i\}_{i \in \mathcal{I}})$, where $T_i = C_i \times \Theta$ is player i's type space, $\kappa_i : C_i \to \mathbb{Z}_+$ is a mapping from player i's cognitive type to his reasoning level, and $\nu_i(t_i) \in \Delta(T_{-i})$ is his belief about other players' types. This formulation is more general in the sense that: we allow (1) correlation between cognitive type and payoff type; (2) player with same reasoning level to have different beliefs about other players' types. In the language of Kajii and Morris (1997), we have considered the restricted class of elaborations such that (1) and (2) are prohibited. It is easily shown that even if we allow (2), our result will not be changed:

Proposition 5.1. *Suppose s^* is a $\boldsymbol{p}$-dominant cognitive equilibrium of incomplete information game with Strzalecki's cognitive type. If there exists a $n \in \mathbb{N}$ such that*

$$\inf_{\{c_i : \kappa_i(c_i) \geq n\}} \sum_{\{c_{-i}:\ 0 \leq \kappa_j(c_j) \leq n-1, \forall j \neq i\}} \mu_i(c_i)(c_{-i}) > \frac{I \cdot p_i - 1}{I - 1}$$

for any $i \in \mathcal{I}$, then s^* is robust.

Heifetz and Kets (2012) consider more general type space which includes both standard universal type space and Strzalecki's universal cognitive type space as belief closed subspaces. By using their type space, we can fully investigate strategic effect of finite depth of reasoning. Here we introduce so called *sophisticated type* and see the flavor of such effect. Assume that there is a likelihood η of sophisticated type for each player i who can iterate best responses infinitely many times, and know the proportion η of each player being such type. Let ∞ denote sophisticated type. The extended cognitive type space for player i is given by $\widehat{K}_i = K_i \cup \{\infty\}$. Level-$\infty$ type of player i forms his belief $\widehat{\mu}_i$ on $\widehat{K}_{-i}$ such that $\widehat{\mu}_i(\{\widehat{k}_j \neq \infty \text{ for all } j \neq i\}) = (1-\eta)^I$. Consider an incomplete information game with extended cognitive type space $\widehat{\mathcal{G}}$. Then our robustness concept can be naturally extended to include sophisticated types by replacing K into $\widehat{K}$ and endowing the sophisticated types with $\widehat{\mu}_i$. Proposition 5.2 shows the strategic effect of finite depth of reasoning; i.e., even if player i is sophisticated type, he behaves as if he is cognitively unsophisticated since his action is affected by believing the other players may not be sophisticated.

Proposition 5.2. *Suppose $\widehat{s} : \widehat{K} \to A$ is a $\boldsymbol{p}$-dominant cognitive equilibrium of $\widehat{\mathcal{G}}$ with $\boldsymbol{p} \in [0,1)^I$. If we have $(1-\eta)^{\mathcal{I}} > \max_{i \in \mathcal{I}}\ p_i$, and if there exists a $n \in \mathbb{N}$ such that*

$$\inf_{k_i \geq n} \sum_{0 \leq k_j \leq n-1, \forall j \neq i} \mu_i(k_i)(k_{-i}) > \frac{I \cdot p_i - 1}{I - 1}$$

for any $i \in \mathcal{I}$, then $\widehat{s}$ is robust to incomplete information.

Proof. By our assumption, we can use Theorem 4.1 and we have, for any $\delta > 0$ and some $\varepsilon > 0$, there exists an event E such that any finite level of any player follows $\widehat{s}$ in $E \subseteq \Omega_{\mathcal{U}}$ with $P(E) > 1 - \delta$ for any (extended) embedding game $\widehat{\mathcal{U}} \in E(\widehat{\mathcal{G}}, \varepsilon)$. Since $(1-\eta)^{\mathcal{I}} > \max_{i \in \mathcal{I}}\ p_i$, we can take $0 < q < 1$ such that $q \cdot (1-\eta_i)^{\mathcal{I}} > \max_{i \in \mathcal{I}}\ p_i$. Define $B^q_*(E) = \cap_{i \in \mathcal{I}} B^q_i(E)$. Then since type ∞ of player i believes other players have finite reasoning levels with probability at least $(1-\eta)^{\mathcal{I}}$, and such types of other players follow $\widehat{s}$ in E, $\widehat{s}_i(\infty)$ becomes his best response at $\theta \in \Omega_{\mathcal{U}}$ if $P(E \mid \pi_i(\theta)) \geq q$ for any $i \in \mathcal{I}$. Thus $\widehat{s}$ is played at any $\theta \in B^q_*(E)$. But then by definition of E and by using Lemma 3.2, for any $\delta' > 0$, there exists some $\varepsilon' > 0$ such that $P(B^q_*(E)) > 1 - \delta'$ for any $\widehat{\mathcal{U}} \in E(\widehat{\mathcal{G}}, \varepsilon)$. Therefore, $\widehat{s}$ is robust. □

5.2 Degree of Knowledge about Players

By definition of robustness, we have implicitly put (at least) two assumptions on the outside observer's knowledge: First, *she does not know anything about players' cognitive types*, and hence the sup-metric is used with respect to reasoning levels. Instead, in experimental literature, it is often assumed that experimentalists know players have the upper bound for their reasoning levels, say $\bar{k}$. If we assume this, we can show that for *any* $\boldsymbol{p} \in [0, 1]$*, any* $\boldsymbol{p}$*-dominant cognitive equilibrium is "robust" to incomplete information.* So our robustness test is valid only when the outside observer is not sure about how players are strategically sophisticated and/or the possibility of *sophisticated types*.

Second, following experimental literature, we have assumed that *the outside observer knows players' beliefs about other players' cognitive types.* Thanks to this formulation, we are able to characterize level-k and cognitive-hierarchy models in terms of the size of robust predictions. However, it may happen that the outside observer has no information about how players model other players. In this situation, since level-k belief yields the most stringent test for robustness, it follows that *a* $\boldsymbol{p}$*-dominant cognitive equilibrium is "robust" if* $\sum_{i \in \mathcal{I}} p_i < 1$.

Appendix A: Contagion under Finite Depth of Reasoning

A.1 Definitions: Contagion and Marginal Belief Potential

In this Appendix, we use the same notations as before whenever no qualification is added. Given any two-player incomplete information game with cognitive type $\mathcal{G} = (\{1, 2\}, (A_i, u_i, K_i, \mu_i)_{i=1,2})$, let us define ε-elaboration $\mathcal{E}$ of $\mathcal{G}$ such that $\Omega_{\mathcal{E}} \equiv \{\theta \in \Theta : u_i(a, \theta') = g_i(a)$ for all $a \in A$, $\theta' \in \pi_i(\theta)$ and $i = 1, 2\}$, $|\pi_i| < \infty$ for any $\pi_i \in \Pi_i$ and $i = 1, 2$, and $E \equiv |\Theta \setminus \Omega_{\mathcal{E}}| < \infty$. A set of ε-elaborations is denoted by $E(\mathcal{G}, \varepsilon)$. Note that if we use i and j at the same time, the j means "*not i*". Remember that player i's strategy in $\mathcal{E}$ is given by $t_i : K_i \times \Theta \to A_i$, and t_i is Π_i-measurable. We say an action pair a^* is contagious in $\mathcal{E}$ if, once a_i^* is played in E by any level of both players, then a^* must be played everywhere as $k_1, k_2 \to \infty$. The formal definition is given by:

Definition A.1. An action pair a^* is said to be *contagious* in $\mathcal{E}$ if any Bayesian cognitive equilibrium of $\mathcal{E}$, t^*, satisfies $P(\{\theta \in \Theta : t^*(\mathbf{k}, \theta) = a^*\}) \to 1$ as $k_1, k_2 \to \infty$ whenever we have $t_i(k_i, E) = \{a_i^*\}$ for any $k_i \in K_i$ and $i = 1, 2$.[23]

Remark A.1. For any action pair a, consider a set of cognitive equilibria $S^*(a)$ of $\mathcal{G}$ such that for any $s^* \in S^*(a)$, there exists some $\mathbf{m} \in K$ such that $s^*(\mathbf{k}) = a$ for any $\mathbf{k} \gg \mathbf{m}$. By definition, if an action pair a^* is contagious in some $\mathcal{E} \in E(\mathcal{G}, \varepsilon)$, then s^* is robust only if $s^* \in S^*(a^*)$. In turn, if $s^* \in S^*(a^*)$ is robust, then other action pairs than a^* cannot be contagious in any $\mathcal{E} \in E(\mathcal{G}, \varepsilon)$.

[23] No difference arises for our results if we only require $t_i(k_i, E) = \{a_i^*\}$ for *some* $k_i \in K_i$ and $i = 1, 2$.

In incomplete information games *only with* payoff type, *the belief potential of event* E yields a sufficient condition on information system for such contagion to operate (Morris et al. 1995). Let us define $H_i^p(E) \equiv B_i^p(B_j^p(E)) \cup E$ as a contagion operator, and let us inductively define $[H_i^p]^k(E) = H_i^p([H_i^p]^{k-1}(E))$ for $k \geq 1$ with $[H_i^p]^0(E) = E$, and $[H_i^p]^\infty(E) = \bigcup_{k=1}^\infty [H_i^p]^k(E)$. The belief 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 A.2. The belief potential of event E is given by $\sigma(E) = \min_{i=1,2} \sigma_i(E)$, where $\sigma_i(E) = \sup\{p \in [0,1] \mid [H_i^p]^\infty(E) = \Theta\}$.

Example A.1. Consider the information structure given in Table 4.1. Let $E = \{1\}$. Since $P(\{n\} \mid \{n, n+1\}) = 1/(2-\varepsilon)$ for all $n \in \mathbb{N}$, $B_2^p(E) = \{1, 2\}$ if $p \leq 1/(2-\varepsilon)$ and $B_2^p(E) = \emptyset$ if $p > 1/(2-\varepsilon)$. But in turn, $H_1^p(E) = B_1^p B_2^p(E) = \{1, 2, 3\}$ if $p \leq 1/(2-\varepsilon)$ and $H_1^p(E) = \emptyset$ if $p > 1/(2-\varepsilon)$. Continuing this argument yields $[H_1^p]^\infty(E) = \Theta$ if $p \leq 1/(2-\varepsilon)$ and $[H_1^p]^\infty(E) = \emptyset$ if $p > 1/(2-\varepsilon)$. We can repeat similar logic for player 2, and hence the belief potential of E, $\sigma(E)$, is $1/(2-\varepsilon)$.

The belief potential of event is useful to know the global strength of the information system for contagion argument to operate. But since contagion proceeds step by step under finite depth of reasoning, it is important for us to know how player i's conditional belief will vary in the course of contagion. The marginal belief potential of event is defined as follows.

Definition A.3. Given the belief potential of event E, $\sigma(E)$, the marginal belief potential of event E for player i is defined by $\xi_{(i,n)}^p(E) \equiv \sup\{q \in [0,1] \mid B_i^q(B_j^p[C_i^p]^{n-1}(E)) \cap [C_i^p]^n \neq \emptyset\}$ for any $p \in (0, \sigma(E)]$, $n \in \mathbb{N}$, and $i = 1, 2$. Let $\xi_i^p(E)$ denote $\sup_n\{\xi_{(i,n)}^p(E)\}$.

Note that $\xi_{(i,n)}^p(E)$ is well defined since $p \in \{q \in [0,1] : B_i^q(B_j^p[C_i^p]^{n-1}(E)) \cap [C_i^p]^n \neq \emptyset\}$ for any $p \in (0, \sigma(E)]$, $n \in \mathbb{N}$, and $i = 1, 2$. It also clearly follows that $\xi_{(i,n)}^p(E) \geq \sigma_i(E)$ holds for any $p \in (0, \sigma(E)]$ and $i = 1, 2$. The following example gives an idea why the marginal belief potential is crucial.

Example A.2. Consider the following information structure which is different from the previous example only in $P(3) = \varepsilon(1-\varepsilon)^2/2$ and $P(4) = \varepsilon(1-\varepsilon)^3 + \varepsilon(1-\varepsilon)^2/2$. Table 6 shows the conditional probability $P(\{n\}|\{n, n+1\})$ for each $n \in \mathbb{N}$. Since $P(3)/(P(3) +$

<table>
<tr><td>player 1</td><td>1</td><td>|</td><td colspan="2">$\frac{2}{3-\varepsilon}$</td><td>|</td><td colspan="2">$\frac{3-2\varepsilon}{2\varepsilon^2-6\varepsilon+5}$</td><td>|</td><td colspan="2">$\frac{1}{2-\varepsilon}$</td><td>|</td><td>...</td></tr>
<tr><td></td><td>1</td><td></td><td>2</td><td>3</td><td></td><td>4</td><td>5</td><td></td><td>6</td><td>7</td><td></td><td>...</td></tr>
<tr><td>player 2</td><td colspan="2">$\frac{1}{2-\varepsilon}$</td><td>|</td><td colspan="2">$\frac{1}{2(2-\varepsilon)}$</td><td>|</td><td colspan="2">$\frac{1}{2-\varepsilon}$</td><td>|</td><td colspan="3">...</td></tr>
</table>

Table 6: Partitions and posteriors.

$P(4)) = 1/2(2-\varepsilon)$, 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 $n \geq 5$. In fact, the information structure of Example A.1 yields $\xi^p_{(i,n)}(E) = 1/(2-\varepsilon)$ for any $p \in (0, 1/(2-\varepsilon)]$, $n \in \mathbb{N}$, and $i = 1, 2$. But now we have $\xi^p_{(1,1)}(E) = 2/(3-\varepsilon)$, $\xi^p_{(1,2)}(E) = (3-2\varepsilon)/(2\varepsilon^2 - 6\varepsilon + 5)$, $\xi^p_{(2,1)}(E) = 1/(2-\varepsilon)$, $\xi^p_{(2,2)}(E) = 1/2(2-\varepsilon)$, and $\xi^p_{(1,n)}(E) = \xi^p_{(2,n)}(E) = 1/(2-\varepsilon)$ for any $n \geq 3$ and $p \in [0, 1/2(2-\varepsilon)]$.

A.2 Contagion under Finite Depth of Reasoning

Suppose a^* is strict p-dominant in $\mathcal{E}$ with $p < \sigma(E)$. Under this condition, the main theorem of Morris et al. (1995) shows that once a^*_i is played in some (Π_i-measurable) event E, a^* is played at any state in any Bayesian Nash equilibrium. If players have level-k belief, the number of iteration of contagion operator increases as reasoning levels get higher. Hence, if a^* satisfies the sufficient condition of Morris et al. (1995) and both players have level-k belief, a^* is contagious in our sense too.

Proposition A.3. *Suppose players have level-k belief. If a^* is strict p-dominant and $p < \sigma(E)$, then a^* is contagious in $\mathcal{E}$.*

Proof. Let $E_{(i,0)} = E$ and define inductively $E_{(i,k_i)} = \{\theta \in \Theta : \sum_{t=0}^{k_i-1} \mu_i(k_i)(t) P(E_{(j,t)} \mid \pi_i(\theta)) > p\} \cup E$. Then for level-$k_i$ type of player i, a^*_i becomes a unique interim best response in $E_{(i,k_i)}$. Suppose players have level-k belief and let us denote $\widehat{E}_{(i,k_i)} = E_{(i,k_i)} = \{\theta \in \Theta : \mu_i(k_i)(k_i - 1) P(E_{(j,k_i-1)} \mid \pi_i(\theta)) > p\} \cup E$. Observe that for any $i = 1, 2$, we have the following: (1) $\widehat{E}_{(i,k_i)}$ is increasing in k_i; (2) $E_{(i,k_i)} \subseteq \widehat{E}_{(i,k_i)}$ for any $k_i \in \mathbb{Z}_+$; (3) $\widehat{E}_{(i,2k)} = [C^p_i]^k(E)$ and $\widehat{E}_{(i,2k+1)} = B^p_i [C^p_j]^k(E) \cup E$ for any $k \in \mathbb{Z}_+$. Since we have $p < \sigma(E)$, $[C^p_i]^{k_i}(E) \uparrow \Theta$ as $k_i \to \infty$ for any $i = 1, 2$. Thus we have $P(\widehat{E}_{(i,k_i)}) \uparrow 1$ as $k_i \to \infty$. □

In contrast, we can show that there exists a class of ε-elaborations such that even if the above conditions are satisfied, there exists a Bayesian cognitive equilibrium in which other actions than a^*_i is played by any level of both players with positive probability, that is, a^* cannot be contagious.

Proposition A.4. *Suppose a^* is strict p-dominant, $\bar{a}$ is strict $(1-p)$-dominant with $p \in (0, \sigma(E)]$ in $\mathcal{E}$. If, for any $i = 1, 2$, there exists a $n \in \mathbb{N}$ such that*

$$\sum_{t=0}^{n-1} \mu_i(k_i)(t) > 1 - \frac{p}{\xi^p_i(E)}$$

for any $k_i \geq n$, then a^* cannot be contagious in $\mathcal{E}$.

Proof. To show this we use the following Lemma A.4. of Oyama and Tercieux (2012)[24]:

Lemma A.1. *Fix an incomplete information game and $p \in (0, 1]$. Then any event $E \in \mathcal{F}_1 \oplus \mathcal{F}_2$ satisfies $P((C^p_i)^K(E)) \leq P(E) \sum_{k=0}^{2K} \left(\frac{1-p}{p}\right)^k$ for all $i = 1, 2$.*

[24] Lemma A.4. of Oyama and Tercieux (2012) allows *non-common priors*. Under a common prior, this lemma immediately follows from the critical path result.

Consider a Bayesian cognitive equilibrium t^* with $t^*(0,\theta) = \bar{a}$ for all $\theta \in \Omega_{\mathcal{E}}$. If we can show that there exists some K such that $E_{(i,k_i)} \subseteq E_{(i,K)}$ for any $k_i \in K_i$, then the result follows from this lemma since $\bar{a}_i$ is played in t^* by level-k_i type of player i in $\Theta \setminus E_{(i,k_i)}$.

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 $\theta \in \widehat{E}_{(i,N)} \setminus \widehat{E}_{(i,N-1)}$.

$\because$) Suppose not. Then $\exists i \in \{1, 2\}$, $\forall K_i \in \mathbb{N}$, $\exists N \geq K_i$ such that $P(\widehat{E}_{(j,n-1)} \mid \pi_i(\theta_N)) > 0$ for some $\theta_N \in \widehat{E}_{(i,N)} \setminus \widehat{E}_{(i,N-1)}$. Define inductively $N_1 = \min\{N \geq 1 : P(\widehat{E}_{(j,n-1)} \mid \pi_i(\theta_N)) > 0$ for some $\theta_N \in \widehat{E}_{(i,N)} \setminus \widehat{E}_{(i,N-1)}\}$, $N_2 = \min\{N \geq N_1 + 1 : P(\widehat{E}_{(j,n-1)} \mid \pi_i(\theta_N)) > 0$ for some $\theta_N \in \widehat{E}_{(i,N)} \setminus \widehat{E}_{(i,N-1)}\}$,... 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 $\theta_{N_m} \in \widehat{E}_{(i,N_m)} \setminus \widehat{E}_{(i,N_m-1)}$ and $\theta_{N_l} \in \widehat{E}_{(i,N_l)} \setminus \widehat{E}_{(i,N_l-1)}$ with $m \neq l$. But then, since $P(\widehat{E}_{(j,n-1)} \mid \pi_i(\theta_{N_m})) > 0$ for all $m \in \mathbb{N}$, we must have $\mid \widehat{E}_{(j,n-1)} \mid = \infty$. On the other hand, since $\mid E \mid < \infty$ and $\mid \pi_i \mid < \infty$ for all $\pi_i \in \Pi_i$ and $i = 1, 2$, we have $\mid \widehat{E}_{(j,n-1)} \mid < \infty$, contradicting. □

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

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

$\because$) 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 $\xi_i^p(E)$. □

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 $i = 1, 2$.

$\because$) 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 $k \leq n - 1$. 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. □

By Claim 3, we have $E_{(i,k_i)} \subseteq \widehat{E}_{(i,2K-1)} \subseteq \widehat{E}_{(i,2K)}$ for any $k_i \in K_i$ and $i = 1,2$. Since $P(\cup_{i=1,2}\widehat{E}_{(i,2K)}) \leq 2\varepsilon \sum_{k=0}^{4K} \left(\frac{1-p}{p}\right)^k$ by Lemma A.1, letting $\varepsilon < 1/2 \sum_{k=0}^{4K} \left(\frac{1-p}{p}\right)^k$ yields the result. □

Remark A.2. Consider the game specified in Section 4.1. It is easy to check that $\xi_i^p(E) = 2-\varepsilon$ for $i = 1,2$, and if players have the mixture of level-k and cognitive-hierarchy beliefs with $\alpha = p(2-\varepsilon)$, (R,R) becomes contagious. Since there exists a $n \in \mathbb{N}$ such that $\sum_{t=0}^{n-1} \mu_i(k_i)(t) = 2-\varepsilon$ for any $k_i \geq n$, our sufficient condition is tight in this sense.

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

Consider a version of email game as in Section 4.1. 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. 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. For the payoff matrix see Table 2, and for the information structure see Table 4.1. Let us consider a cognitive equilibrium in which level-0 types choose *Attack* at any states other than $(0,0)$.[25] By using Proposition A.4, 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 A.2. *Take any constant M (> 0). For any $i = 1,2$, there exists a $n \in \mathbb{N}$ such that $\sum_{t=n-1}^{k-1} \mu_i(k)(t) < M$ for any $k \geq n$ 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. (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, $\mu_i(k^{m'})(\{k^j \geq k^{m'-1}\}) < M$. □

Corollary A.1. (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.

[25] Same result follows if we assume level-0 players choose *Attack* after receiving a certain number of messages.

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 A.2 and Proposition A.4, 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*. □

Appendix B: Omitted Proofs

Proof of Lemma 4.2

Proof. We show this by induction. First, $E_{(i,1)} = \widehat{E}_{(i,1)} = B_i^{p_i}(E_0)$ for any $i \in \mathcal{I}$ by definition. Next, suppose that for any $i \in \mathcal{I}$, we have $\widehat{E}_{(i,k_i)} \subseteq E_{(i,k_i)}$ for all $0 \le k_i \le k \in \mathbb{N}$. Then since $\widehat{E}_{k_{-i}} \subseteq E_{k_{-i}}$ for any $0 \le k_j \le k$ and $j \neq i$, we have

$$\begin{aligned} E_{(i,k+1)} &= \left\{\theta \in \Theta : \sum_{0 \le k_j \le k, \forall j \neq i} \mu_i(k+1)(k_{-i}) P(E_{k_{-i}} \mid \pi_i(\theta)) \ge p_i\right\} \cap E_0 \\ &\supseteq \left\{\theta \in \Theta : \sum_{0 \le k_j \le k, \forall j \neq i} \mu_i(k+1)(k_{-i}) P(\widehat{E}_{k_{-i}} \mid \pi_i(\theta))\} \ge p_i\right\} \cap E_0 \\ &\supseteq \{\theta \in \Theta : P(\widehat{E}_{\mathbf{k}} \mid \pi_i(\theta)) \ge p_i\} \cap E_0 \\ &= \widehat{E}_{(i,k+1)}. \end{aligned}$$

The third inequality follows since $\widehat{E}_{(i,k_i)}$ is decreasing in k_i. By our induction hypothesis the result follows. □

Proof of Theorem 4.1

Proof. By Proposition 4.1, it suffices to show the case where $\sum_{i \in \mathcal{I}} p_i \ge 1$. Suppose there exists a $n \in \mathbb{N}$ such that $\inf_{k_i \ge n} \sum_{t=0}^{k_i - 1} \mu_i(k_i)(t) > (Ip_i - 1)/(I-1)$ for any $i \in \mathcal{I}$. Let us define $\underline{S} \equiv \{i \in \mathcal{I} : p_i < 1/I\}$ and $\overline{S} \equiv \{i \in \mathcal{I} : p_i \ge 1/I\}$. For any $i \in \underline{S}$, there exists a $\omega_i > 0$ such that $p_i < 1/(I + \omega_i)$. On the other hand, for any $i \in \overline{S}$, there exists a $\omega_i > 0$ such that $\sum_{t=0}^{k_i-1} \mu_i(k_i)(t) > \{(I+\omega_i)p_i - 1\}/(I-1)$. Let $\omega \equiv \min_{i \in \mathcal{I}} \omega_i$. Define $\psi = I/(I+\omega)$, and $\sigma = 1/(I+\omega)$. It is easy to check that we have $0 < \psi < 1$, and $0 < \sigma < \psi$. Take any $\delta > 0$, and let

$$\varepsilon = \frac{\delta}{\left(1 + \frac{I\psi}{1-\psi}\right)\left(1 + \sum_{i \in \mathcal{I}} \frac{p_i}{1-p_i}\right)^{n-1}\left(\frac{1-\sigma}{1-I\sigma}\right)}.$$

Fix any embedding game $\mathcal{U} \in E(\mathcal{G}, \varepsilon)$. Let $\widehat{E}^{(n)}$ denote $\bigcap_{i \in \mathcal{I}} \widehat{E}_{(i,n)}$. By using ψ and σ, let us inductively construct a sequence of events $(F_{(i,k_i)})_i^{k_i}$ as follows:

$$\begin{aligned} F_{(i,0)} &= B_*^{\psi}(\widehat{E}^{(n-1)}) \cap \widehat{E}^{(n-1)} = B_*^{\psi}(\widehat{E}^{(n)}) = F_0 \text{ for any } i \in \mathcal{I}\ ; \\ F_{(i,k)} &= B_i^{\sigma}(F_{k-1}) \cap F_0, \text{where } F_k = \cap_{i \in \mathcal{I}} F_{(i,k)} \text{ for any } k \in \mathbb{N} \text{ and } i \in \mathcal{I}. \end{aligned}$$

Observe that by Lemma 3.2, we have

$$P(F_0) \geq 1 - \left(1 + \frac{I\psi}{1-\psi}\right)(1 - P(\widehat{E}^{(n-1)})), \text{ and } P(\widehat{E}^{(n-1)}) \geq 1 - \varepsilon\left(1 + \sum_{i\in\mathcal{I}} \frac{p_i}{1-p_i}\right)^{n-1}.$$

Thus

$$P(F_0) \geq 1 - \varepsilon\left(1 + \frac{I\psi}{1-\psi}\right)\left(1 + \sum_{i\in\mathcal{I}} \frac{p_i}{1-p_i}\right)^{n-1}.$$

This $(F_{(i,k_i)})_i^{k_i}$ gives a lower bound for the ex-ante probability of $(E_{(i,k_i)})_i^{k_i}$ under our restriction on $(\mu_i)_{i\in\mathcal{I}}$.

Claim: $F_{(i,k)} \subseteq E_{(i,n+k)}$ for any $k \in \mathbb{Z}_+$ and $i \in \mathcal{I}$.
$\because$) Fix any $i \in \mathcal{I}$. If $i \in \underline{S}$, since $p_i < \sigma$ and $F_{(i,0)} \subseteq E_{(i,n)}$, the claim immediately follows. Consider the case of $i \in \overline{S}$. We show the claim by induction. By definition of $F_{(i,0)}$, we have $F_{(i,0)} \subseteq E_{(i,n)}$ for any $i \in \mathcal{I}$. Suppose $F_{(i,k)} \subseteq E_{(i,n+k)}$ for any $0 \leq k \leq m$ and $i \in \mathcal{I}$. Our goal is to show $F_{(i,m+1)} \subseteq E_{(i,n+m+1)}$ for any $i \in \mathcal{I}$. Suppose, in negation, that there exists a $\theta \in F_{(i,m+1)} \setminus E_{(i,n+m+1)}$ for some $i \in \mathcal{I}$. Then we have

$$\begin{aligned}
& \sum_{0\leq k_j\leq n+m, \forall j\neq i} \mu_i(n+m+1)(t)P(E_{k_{-i}} \mid \pi_i(\theta)) \\
= & \sum_{0\leq k_j\leq n-1, \forall j\neq i} \mu_i(n+m+1)(t)P(E_{k_{-i}} \mid \pi_i(\theta)) + \sum_{n\leq k_j\leq n+m, \forall j\neq i} \mu_i(n+m+1)(t)P(E_{k_{-i}} \mid \pi_i(\theta)) \\
\geq & \sum_{0\leq k_j\leq n-1, \forall j\neq i} \mu_i(n+m+1)(t)P(\widehat{E}^{(n-1)} \mid \pi_i(\theta)) + \sum_{n\leq k_j\leq n+m, \forall j\neq i} \mu_i(n+m+1)(t)P(F_m \mid \pi_i(\theta)) \\
> & [\{(I+\omega)p_i - 1\}/(I-1)] \cdot \psi + [\{I - (I+\omega)p_i\}/(I-1)] \cdot \sigma \\
= & p_i.
\end{aligned}$$

The third inequality follows since $\theta \in B_i^{\psi}(\widehat{E}^{(n-1)})$, $\theta \in F_{(i,m+1)} \subseteq B_i^{\sigma}(F_m)$, and F_t is decreasing in t. Hence, $\theta \in E_{(i,n+m+1)}$, a contradiction.

Since F_k is decreasing in k and $\sum_{i\in\mathcal{I}} \sigma < 1$, the above claim and Lemma 3.1 imply $P(E_{n+k}) \geq P(F_k) \geq P(C^{\sigma}(F_0)) \geq 1 - (1 - P(F_0))\left(\frac{1-\sigma}{1-I\sigma}\right)$ for any $k \in \mathbb{Z}_+$. But then, for any $k \in \mathbb{Z}_+$,

$$P(E_k) \geq P(F_k) \geq 1 - \varepsilon\left(1 + \frac{I\psi}{1-\psi}\right)\left(1 + \sum_{i\in\mathcal{I}} \frac{p_i}{1-p_i}\right)^{n-1}\left(\frac{1-\sigma}{1-I\sigma}\right) > 1 - \delta.$$

By Lemma 4.1, there exists a Bayesian cognitive equilibrium t^* of $\mathcal{U}$ that satisfies $P\big(\{\theta \in \Theta \mid t^*(\mathbf{k}, \theta) = s^*(\mathbf{k})\big) > P(E_k) > 1 - \delta$ for any $\mathbf{k} \in \mathbf{K}$. Therefore, we are done. □

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.

[2] Arad, A., and Rubinstein, A. (2012). "The 11-20 Money Request Game: A Level-k Reasoning Study." *American Economic Review*, 102, 3561-3573.

[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., and Iriberri, N. (2007). "Fatal attraction: Salience, naivete, and sophistication in experimental" Hide-and-Seek" games." *American Economic Review*, 1731-1750.

[9] 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.

[10] 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.

[11] Heifetz, A. and Kets, W. (2012). "All types naive and canny." mimeo.

[12] Kajii, A., Morris, S. (1997). "The Robustness of Equilibria to Incomplete Information." *Econometrica*, 65, 1283-1309.

[13] Kneeland, T. (2012). "Coordination under Limited Depth of Reasoning." mimeo.

[14] Monderer, D., Samet, D. (1989). "Approximating Common Knowledge with Common Beliefs." *Games and Economic Behavior*, 1, 170-190.

[15] Morris, S., Rob, R., and Shin, H.S. (1995). "p-Dominance and belief potential." *Econometrica*, 63, 145-157.

[16] Nagel, R. (1995). "Unraveling in guessing games: An experimental study." *The American Economic Review*, 85, 1313-1326.

[17] Oyama, D., Tercieux, O. (2012). "On the strategic impact of an event under non-common priors." *Games and Economic Behavior*, 74, 321-331.

[18] Penczynski, S. P. (2011). "Strategic Thinking: The Influence of the Game." Working Paper.

[19] Rubinstein, A. (1989). "The Electronic Mail Game: Strategic Behavior Under "Almost Common Knowledge." *The American Economic Review*, 79, 385-391.

[20] 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.

[21] Strzalecki, T. (2010). "Depth of Reasoning and Higher Order Belief." mimeo.

[22] Ui, T. (2001). "Robust equilibria of potential games." *Econometrica*, 69, 1373-1380.