Discussion Paper No. 650

MUTUAL KNOWLEDGE OF RATIONALITY IN THE ELECTRONIC MAIL GAME

Koji Takamiya
and
Akira Tanaka

March 2006

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Mutual knowledge of rationality in the electronic mail game*

Koji TAKAMIYA†
Institute of Social and Economic Research
Osaka University

Akira TANAKA‡
Division of Computer Science
Graduate School of Information Science and Technology
Hokkaido University

This version, Oct 2004

Abstract

This paper reexamines the paradoxical aspect of the electronic mail game (Rubinstein, 1989). The electronic mail game is a coordination game with payoff uncertainty. At a Bayesian Nash equilibrium of the game, players cannot achieve the desired coordination of actions even when a high order of mutual knowledge of payoff functions obtains. We want to make explicit the role of knowledge about rationality of players, not only that of payoff functions. For this purpose, we use an extended version of the belief system model developed by Aumann and Brandenburger (1995). We propose a certain way of embedding the electronic mail game in an belief system. And we show that for rational players to coordinate their actions, for any embedding belief systems, it is necessary that the upper bound order of mutual knowledge of payoff functions exceeds the upper bound order of mutual knowledge of rationality. This result implies that under common knowledge of rationality, the coordination never occurs, which is similar to Rubinstein's result. We point out, however, that there exists a class embedding belief systems for which the above condition is also sufficient for the desired coordination.

JEL Classification— C72, D81, D82

Keywords— electronic mail game, mutual knowledge, common knowledge, rationality, interactive belief system.

---

*This paper is a largely revised version of our paper entitled "The role of small irrationality in communication and strategic decisions: An example with the electronic mail game" (2002). An earlier version was presented at the *International Conference on Game Theory and Its Applications* held at the Taj Mahal hotel in Mumbai, India on 8–10 January of 2003. We thank participants of the above conference, and the seminars at Hokkaido University and Otaru University of Commerce. And the first author thanks Akihiko Matsui for motivating the research.

†Corresponding author. Institute of Social and Economic Research, Osaka University. 6-1 Mihogaoka Ibaraki Osaka 567-0047 JAPAN. E-mail: takamiya@iser.osaka-u.ac.jp.

‡Division of Computer Science, Graduate School of Information Science and Technology, Hokkaido University. Kita14 Nishi9 Sapporo 060-0814 JAPAN.

0 Introduction

The purpose of this research is to study the role of **knowledge about players' rationality** in the context of the **electronic mail game**, which was proposed by Rubinstein (1989). We analyze knowledge of rationality by an extended version of the model called an **interactive belief system** introduced by Aumann and Brandenburger (1995).

The **electronic mail game**, the **e-mail game** for short, (Rubinstein, 1989) is a paradoxical example regarding risky coordination of actions under incomplete information. In this game, two players have to play a coordination game whose payoff matrices are randomly chosen from two possible types, say α and β. Both players has two available actions, say A and B. When the true payoff matrix is α, the players coordinating their actions on (A, A) is profitable to both. But when the true payoff matrix is β, the profitable coordination is (B, B). In both cases, when one player chooses A and the other takes B, the former neither gains nor loses whereas the latter suffers a loss. Thus the coordination on (B, B) is risky. The players share the information about the true payoff matrix, α or β, according to the following "e-mail communication" scenario: Only one of the players is initially informed of the true payoff matrix. Then they share the information of the true payoff matrix by sending signals to each other before taking actions. If the true payoff matrix is α, then no signal is to be sent, and the players will go for taking actions immediately. But if it is β, the informed player (say Player1) sends a signal to the other player (Player2). And if Player2 has received the signal, then she acknowledges the signal by sending another signal to Player1. Then if Player1 has received this signal, he sends back another signal, and so on. Each signal, however, may be lost and never arrives with a small probability. Each player sends signals **automatically**. It is not a strategic decision whether or not one player sends a signal. So once β realizes, both players continue to send signals by turns until one signal gets lost. And once a signal has been lost, the communication stops and the players will take actions. It is interpreted that the more signals the players exchange, the higher is the order (level) of knowledge that "the true payoff matrix is β" the players share. The e-mail game is formulated as a Bayesian game, in which a type of a player is identified with the number of signals that the player has sent in the communication phase. A strategy of a player in this context is a plan of action, which associates one action with each type of the player.

The major problem about the e-mail game is whether the players can achieve the desired coordination (B, B) when the true payoff matrix is β, provided that Player1 is to take the action A when the true payoff matrix is α. (Recall that Player1 is always to know the true payoff matrix.) It should be noted that we are not interested in the case where Player1 chooses B for the payoff matrix α. In this case, even if the players can coordinate on (B, B) for the payoff matrix β, the coordination must be at the expense of the failure of the coordination on (A, A) for the the payoff matrix α. This is clearly undesirable since the payoff matrix α is more likely to occur. Also, when the true payoff matrix is α, taking the action A is weakly dominant to Player1. So there is no benefit to Player1 from taking the action B.

Rubinstein (1989) searched for the possibility of the (B, B) coordination in

Nash equilibrium. He showed that there is only one Nash equilibrium in which Player1 takes the action A when the true payoff matrix is α, and in that equilibrium both players take the action A no matter how many signals have been exchanged. Thus in equilibrium, they cannot coordinate on (B, B) when the payoff matrix is β regardless of the order of mutual knowledge of the payoff matrix as long as it is finite. The (B, B) coordination would be possible if it were common knowledge that the true payoff matrix is β. Thus a finite order mutual knowledge works very differently than common knowledge however high its order is. Rubinstein regarded the above observation as paradoxical. In his paper, he asks: "What would you do if the number on your screen is 17? It is hard to imagine that when L is slightly above M and ε [the probability for a signal to get lost] is small a player will not play B."[1] And he remarks: "Systematic explanation of our intuition that we will play B when the number on our screen is 17 [...] is definitely a most intriguing question."

We want to rethink this e-mail game "paradox" by examining the effect of **knowledge about rationality of players**. Since the "paradox" is concerned with the behavior at a Nash equilibrium, the role of knowledge about rationality is not explicitly considered. The standard assumption of game theory says that it is implicit that players' rationality is common knowledge unless it is specified. However, common knowledge of rationality is highly idealized assumption and hardly satisfied in "real-life" environments. So if we regard the behavior at a Nash equilibrium as counterintuitive, we have to explicitly test the effect of (the lack of) common knowledge of rationality. In this paper, we look at whether the desired coordination behavior is possible when the order of mutual knowledge about rationality is bounded to finite numbers.

To analyze knowledge about rationality, we will use an extended version of an **interactive belief system** as in Aumann and Brandenburger (1995). An interactive belief system is an formal apparatus to analyze players' epistemic states. In this system, each player faces one of possible **types**, which corresponds to a configuration of all the parameters relevant to the player including possible payoff functions, an action to be taken, and knowledge and beliefs. Each player, for each type, has a probability distribution (belief) over the other player's types with which one can derive all information that the player possesses at each type. In particular, a type also specifies whether the player is **rational** or not at that type. (We will explain how we define players' rationality shortly.) Thus one draws one player's information about the rationality of the other player, and information about the other player's information about the rationality of the player, and so on.

In our analysis, players are subject to two kinds of uncertainty: One is **players' rationality**. The other is **payoff functions** (the type of the true payoff matrix). The rationality of players in this context should be concerned with the choice of **strategies** (i.e. plans of action) as of a Bayesian game, rather than the choice of actions, which the players choose in advance of the "e-mail communication." We assumes that players have two distinct sources of information which are independent of each other. One source is the "e-mail communication," where they receive knowledge about payoff functions. From the other source, the players obtain all

[1] In Rubinstein's original scenario, players' types are displayed on their computer screens.

other information, including knowledge about the rationality of the players. This is because since the strategic interaction is "one-shot" in the e-mail game, there is no possibility that a player learns about the other player's characteristics through the interaction. Thus we have to assume that knowledge about players' rationality comes from somewhere else.

To describe the above setting formally, we **embed the e-mail game into an interactive belief system** in the following manner. We split each player's type into two components, **rationality type** and **payoff type**. The payoff types are identical with the types in the e-mail game as a Bayesian game. Also, their probability distributions are the same. Payoff functions at each type profile depends only on the payoff types and are identical with the true payoff matrix at the corresponding type in the e-mail game. Payoff types summarize the information associated with the "e-mail communication."

On the other hand, rationality types represent all the information from somewhere other than the "e-mail communication." Since we assume this information comes from a distinct and independent source, the probability distribution of rationality types and that of payoff types are independent of each other. In the interactive belief system, one type of a player is associated with one **action**. But one rationality type is associated with one **strategy** (since without the information represented by payoff types, players cannot pin down to a single action.)

The standard definition says that a player is called **rational** at a type if he maximizes the expectation of his payoff with respect to the belief at that type. Thus in this definition, rationality is a question of whether the **action** to be taken at a certain type is payoff-maximizing or not. However, as we have mentioned above, for our purpose, we want to call a player rational if his choice of **strategies**, rather than that of actions, is payoff-maximizing. Thus the relevant definition of rationality should be contingent only upon rationality types. This explains why we call them "rationality types." We define that a player is called **rational** at a type if he is rational in the sense of the standard definition at every payoff type given the rationality type that he is at now.

In short, players' types consist of two independent components: rationality types contain the information about the choice of strategies and so about rationality, whereas payoff types contain the information about the true payoff matrix and actions resulting from strategies.

The results of our analysis suggests the "paradox" of the e-mail game at least partially results from the strength of common knowledge of rationality. The first result of this paper says that **in any interactive belief system embedding the e-mail game, when the upper bound order of mutual knowledge of rationality is at least as large as that of mutual knowledge of the true payoff matrix, the desired coordination on** (B, B) **by rational players must fail**. Thus when common knowledge of rationality (which has the infinite order) is assumed, the paradoxical behavior as observed in Rubinstein (1989) follows. But our second result points out that **there are some interactive belief systems in which when the above condition does not hold, rational players do coordinate their action on** (B, B). Since the order of mutual knowledge of the true payoff matrix is determined by how many signals has been exchanged in the "e-mail communication," these results suggest that the desired coordination may

occur if a particular amount of communication has taken place, and that it does not occur if not. We consider that this observation fits our intuition of how we would actually play this game in Rubinstein's original scenario.

The direction of research that we take was first proposed by Aumann (1992). He used a model called an **information system**, which is a special case of an interactive belief system. (In information systems, payoff functions do not vary across types.) With his model, Aumann examined implications of replacing common knowledge of rationality with mutual knowledge of rationality of finite orders. By an example based on a version of the "centipede game" (Rosenthal 1982, Megiddo 1986), he demonstrated that a minute departure from common knowledge of rationality may resolve the "backward induction paradox." Basically, we follow the same method as in Aumann's paper in analyzing players' rationality in the e-mail game. However, the major departure of our analysis from Aumann's is that we analyze an incomplete information game (the e-mail game) whereas Aumann dealt with strategic games. This is the reason why we need a more elaborate model. Further, more importantly, we can analyze the **relationship** between two kinds of knowledge (payoff functions and rationality).

To the best of our knowledge, most of the preceding literature on the e-mail game following Rubinstein (1989) further examines implications of the lack of common knowledge of payoff functions by considering variations on (or generalization of) the original game (such as Binmore and Samuelson (2001), Dimitri (2003), (2004), Morris (2001)). However, most of them do not directly refer to players' rationality.[2] On the contrary, we will not change the original setting of the game at all. And we explicitly consider epistemic states of players' rationality. We consider that the present work offers a direct resolution to the "paradox" of the e-mail game.

The plan of the paper is as follows: Section 1 introduces the basic concepts, including the formal presentation of the e-mail game, interactive belief systems, and embedding of the e-mail game into an interactive belief system. Section 2 states and proves the results. Section 3 concludes the paper with some remarks.

1 Basic concepts

1.1 The Electronic mail game

The **electronic mail game** (the **e-mail game**, henceforth) was introduced by Rubinstein (1989). Given the underlying scenario provided in Introduction, here we present only a formal description. (See Rubinstein's paper for details.) The game is formulated as a Bayesian game, $\mathcal{G} = (A_1, A_2; T_1, T_2; u_1, u_2; \rho)$. There are two players, Players1 and 2. Each Player i ($i = 1, 2$) has two available **actions**, namely A and B. A_i denotes the action set $\{A, B\}$.

For each player, a **payoff type** is a non negative integer $0, 1, 2, \cdots$.[3] That Player i's payoff type equals t_i means that when the "e-mail communication" has

[2] One exception is Dulleck (2002), which introduced boundedly rational players who cannot count the number of signals correctly.

[3] We avoid the term "types" here not to arise confusion with "types" in the context of "interactive belief systems" to be introduced in Section 1.2.

ended, he has sent signals t_i times. T_i denotes the set of payoff types of Player i. That is, $T_i = \{0, 1, 2, \cdots\}$. Denote by ρ the probability distribution over the set of payoff type profiles, which is depicted in Table 2. Note that for any payoff type profile (t_1, t_2), the probability of (t_1, t_2) is positive if and only if $t_1 = t_2$ or $t_1 = t_2 + 1$. ε is the probability with which an e-mail signal gets lost. Assume $p, \varepsilon < \frac{1}{2}$.

There are two possible types of payoff matrices, α and β. These payoff matrices are depicted in Table 1. α occurs with probability $1-p$, and β does with p. Assume $L > M > 0$. And the payoff type profile (t_1, t_2) equals $(0, 0)$ if and only if the true payoff matrix is α. And any other payoff type profile having a positive probability realizes only if the true payoff matrix is β. Since our analysis is in expectation term throughout, the payoff matrices associated with the payoff type profiles with 0-probability play no role so that they can be arbitrary.

u_i $(i = 1, 2)$ is the payoff function of Player i, $u_i : A_1 \times A_2 \times T_1 \times T_2 \to \mathcal{R}$. Given the relationship between the payoff types and the two payoff matrices, abusing the notation, we sometimes regard u_i as a function defined on $A_1 \times A_2 \times \{\alpha, \beta\}$.

A **strategy** of Player i is a function from T_i to the set of actions A_i.

It is clear from the payoff matrices that the two players want to coordinate their actions on (A, A) when the true payoff matrix is α, and on (B, B) when it is β. We are interested in the problem of whether the players can achieve the desired coordination (B, B) when the true payoff matrix is β, provided that Player1 is to take the action A when the true payoff matrix is α. We will not be concerned with the case where Player1 chooses B for the payoff matrix α. In this case, even if the players can coordinate on (B, B) for the payoff matrix β, the coordination must be at the expense of the failure of the coordination on (A, A) for the the payoff matrix α. This is clearly undesirable since the payoff matrix α is more likely to occur. Also since in the payoff matrix α A is the weakly dominant action to Player1, there is no benefit to Player1 from taking B.

Rubinstein (1989) proved that a strategy profile (σ_1, σ_2) is a Bayesian Nash equilibrium and satisfies $\sigma_1(0) = A$ if and only if for each $i = 1, 2$ and each $t = 0, 1, 2, \cdots$, $\sigma_i(t) = A$.

Table 1: The payoff matrices of the e-mail game

α
probability $(1-p) > \frac{1}{2}$

	A	B
A	M, M	$0, -L$
B	$-L, 0$	$0, 0$

β
probability $p < \frac{1}{2}$

	A	B
A	$0, 0$	$0, -L$
B	$-L, 0$	M, M

Table 2: The probability distribution ρ over the set of payoff type profiles $T_1 \times T_2$

Payoff types of Player1/Payoff types of Player2

	0	1	2	3	$\cdots$
0	$(1-p)$				
1	εp	$\varepsilon(1-\varepsilon)p$			
2		$\varepsilon(1-\varepsilon)^2 p$	$\varepsilon(1-\varepsilon)^3 p$		
3			$\varepsilon(1-\varepsilon)^4 p$	$\varepsilon(1-\varepsilon)^5 p$	
$\vdots$				$\ddots$	$\ddots$

1.2 Interactive belief system

An interactive belief system is formal apparatus to analyze players' knowledge and beliefs. The following is an extension of the definition by Aumann and Brandenburger (1995). Here we consider only two person cases. An **interactive belief system** is a list $(A_1, A_2; S_1, S_2; \boldsymbol{a}_1, \boldsymbol{a}_2; \boldsymbol{g}_1, \boldsymbol{g}_2; \mu)$. Here A_i $(i = 1, 2)$ are the **action sets**, which are nonempty. S_i $(i = 1, 2)$ is the set of **types** of Player i. A type of Player i represents a specification of all the relevant parameters to Player i. Assume S_i is a nonempty countable set.

Call any element of $S_1 \times S_2$ a **state of the world** (**SOW**, for short). Denote a SOW (s_1, s_2) by s, (s_1', s_2') by s' and so on. Denote by Ω the set $S_1 \times S_2$.

$\boldsymbol{a}_i$ is a function from S_i to A_i $(i = 1, 2)$. This means that Player i takes the action $\boldsymbol{a}_i(s_i)$ when his type is s_i.

$\boldsymbol{g}_i$ $(i = 1, 2)$ is a function from Ω to the set of real-valued functions defined on $A_1 \times A_2$. (Aumann and Brandenburger define the domain of $\boldsymbol{g}_i$ to be S_i, not Ω. This is the only difference to our formulation.) For each $s \in \Omega$, $\boldsymbol{g}_i(s)$ represents the payoff function of Player i when the SOW is s.

μ is a **probability distribution** over Ω. We postulate the **common prior assumption**.[4] One derives each player's knowledge and beliefs at each type from this single probability distribution. Call a SOW **possible** if it is assigned non-zero probability by μ. Denote by $\Omega^\star$ the set of the possible SOWs. For the sake of simplicity, assume that for any $i = 1, 2$ and for any $s_i \in S_i$, μ assigns the set $\{s_i\} \times S_j$, where $i \neq j$, a non-zero probability.

Since (Ω, μ) is a (countable) probability space, we can regard the functions $\boldsymbol{a}_i, \boldsymbol{g}_i$ as random variables.

Let us introduce concepts of interactive knowledge and beliefs. Let E be a subset of Ω. Call E an **event**. Let $s \in \Omega$. Then say that at s, Player i **believes** E with probability $\mu(E \mid s_i)$. This defines players' **belief**. In particular, if $\mu(E \mid s_i) = 1$, we say that Player i **knows** E at s. This defines players' **knowledge**.[5]

[4] Aumann and Brandenburger do not include the common prior assumption in the definition. There is a numerous papers on the reasonability of the assumption. We do not intend to argue about it here.

[5] We do not treat knowledge in the sense of absolute certainty.

These definitions allow us to say like "Player i knows (or believes) E at s_i" with $s_i \in S_i$.

Let E be an event. Denote the event "Player i knows an event E" by K_iE, i.e. $K_iE = \{s \in \Omega \mid \mu(E \mid s_i) = 1\}$. Let K^1E denote $K_1E \cap K_2E$. Also let K^2E denote K^1K^1E, K^3E denote K^1K^2E, and so on.

Let E be an event, and $s \in \Omega$. Let m be a positive integer. Call E **m-th order mutual knowledge** at s if $s \in K^mE$. If E is m-th order mutual knowledge with $m \geq 2$ at s, then E is also $(m-1)$-th order mutual knowledge at s (Lemma 5 below). Suppose E is mutual knowledge of some order at s. Then we say that m is the **upper bound order of mutual knowledge of E at s** if $s \in K^mE$ and $s \notin K^{m+1}E$. Denote by $\#MKE(s)$ the upper bound order of mutual knowledge of E at s. For convenience, define $\#MKE(s) = 0$ if E is not mutual knowledge of any order at s. If E is m-th order mutual knowledge for any positive integer m at s, we say that E is **common knowledge** at s. Denote by CKE the event "E is common knowledge." That is, $CKE = K^1E \cap K^2E \cap K^3E \cap \cdots$. For the case $s \in CKE$, define $\#MKR(s) = \infty$.

Let $s \in \Omega$. Call Player i **rational at** s if the action $\boldsymbol{a}_i(s_i)$ that the player takes at s maximizes the expectation of his payoff given his posterior, that is, if it holds true that $\boldsymbol{a}_i(s_i) \in \arg\max_{a_i \in A_i} \boldsymbol{E}\boldsymbol{g}_i(a_i, \boldsymbol{a}_j \mid s_i)$.[6]

For our purpose, we need to extend the above definition of rationality to the following one. Let a partition $\mathcal{Q}$ of Ω be given. Let $s \in \Omega$. Call Player i **rational at s relative to $\mathcal{Q}$** if he is rational is at any SOW in the cell of $\mathcal{Q}$ which contains s. This definition of rationality is stronger than the one in the above. So to speak, the first definition is "pointwise," and the second one is "setwise." In the next subsection, we will see the "setwise" definition of rationality is relevant to our purpose.

Note that by these definitions, the rationality of a player is contingent only upon his types (regardless of "pointwise" or "setwise"). Thus each player knows whether he is rational or not at any type. Also we may say like "Player i is rational (or rational relative to $\mathcal{Q}$) at s_i" with $s_i \in S_i$.

We give some preliminary lemmas. Proofs of Lemmas 1 and 2 are straightforward thus omitted.

Lemma 1: *Let $\{E_j\}$ be a family of events. $\bigcap_j K_iE_j = K_i(\bigcap_j E_j)$.*

Lemma 2: *If $E \subset F$, then $K_iE \subset K_iF$.*

It is immediate from Lemma 1 that for any positive integer m, $\bigcap_j K^mE_j = K^m(\bigcap_j E_j)$.

Note that since we have defined "knowledge" to be "belief with probability 1," in general it is not true that $K_iE \subset E$ (i.e. if Player i knows E, then E is true). But it holds true that $K_iE \cap \Omega^\star \subset E \cap \Omega^\star$.

There is a class of events such that $K_iE = E$ (i.e. Player i knows E if and only if E is true). For $i = 1, 2$, let $\mathcal{P}_i$ be the partition of Ω such that any two SOW s and s' are in the same cell of $\mathcal{P}_i$ if and only if $s_i = s_i'$. Denote by $\mathcal{F}_i$ the σ-field

[6]Here and in the sequel, $\boldsymbol{E}$ is the operator of expectation.

generated by $\mathcal{P}_i$. (i.e. $\mathcal{F}_i$ consists of all the unions of elements of $\mathcal{P}_i$.)

Lemma 3: *If $E \in \mathcal{F}_i$, then $K_iE = E$.*

Proof: (i) We prove $K_iE \subset E$. Suppose that there exists a SOW s such that $s \in K_iE$ and $s \notin E$. Then the former implies $\mu(E \mid s_i) = 1$. The latter implies $(\{s_i\} \times S_j) \cap E = \emptyset$, where $i \neq j$, since $E \in \mathcal{F}_i$. This implies $\mu(E \mid s_i) = 0$. This is a contradiction.

(ii) We prove $E \subset K_iE$. Suppose that there exists a SOW s such that $s \in E$ and $s \notin K_iE$. Then the latter implies $\mu(E \mid s_i) \neq 1$. The former implies $(\{s_i\} \times S_j) \cap E = \{s_i\} \times S_j$, where $i \neq j$, since $E \in \mathcal{F}_i$. This implies $\mu(E \mid s_i) = 1$. This is a contradiction.
□

Important examples of such events E that $E \in \mathcal{F}_i$ are the event "Player i is rational," and the event "Player i takes the action A (or B)." Thus Player i knows he is rational if and only if he is actually rational. And Player i knows he takes the action A (or B, resp.) if and only if he actually takes A (or B, resp.). Also it is important to note that for any event E, $K_iE \in \mathcal{F}_i$. Then the following lemma is immediate.

Lemma 4: $K_iE = K_iK_iE$.

Lemma 5: *Let $m \geq 2$. $K^mE \subset K^{m-1}E$.*

Proof: By induction regarding m.
(i) Let $m = 2$. $K^2E = K^1(K_1E \cap K_2E)$. Then by Lemma 1, this equals $K_1K_1E \cap K_1K_2E \cap K_2K_1E \cap K_2K_2E$. By Lemma 4, this equals $K_1E \cap K_2E \cap K_1K_2E \cap K_2K_1E$. Clearly this is a subset of $K_1E \cap K_2E = K^1E$.

(ii) Suppose that for some $k \geq 2$, $K^kE \subset K^{k-1}E$. We have $K^{k+1}E = K^1K^kE = K_1K^kE \cap K_2K^kE$. Then by the induction hypothesis and Lemma 2, this is a subset of $K_1K^{k-1}E \cap K_2K^{k-1}E$, which equals K^kE. Thus we have $K^{k+1}E \subset K^kE$.
□

Lemma 6: $K^mCKE = CKE$.

Proof: $K^mCKE = K^m(K^1E \cap K^2E \cap \cdots)$. Lemma 1 implies this equals $K^{m+1}E \cap K^{m+2}E \cap \cdots$. Lemma 5 implies this equals $K^1E \cap K^2E \cap \cdots = CKE$. □

The following lemma will be useful in proving some of the results to appear in the next section. To state the lemma, we need one more definition. Let $s, s' \in \Omega^\star$. A **path** from s to s' is a sequence $(s^\nu)_{\nu=0}^m$ with each $s^\nu \in \Omega^\star$ satisfying
(i) $s^0 = s$ and $s^m = s'$; and
(ii) for some $i = 1$ or 2, letting $i \neq j$, $s_i^0 = s_i^1, s_j^1 = s_j^2, s_i^2 = s_i^3, s_j^3 = s_j^4 \cdots$.
Call the number m the **length of the path**. Note that there are infinitely many

paths from s to s', for any given s and s' in $\Omega^\star$.[7] Also note that by definition a path must lie within the set of the **possible** SOWs. Let $s \mapsto s'$ denote a path from from s to s'. Denote by $\mid s \mapsto s' \mid$ the length of $s \mapsto s'$.

Lemma 7: *Let E be an event. And let $s \in \Omega^\star$. Then $s \in K^m E$ if and only if for any $s' \in \Omega^\star \setminus E$, there does not exist $s \mapsto s'$ with $\mid s \mapsto s' \mid \leq m$.*

Proof: By induction regarding m.
(i) Let $m = 1$. Let $s \in \Omega^\star$.

(if) Suppose $s \notin K^1 E$. Then for some $i = 1, 2$, $s \notin K_i E$. Thus there exists some $s' \in \Omega^\star \setminus E$ such that $s'_i = s_i$ and $\mu(s' \mid s_i) > 0$. Then (s, s') is a path whose length is 1.

(only if) Assume $s \in K^1 E$. Suppose that for some $s' \in \Omega^\star \setminus E$, there exists $s \mapsto s'$ with $\mid s \mapsto s' \mid \leq 1$. Then since $s \neq s'$, $\mid s \mapsto s' \mid = 1$. Then by the definition of paths, $\mu(s' \mid s_i) > 0$ for some $i = 1, 2$. Thus $\mu(E \mid s_i) \neq 1$, that is, $s \notin K_i E$. This implies $s \notin K^1 E$. This is a contradiction.

(ii) Suppose that for some $k \geq 1$, for any event E and any $s \in \Omega^\star$, we have $s \in K^k E$ if and only if for any $s' \in \Omega^\star \setminus E$, there does not exist $s \mapsto s'$ with $\mid s \mapsto s' \mid \leq k$.

(if) Let $s \in \Omega^\star$. Suppose $s \notin K^{k+1} E$. Note that $K^{k+1} E = K^1(K^k E)$. Then by the case (i), for some $\hat{s} \in \Omega^\star \setminus K^k E$, there exists some path $s \mapsto \hat{s}$ with the length 1. Also by the induction hypothesis, $\hat{s} \in \Omega^\star \setminus K^k E$ implies that for some $s' \in \Omega^\star \setminus E$, there exists some path $\hat{s} \mapsto s'$ such that $\mid \hat{s} \mapsto s' \mid \leq k$. This implies that there exists some path $s \mapsto s'$ such that $\mid s \mapsto s' \mid \leq k + 1$.

(only if) Let $s \in \Omega^\star$. Assume $s \in K^{k+1} E$. Suppose that for some $s' \in \Omega^\star \setminus E$, there exists some path $s \mapsto s'$ such that $\mid s \mapsto s' \mid \leq k + 1$. Denote this path by $(s^0, s^1, \cdots, s^l)$, where $s^0 = s$ and $s^l = s'$. Now note that $K^{k+1} E = K^1(K^k E)$, and (s^0, s^1) is a path with the length 1. Then the case (i) implies $s^1 \in K^k E$. Then $(s^1, s^2, \cdots, s^l)$ is a path from $s^1 \in K^k E$ to $s' \in \Omega^\star \setminus E$ with the length $l - 1$, which is not more than k. This contradicts the induction hypothesis.
□

The following Lemma 8 is immediate from Lemma 7 but worth noting.

Lemma 8: *Let E be an event, and $s, s' \in \Omega^\star$. Then if $s \in CKE$ and there exists some $s \mapsto s'$, then $s' \in CKE$.*

Proof: Denote $\mid s \mapsto s' \mid$ by m. Suppose $s' \notin CKE$. Then Lemma 7 implies $s \notin K^m CKE$. By Lemma 6, we have $s \notin CKE$. □

[7] For example, let $s = s'$. Then (s), (s, s), $(s, s, s) \cdots$ are all different paths from s to s'.

1.3 Embedding the e-mail game in an interactive belief system

In our analysis, we are concerned with two kinds of players' knowledge. One is knowledge about **payoff functions** which is implicit in the formulation of the e-mail game. The other is knowledge about **rationality of players**. We assume that the two kinds of knowledge are elicited from distinct and independent sources of information, respectively, for the following reason. We suppose that the rationality in this context is concerned with the choice of strategies (plans of action), rather than the choice of actions, which the players make in advance of the "e-mail communication." And the "e-mail communication" is an "automatic device" so the strategic interaction is one-shot, thus it has to be assumed that each player obtains information about rationality independent of the "e-mail communication." The following formulation reflects our viewpoint.

Let the e-mail game $\mathcal{G} = (A_1, A_2; T_1, T_2; u_1, u_2; \rho)$ be given. Then an **interactive belief system which embeds** $\mathcal{G}$ (**embedding belief system**, for short) is an interactive belief system $\mathcal{B}(\mathcal{G}) = (A_1, A_2; S_1, S_2; \boldsymbol{a}_1, \boldsymbol{a}_2; \boldsymbol{g}_1, \boldsymbol{g}_2; \mu)$. Here A_i $(i = 1, 2)$ are taken from $\mathcal{G}$.

$S_i = T_i^\star \times T_i$ $(i = 1, 2)$, where T_i is taken from $\mathcal{G}$, and $T_i^\star$ is a nonempty countable set. Call an element of $T_i^\star$ a **rationality type**, and that of T_i a **payoff type**, of Player i. Recall that Ω denotes $S_1 \times S_2$. Thus $\Omega = T_1^\star \times T_1 \times T_2^\star \times T_2$. Also denote by T the set $T_1 \times T_2$, and by $T^\star$ the set $T_1^\star \times T_2^\star$. So we denote $(t_1^\star, t_1, t_2^\star, t_2) \in \Omega$ by s, $(t_1^{\star\prime}, t_1', t_2^{\star\prime}, t_2')$ by s' and so on. And similarly denote $t = (t_1, t_2), t' = (t_1', t_2') \in T_1 \times T_2$, $t^\star = (t_1^\star, t_2^\star), t^{\star\prime} = (t_1^{\star\prime}, t_2^{\star\prime}) \in T_1^\star \times T_2^\star$ and so on.

Let $\boldsymbol{t}_i$ $(i = 1, 2)$ denote the functions from Ω to T_i that pick up for each $s \in \Omega$ the component t_i, i.e. $\boldsymbol{t}_i(s) = t_i$ for any $s \in \Omega$. Also, we define functions $\boldsymbol{t}_i^\star : \Omega \to T_i^\star$ $(i = 1, 2)$ similarly. Then $\boldsymbol{t}_i$, $\boldsymbol{t}_i^\star$ are regarded as random variables.

Given ρ of $\mathcal{G}$, the probability distribution μ satisfies the following condition:

(1.1) There exists a probability distribution $\rho^\star$ on $T^\star$ that satisfies for any $s \in \Omega$, $\mu(s) = \rho^\star(t^\star)\rho(t)$.

That is, a rationality type profile and a payoff type profile are independent of each other when viewed as random variables.

The functions $\boldsymbol{g}_i$ $(i = 1, 2)$ satisfy the following.

(1.2) For any $s \in \Omega$, if $\rho(t) > 0$, then $\boldsymbol{g}_i(s) = u_i(\cdot, \cdot, t_1, t_2) : A_1 \times A_2 \to \mathcal{R}$.

That is, rationality types are payoff-irrelevant. And the payoff functions u_i of $\mathcal{G}$ are essentially inherited. Given (1.2), by the definition of u_i, $(\boldsymbol{g}_1, \boldsymbol{g}_2)$ takes only two values, namely $(u_1(\cdot, \cdot, \alpha), u_2(\cdot, \cdot, \alpha))$ and $(u_1(\cdot, \cdot, \beta), u_2(\cdot, \cdot, \beta))$, for any SOW s such that $\rho(t) > 0$. For the former case, say **"the true payoff matrix is** α**,"** and for the latter case, say **"the true payoff matrix is** β**."** Note that these expressions do not apply to such SOWs s that $\rho(t) = 0$. We are not concerned with the values of $(\boldsymbol{g}_1, \boldsymbol{g}_2)$ for such SOWs. Because they are not possible SOWs, and our analysis is in terms of expectation throughout.

Let $\mathcal{Q}^\star$ be the partition of Ω such that any two SOWs s and s' are in the same

cell of $\mathcal{Q}^\star$ if and only if $t^\star = t^{\star\prime}$. Denote by $\mathcal{F}^\star$ the σ-field generated by $\mathcal{Q}^\star$. Then $E \in \mathcal{F}^\star$ means that E is an event which is contingent only upon the component $t^\star$ of SOWs s. In other words, such events E are those events that are determined independent of the "e-mail communication." The condition (1.1) implies that for any event E, $E \in \mathcal{F}^\star$ implies $K_i E \in \mathcal{F}^\star$ thus $K^m E, CKE \in \mathcal{F}^\star$.

The rationality concept relevant to this context is the "setwise" one that we have introduced in Section 1.2. In the following, we are concerned with each player's rationality relative to $\mathcal{Q}^\star$. So unless noted, we say that Player i is **rational at** s meaning that Player i is **rational at** s **relative to** $\mathcal{Q}^\star$. Thus whether Player i is rational or not at a SOW s depends only on the component $t_i^\star$ of s. So we say like "Player i is rational at $t_i^\star$ (or at $t^\star$)." This definition of rationality explains why we call elements of $T_i^\star$ "rationality types." When we want to refer to the "pointwise" definition, we will use phrases such as "pointwise rationality."

The reason for adopting the above definition of rationality should be clear. As mentioned, we do not consider rationality to be an attribute of actions but of strategies (plans of action) in this context. So it should be determined independent of the "e-mail communication." Thus the relevant definition of rationality has to be independent of the component t of SOWs s.

In the sequel, we are interested in three kinds of events: The first is the event that **"the true payoff matrix is** β**."** Denote this event by P_β. Formally, $P_\beta = \{s \in \Omega \mid \rho(t) > 0 \text{ and } (t_1, t_2) \neq (0,0)\}$. The second is the event that **"both players are rational."** Denote this event by R. Note that $R \in \mathcal{F}^\star$. The third is the event "if Player1's payoff type is 0, then he chooses the action A." Denote this event by $\tilde{E}$. Formally, $\tilde{E} = \{s \in \Omega \mid t_1 = 0 \Rightarrow \boldsymbol{a}_1(s_1) = A\}$. Note that $\tilde{E} \in \mathcal{F}^\star$. Significance of P_β and R is evident. The motivation for introducing $\tilde{E}$ is explained after stating Theorem 1 in the next section.

2 Results

2.1 General results

Throughout this subsection, let an embedding belief system $\mathcal{B}(\mathcal{G})$ be given. Lemma 9 below formally states the intended interpretation of the "e-mail communication" of the e-mail game. (See Rubinstein (1989).)

Lemma 9: *Let $s \in \Omega^\star$. If $t_1 \geq 1$, then $\#MKP_\beta(s) = t_1 + t_2 - 1$.*

Proof: Index those elements of T that ρ gives positive probabilities as follows: Let $t \in T$. Let $h = t_1 + t_2$. Then refer to t as $t^{(h)}$.

Let $s \in \Omega^\star$. Then say t's index is h i.e. $t = t^{(h)}$. Consider a path $(s^\nu)_{\nu=0}^h$ such that

$s^0 = s$,
$s^1 = (t_1^\star, t_1^{(h-1)}, t_2^\star, t_2^{(h-1)})$,
$s^2 = (t_1^\star, t_1^{(h-2)}, t_2^\star, t_2^{(h-2)})$,
$\vdots$
$s^h = (t_1^\star, t_1^{(0)}, t_2^\star, t_2^{(0)})$.

Note that $t^{(0)} = (0,0)$. Then clearly, $s^0, s^1, \cdots, s^{h-1} \in P_\beta$ and $s^h \notin P_\beta$. Clearly, this is the shortest path from s to some possible SOW outside P_β, and its length is h. Then Lemma 7 implies $\#MKP_\beta(s) = h - 1 = t_1 + t_2 - 1$. □

Lemma 10: *Let $s \in \Omega$. Then*
(i) if Player1 is rational at s, $t_1 \geq 1$ and $\boldsymbol{a}_1(s_1) = B$, then there exists some $s' \in \Omega^\star$ such that $s_1 = s'_1$, $t'_2 = t_1 - 1$ and $\boldsymbol{a}_2(s'_2) = B$; and
(ii) if Player2 is rational at s, $\boldsymbol{a}_1(s_2) = B$, then there exists some $s' \in \Omega^\star$ such that $s_2 = s'_2$, $t'_1 = t_2$ and $\boldsymbol{a}_1(s'_1) = B$.

Proof:
Proof of (i): Let $s \in \Omega$ satisfy the assumptions in the statement. Suppose that for any $s' \in \Omega^\star$ such that $s_1 = s'_1$ and $t'_2 = t_1 - 1$, we have $\boldsymbol{a}_2(s'_2) = A$. Then since we assumed $\boldsymbol{a}_1(s_1) = B$, we have $\boldsymbol{Eg}_1(\boldsymbol{a}_1, \boldsymbol{a}_2 \mid s_1) = \frac{1}{1+(1-\varepsilon)} u_1(B, A; \beta) + \frac{1-\varepsilon}{1+(1-\varepsilon)} \sum_{\tilde{t}_2^\star \in T_2^\star, \tilde{t}_2 = t_1} u_1(B, \boldsymbol{a}_2(\tilde{t}_2^\star, \tilde{t}_2); \beta) \rho^\star(\tilde{t}_2^\star \mid t_1) \leq \frac{1}{1+(1-\varepsilon)}(-L) + \frac{1-\varepsilon}{1+(1-\varepsilon)} M < 0$. On the other hand, $\boldsymbol{Eg}_1(A, \boldsymbol{a}_2 \mid s_1) = 0$. This contradicts the rationality of Player1.

Proof of (ii): Let $s \in \Omega$ satisfy the assumptions in the statement. We have two cases, namely the case $t_2 \geq 1$ and the case $t_2 = 0$.

(a) Assume $t_2 \geq 1$. Suppose that for any $s' \in \Omega^\star$ such that $s_2 = s'_2$ and $t'_1 = t_2$, we have $\boldsymbol{a}_1(s'_1) = A$. Then since we assumed $\boldsymbol{a}_2(s_2) = B$, we have $\boldsymbol{Eg}_2(\boldsymbol{a}_1, \boldsymbol{a}_2 \mid s_2) = \frac{1}{1+(1-\varepsilon)} u_2(A, B; \beta) + \frac{1-\varepsilon}{1+(1-\varepsilon)} \sum_{\tilde{t}_1^\star \in T_1^\star, \tilde{t}_1 = t_2+1} u_2(\boldsymbol{a}_1(\tilde{t}_1^\star, \tilde{t}_1), B; \beta) \rho^\star(\tilde{t}_1^\star \mid t_2) \leq \frac{1}{1+(1-\varepsilon)}(-L) + \frac{1-\varepsilon}{1+(1-\varepsilon)} M < 0$. On the other hand, $\boldsymbol{Eg}_2(\boldsymbol{a}_1, A \mid \tilde{s}_2) = 0$. This contradicts the rationality of Player2.

(b) Assume $t_2 = 0$. Suppose that for any $s' \in \Omega^\star$ such that $s_2 = s'_2$ and $t'_1 = t_2$, we have $\boldsymbol{a}_1(s'_1) = A$. Then since we assumed $\boldsymbol{a}_2(s_2) = B$, we have $\boldsymbol{Eg}_2(\boldsymbol{a}_1, \boldsymbol{a}_2 \mid s_2) = \frac{1-p}{(1-p)+\varepsilon p} u_2(A, B; \alpha) + \frac{\varepsilon p}{(1-p)+\varepsilon p} \sum_{\tilde{t}_1^\star \in T_1^\star, \tilde{t}_1 = t_2+1} u_2(\boldsymbol{a}_1(\tilde{t}_1^\star, \tilde{t}_1), B; \beta) \rho^\star(\tilde{t}_1^\star \mid t_2) \leq \frac{1-p}{(1-p)+\varepsilon p}(-L) + \frac{\varepsilon p}{(1-p)+\varepsilon p} M < 0$. (Recall that $p < \frac{1}{2}$.) On the other hand, $\boldsymbol{Eg}_2(\boldsymbol{a}_1, A \mid \tilde{s}_2) = 0$. This contradicts the rationality of Player2.
□

The following is the first of our main results.

Theorem 1: *Let $s \in \Omega^\star$. Assume that each player is rational at s. And assume $s \in CK\tilde{E}$. Then if $\boldsymbol{a}_i(s) = B$ for each $i = 1, 2$, then $\#MKP_\beta(s) > \#MKR(s)$.*

Note that Theorem 1 is concerned only with what happens at **possible** SOWs. In Theorem 1, we assume that each player is rational at s. This assumption can be replaced with the requirement that s is such that $t_1, t_2 \geq 1$. (If replaced, however, we would be silent about the case $t = (1,0)$.) Because, if s satisfies $t_1, t_2 \geq 1$, then $\#MKP_\beta(s) \geq 1$. And if at least one player is not rational at s, then $\#MKR(s) = 0$. So we have $\#MKP_\beta(s) > \#MKR(s)$ trivially. Thus in the case of $t_1, t_2 \geq 1$, it is superfluous to assume that both players are rational.

In words, Theorem 1 says that at any possible SOW, provided that it is common knowledge that [Player1 chooses the action A if the true payoff matrix is α], it

is a necessary condition for rational players to achieve the (B, B) coordination that [the upper bound order of mutual knowledge of "the true payoff matrix is β" exceeds the upper bound order of mutual knowledge of rationality].

The motivation for the assumption $s \in CK\tilde{E}$ in Theorem 1 should be made clear. As we have discussed, we are interested in whether players can coordinate their actions on (B, B) when the true payoff matrix is β provided that Player1 chooses A for the payoff matrix α. Thus we exclude the cases where there is a possibility that Player1 takes the action B when the true payoff matrix is α. This entails common knowledge of the event $\tilde{E}$.

The significance of Theorem 1 is that it relates the possibility of the (B, B) coordination to the comparison between the orders of mutual knowledge about players' rationality and the true payoff matrix. In particular, the result opens the possibility for the (B, B) coordination as a result of rational plans of action when the order of mutual knowledge of rationality is bounded. We search for this possibility in the next subsection.

It should be noted that the condition that Theorem 1 gives is necessary but not sufficient for the (B, B) coordination. The following counterexample shows this.

Example 1: Consider the embedding belief system satisfying the following: $T_1^\star = \{1\}$ and $T_2^\star = \{1, 2\}$. The probability distribution $\rho^\star$ over $T^\star$ assigns a probability $1 - \epsilon$ to $(1, 1)$ and ϵ to $(1, 2)$, where $\epsilon > 0$ is negligibly small. ($\rho^\star$ is as depicted in Table 3. The row is the rationality type of Player1 whereas the columns are those of Player2.) For each player i, $\boldsymbol{a}_i(1, t_i) = A$ for all $t_i \in T_i$. And $\boldsymbol{a}_2(2, t_2) = B$ for all $t_2 \in T_2$. Then Player1 is rational everywhere. And Player2 is rational at $(1, 1)$, and irrational at $(1, 2)$. Thus $\#MKR(s) = 0$ for any $s \in \Omega^\star$. Because Player1 is never sure that Player2 is rational. On the other hand, Player1' s rationality is common knowledge.[8] Note that for any $s \in \Omega$, $s \in CK\tilde{E}$. Now look at any $s \in \Omega^\star$ satisfying $t^\star = (1, 1)$ and $t_1, t_2 \geq 1$. It is satisfied that $\#MKP_\beta(s) > \#MKR(s)$. But neither player takes the action B. □

Table 3: A counterexample

	1	2
1	$1 - \epsilon$	ϵ

Proof of Theorem 1: Index those elements of T that ρ gives positive probabilities as follows: Let $t \in T$. Let $h = t_1 + t_2$. Then refer to t as $t^{(h)}$.

Let $s \in \Omega^\star$.

(i) Assume $t_1, t_2 \geq 1$. Say $t = t^{(h)}$. Then by Lemma 9, $\#MKP_\beta(s) = h - 1$. Suppose that $\#MKR(s) \geq h - 1$. Then Lemma 7 implies that on any path from

[8] In the case of Example 1, common $(1 - \epsilon)$-belief of rationality holds at any $s \in \Omega^\star$ with $t^\star = (1, 1)$. See Monderer and Samet (1989) for the concept of common p-belief.

s whose length is not more than $h-1$, both players are rational. Assume for each $i=1,2$, $\boldsymbol{a}_i(s_i)=B$. Then applying Lemma 10 repeatedly, we obtain a path $(s^\nu)_{\nu=0}^{h-1}$ such that for some $i=1,2$, letting $j\neq i$,
$s^0=s$,
$s_i^1=s_i^0$, and $t^1=t^{(h-1)}$,
$s_j^2=s_j^1$, and $t^2=t^{(h-2)}$,
$\vdots$
$s_2^{h-2}=s_2^{h-3}$, and $t^{h-2}=t^{(2)}$,
$s_1^{h-1}=s_1^{h-2}$, and $t^{h-1}=t^{(1)}$,
where for each $\nu=0,1,\cdots,h-1$ and each $k=1,2$, we have $\boldsymbol{a}_k(s_k^\nu)=B$. Note that the length of the path is $h-1$ and both players are rational at any point on this path. By Lemma 8, the assumption $s\in CK\tilde{E}$ implies $s^{h-1}\in CK\tilde{E}$. Thus at s^{h-1} Player2 knows if $\boldsymbol{t}_1=0$, $\boldsymbol{a}_1=A$. Note that $t_2^{h-1}=0$. Then by (ii) of Lemma 10, the rationality of Player2 at s^{h-1} implies $\boldsymbol{a}_2(s_2^{h-1})=A$. This is a contradiction.

(ii) Assume $t=(1,0)$. The assumption $s\in CK\tilde{E}$ implies that at s Player2 knows that if $\boldsymbol{t}_1=0$, $\boldsymbol{a}_1=A$. Thus Lemma 10 (ii) implies $\boldsymbol{a}_2(s_2)=A$.

$s\in CK\tilde{E}$ implies that Player 1 knows that Player2 knows that if $\boldsymbol{t}_1=0$, $\boldsymbol{a}_1=A$. Thus at s Player1 knows that if $\boldsymbol{t}_2=0$, $\boldsymbol{a}_2=A$. Then Lemma 10 (i) implies $\boldsymbol{a}_1(s_1)=A$. Thus it never be the case that for each $i=1,2$, $\boldsymbol{a}_i(s_i)=B$. □

Assuming common knowledge of rationality, Theorem 1 leads to the "paradoxical behavior" as observed in Rubinstein (1989). Corollary 1 below asserts that provided that Player1 choosing A for the payoff type 0 is common knowledge, if common knowledge of rationality obtains, then both players must choose A for all payoff types. Denote by $\hat{E}$ the event $\{s\in\Omega \mid \text{for each } i=1,2, \boldsymbol{a}_i(s_i)=A\}$.

Corollary 1: $CKR\cap CK\tilde{E}\subset CK\hat{E}\subset\hat{E}$.

Note that $CKR\cap CK\tilde{E}\in\mathcal{F}^\star$. Thus Corollary 1 implies that if $(t_1^\star,t_1,t_2^\star,t_2)\in CKR\cap CK\tilde{E}$, then for any $t'\in T$, we have $(t_1^\star,t_1',t_2^\star,t_2')\in CKR\cap CK\tilde{E}$, thus $(t_1^\star,t_1',t_2^\star,t_2')\in CK\hat{E}$. In other words, if at a SOW s, the component $t^\star$ of s is such that R and $\tilde{E}$ are both common knowledge (the component t is irrelevant), then each Player i takes the action A for all t_i, and this is common knowledge.

Proof: Let us denote for $i=1,2$, $\hat{E}_i=\{s\in\Omega \mid \boldsymbol{a}_i(s_i)=A\}$. We will prove $CKR\cap CK\tilde{E}\subset CK\hat{E}$ in (i) and (ii) in the below, and prove $CK\hat{E}\subset\hat{E}$ in (iii).

Let $s\in CKR\cap CK\tilde{E}$ be given.

(i) We prove $CKR\cap CK\tilde{E}\subset CK\hat{E}_2$. First we show $CKR\cap CK\tilde{E}\subset\hat{E}_2$ in (a) in the below. And we point out $CKR\cap CK\tilde{E}\subset CK\hat{E}_2$ in (b).

(a) Suppose $\boldsymbol{a}_2(s_2)=B$. By Lemma 3, $s\in CKR$ implies Player2 is rational at s. Then Lemma 10 (ii) implies that there exists some $s'\in\Omega^\star$ such that $s_2'=s_2$ and $\boldsymbol{a}_1(s_1')=B$. On the other hand, since we have $\#MKR(s)=\infty$, Theorem 1 implies for any $s'\in\Omega^\star$ such that $s_2'=s_2$, we have $\boldsymbol{a}_1(s_1')=A$. This is a

contradiction.

(b) By (a) in the above, we have $CKR \cap CK\tilde{E} \subset \hat{E}_2$. Lemma 1 implies $CKR \cap CK\tilde{E} = CK(R \cap \tilde{E})$. Then Lemma 2 implies $K^1CK(R \cap \tilde{E}) \subset K^1\hat{E}_2$. Note that $K^1CK(R \cap \tilde{E}) = CK(R \cap \tilde{E})$ by Lemma 6. Then we have $CK(R \cap \tilde{E}) \subset K^1\hat{E}_2$. Repeating the same argument, for any positive integer m, we can obtain $CK(R \cap \tilde{E}) \subset K^m\hat{E}_2$. This implies $CK(R \cap \tilde{E}) \subset CK\hat{E}_2$.

(ii) We prove $CKR \cap CK\tilde{E} \subset CK\hat{E}_1$.

(a) Assume $t_1 = 0$. By Lemma 3, $s \in CK\tilde{E}$ implies $\boldsymbol{a}_1(s_1) = A$.

(b) Assume $t_1 \geq 1$. Since we have proved in the above that $CKR \cap CK\tilde{E} \subset CK\hat{E}_2$, Player1 knows at s that $\boldsymbol{a}_2 = A$. Also $s \in CKR$ and Lemma 3 imply Player1 is rational at s. Then Lemma 10 (i) implies $\boldsymbol{a}_1(s_1) = A$.

(c) Similarly to (i) (b) above, we obtain $CKR \cap CK\tilde{E} \subset CK\hat{E}_1$.

Now we have $CKR \cap CK\tilde{E} \subset CK\hat{E}_1 \cap CK\hat{E}_2$. Then Lemma 1 implies $CKR \cap CK\tilde{E} \subset CK\hat{E}$.

(iii) We show that $CK\hat{E} \subset \hat{E}$. Suppose that there exists some SOW s such that $s \in CK\hat{E}$ and $s \notin \hat{E}$. The latter implies for some $i = 1, 2$, $s \notin \hat{E}_i$. Then since $\hat{E}_i \in \mathcal{F}_i$, Lemma 3 implies $s \notin K_i\hat{E}_i$. On the other hand, the former ($s \in CK\hat{E}$) implies $s \in K_i\hat{E}_i$. This is a contradiction.
□

Now consider the case where there is only one rationality type profile (i.e. $T^\star$ is a singleton) in Corollary 1. Then we obtain the result in Rubinstein (1989) which appeared in Section 1.1. Because now the behavior of $(\boldsymbol{a}_1, \boldsymbol{a}_2)$ induces only one strategy profile of the e-mail game. Then common knowledge of rationality implies that it constitutes a Bayesian Nash equilibrium of the e-mail game.

Corollary 2 (Rubinstein 1989): *Let (σ_1, σ_2) be a strategy profile of the e-mail game. Assume $\sigma_1(0) = A$. Then (σ_1, σ_2) is a Bayesian Nash equilibrium if and only if for each $i = 1, 2$, for all $t \in \{0, 1, 2, \cdots\}$, $\sigma_i(t) = A$.*

2.2 Special case

In the previous subsection, we gave a necessary condition for the (B, B) coordination by rational players under the assumption that "Player1 chooses the action A for the payoff type 0" is common knowledge. But as seen in Example 1, the condition is not sufficient for the coordination. Here we show that there is a class of embedding belief systems where the condition is also sufficient. Theorem 2 in the below states this.

Theorem 2: *Let an even positive integer m be given. Then there exists an embedding belief system $\mathcal{B}(\mathcal{G})$ which satisfies the following:*
(i) For any $s \in \Omega$, $s \in CK\tilde{E}$.
(ii) For any $s \in \Omega^\star \cap R$, we have $\#MKP_\beta(s) > \#MKR(s)$ if and only if

$\boldsymbol{a}_i(s_1) = B$ *for each* $i = 1, 2$.
(iii) For any k *with* $0 \leq k \leq m$, *there exists some* $s \in \Omega^\star$ *at which* $\#MKR(s) = k$.
(iv) For any $s \in \Omega^\star$, $\#MKR(s) \leq m$.

Note that Theorem 2 (ii), (iii) and (iv) are concerned only with what happens at **possible** SOWs. Theorem 2 asserts the existence of an embedding belief system in which at any SOW, it is common knowledge that Player1 chooses the action A for the rationality type 0, and at any possible SOW where each player is rational, the (B, B) coordination occurs if and only if the upper bound order of mutual knowledge of "the true payoff matrix is β" exceeds the upper bound order of mutual knowledge of rationality. Moreover, in the embedding belief system, mutual knowledge of rationality can obtain at any order up to some given number.

In the following, we give a recipe for constructing embedding belief systems that are to satisfy the properties (i)-(iv) in the statement of Theorem 2. And then we prove these embedding belief systems actually satisfy these properties.

Construction: Let a positive integer K be given. Then fix any embedding belief system that satisfies the following (1)-(3). Let us call this embedding belief system $\mathcal{B}^K(\mathcal{G})$.

(1) $T_1^\star = \{1, 2, \cdots, K\}$, and
$T_2^\star = \{0, 1, 2, \cdots, K\}$.

(2) The functions $\boldsymbol{a}_i (i = 1, 2)$ satisfy for any $s \in \Omega$,
$\boldsymbol{a}_i(t_i^\star, t_i) = A$ if $0 \leq t_i < t_i^\star$, and
$\boldsymbol{a}_i(t_i^\star, t_i) = B$ otherwise.

e.g. Let $t_i^\star = 4$.
Then Player i takes the actions
$\boldsymbol{a}_i(4, 0) = A$,
$\boldsymbol{a}_i(4, 1) = A$,
$\boldsymbol{a}_i(4, 2) = A$,
$\boldsymbol{a}_i(4, 3) = A$,
$\boldsymbol{a}_i(4, 4) = B$, $\leftarrow$ switching from A to B
$\boldsymbol{a}_i(4, 5) = B$,
$\boldsymbol{a}_i(4, 6) = B$,
$\vdots$

(3) The probability distribution $\rho^\star$ over $T^\star$ satisfies the following:
(i) For any $t^\star \in T^\star$, $\rho^\star(t^\star) > 0$ if and only if $t_1^\star = t_2^\star$ or $t_1^\star = t_2^\star + 1$.
(ii) Let $t^\star \in T^\star$ be such that $\rho^\star(t^\star) > 0$. Denote $t^\star$ by $t^{\star(h)}$ where $h = t_1^\star + t_2^\star$.
Then $\rho^\star$ satisfies $\frac{\rho^\star(t^{\star(h)})}{\rho^\star(t^{\star(h+1)})} \geq \frac{L-(1-\varepsilon)M}{(2-\varepsilon)M}$ for each $h = 1, 2, \cdots, 2K$. □

Example 2: Table 4 depicts the shape of the set $T^\star$ for the case $K = 3$. The rows are the rationality types of Player1 whereas the columns are those of Player2.

$t^{\star(h)}(h = 1, 2, \cdots, 6)$ are the rationality type profiles that are assigned positive probabilities. Note that they exhibit a "zigzag" structure. □

Table 4: A embedding belief system

	0	1	2	3
1	$t^{\star(1)}$	$t^{\star(2)}$		
2		$t^{\star(3)}$	$t^{\star(4)}$	
3			$t^{\star(5)}$	$t^{\star(6)}$

Now we show the embedding belief system $\mathcal{B}^K(\mathcal{G})$ satisfies the properties in Theorem 2 with an appropriate choice of the number K.

Lemma 11: *Let K be a positive integer. Let $\mathcal{B}^K(\mathcal{G})$ be given. Then the event R equals $\{s \in \Omega \mid t_2^\star \neq 0\}$.*

Paraphrasing, Lemma 11 says that in $\mathcal{B}^K(\mathcal{G})$, Player2 is not rational if her rationality type is 0, and is rational otherwise, and that Player1 is rational everywhere.

Proof: In the following, we will first prove that both players are rational in $\{s \in \Omega \mid t_2^\star \neq 0\}$ ((i) and (ii) in the below). Then we will show that Player2 is not rational outside of $\{s \in \Omega \mid t_2^\star \neq 0\}$ ((iii) in the below).

Let $s \in \Omega$ be given.

(i) We show that Player1 is rational at $t_1^\star$. Then all we have to do is check that Player1 is pointwise rational at the type $(t_1^\star, \tilde{t}_1)$ for each $\tilde{t}_1 = 0, 1, 2, \cdots$. Denote $k = t_1^\star$. The construction (3) of $\mathcal{B}^K(\mathcal{G})$ implies that at $(t_1^\star, \tilde{t}_1)$, Player1 knows $\boldsymbol{t}_2^\star = k - 1$ or k, thus he knows that $\boldsymbol{a}_2 = B$ if $\boldsymbol{t}_2 \geq k$, and that $\boldsymbol{a}_2 = A$ if $\boldsymbol{t}_2 \leq k - 2$. We examine three cases.

(a) Assume $\tilde{t}_1 \geq k + 1$. Then at $(t_1^\star, \tilde{t}_1)$, Player1 knows $\boldsymbol{t}_2 \geq k$, thus he knows $\boldsymbol{a}_2 = B$. Also he knows that the true payoff matrix is β. Thus at this type of Player1, his best reply is the action B. The construction (2) says that Player1 indeed takes B at this type. Thus Player1 is pointwise rational at $(t_1^\star, \tilde{t}_1)$ with $\tilde{t}_1 \geq k + 1$.

(b) Assume $\tilde{t}_1 \leq k-1$. At $(t_1^\star, \tilde{t}_1)$, Player1 knows $\boldsymbol{t}_2 = \tilde{t}_1$ or $\tilde{t}_1 - 1$. If $\boldsymbol{t}_2 = \tilde{t}_1 - 1$, then $\boldsymbol{t}_2 \leq k - 2$. Thus he knows that $\boldsymbol{a}_2 = A$ if $\boldsymbol{t}_2 = \tilde{t}_1 - 1$. Thus by Lemma 10 (i), the best reply of Player1 at this type is the action A. The construction (2) says that Player1 indeed does so. Thus Player1 is pointwise rational at $(t_1^\star, \tilde{t}_1)$ with $\tilde{t}_1 \leq k - 1$.

(c) Assume $\tilde{t}_1 = k$. Since $k \geq 1$, at $(t_1^\star, \tilde{t}_1)$ Player1 knows that the true payoff matrix is β. And at this type, Player1 believes $\boldsymbol{t}_2 = k - 1$ with probability $\frac{1}{1+(1-\varepsilon)}$, and $\boldsymbol{t}_2 = k$ with probability $\frac{1-\varepsilon}{1+(1-\varepsilon)}$. Also Player1 believes $\boldsymbol{t}_2^\star = k - 1$ with probability $\rho^\star(k, k-1)/\gamma$, and $\boldsymbol{t}_2^\star = k$ with probability $\rho^\star(k, k)/\gamma$, where

$\gamma = \rho^\star(k,k-1) + \rho^\star(k,k)$. The construction (2) implies $\boldsymbol{a}_2(t_2^\star, t_2) = B$ if $(t_2^\star, t_2) = (k-1,k-1), (k-1,k), (k,k)$, and $\boldsymbol{a}_2(t_2^\star, t_2) = A$ if $(t_2^\star, t_2) = (k, k-1)$. Thus at this type, Player1 believes $\boldsymbol{a}_2 = B$ with probability $\frac{1}{\gamma}\{\rho^\star(k,k-1) + \rho^\star(k,k)\frac{1-\varepsilon}{1+(1-\varepsilon)}\}$, and $\boldsymbol{a}_2 = A$ with probability $\frac{\rho^\star(k,k)}{\gamma}\frac{1}{1+(1-\varepsilon)}$. Note that by the construction (3), $\frac{\rho^\star(k,k-1)}{\rho^\star(k,k)} \geq \frac{L-(1-\varepsilon)M}{(2-\varepsilon)M}$. Thus $\frac{\mu(\boldsymbol{a}_2=B|(t_1^\star,\tilde{t}_1))}{\mu(\boldsymbol{a}_2=A|(t_1^\star,\tilde{t}_1))} \geq \frac{L-(1-\varepsilon)M}{(2-\varepsilon)M}\{1+(1-\varepsilon)\} + (1-\varepsilon)$. Denote $q = \frac{L-(1-\varepsilon)M}{(2-\varepsilon)M}(1+(1-\varepsilon)) + (1-\varepsilon)$. Then $\boldsymbol{Eg}_1(B, \boldsymbol{a}_2 \mid (t_1^\star, \tilde{t}_1)) \geq \frac{q}{1+q}M + \frac{1}{1+q}(-L)$. A simple calculation shows $\frac{q}{1+q}M + \frac{1}{1+q}(-L) = 0$. On the other hand, $\boldsymbol{Eg}_1(A, \boldsymbol{a}_2 \mid (t_1^\star, \tilde{t}_1)) = 0$. The construction (2) says that Player1 takes B at this type. Thus Player1 is pointwise rational at $(t_1^\star, \tilde{t}_1)$ with $\tilde{t}_1 = k$.

Now we conclude that Player1 is pointwise rational at $(t_1^\star, \tilde{t}_1)$ for each $\tilde{t}_1 = 0, 1, 2 \cdots$.

(ii) We show that Player2 is rational at $t_2^\star$ if $t_2^\star \geq 1$. Assume $t_2^\star \geq 1$. Then all we have to do is check that Player2 is pointwise rational at the type $(t_2^\star, \tilde{t}_2)$ for each $\tilde{t}_2 = 0, 1, 2, \cdots$. Denote $k = t_2^\star$. The construction (3) of $\mathcal{B}^K(\mathcal{G})$ implies that at $(t_2^\star, \tilde{t}_2)$, Player2 knows $\boldsymbol{t}_1^\star = k$ or $k+1$, thus she knows that $\boldsymbol{a}_1 = B$ if $\boldsymbol{t}_1 \geq k+1$, and that $\boldsymbol{a}_1 = A$ if $\boldsymbol{t}_1 \leq k-1$. We examine three cases.

(a) Assume $\tilde{t}_2 \geq k+1$. Then at $(t_2^\star, \tilde{t}_2)$, Player2 knows $\boldsymbol{t}_1 \geq k+1$, thus she knows $\boldsymbol{a}_1 = B$. Also she knows that the true payoff matrix is β. Thus at this type, Player2's best reply is the action B. The construction (2) says that Player2 indeed takes B at this type. Thus Player2 is pointwise rational at $(t_2^\star, \tilde{t}_2)$ with $\tilde{t}_2 \geq k+1$.

(b) Assume $\tilde{t}_2 \leq k-1$. Then at $(t_2^\star, \tilde{t}_2)$, Player2 knows $\boldsymbol{t}_1 = \tilde{t}_2$ or $\tilde{t}_2 + 1$. If $\boldsymbol{t}_1 = \tilde{t}_2$, then $\boldsymbol{t}_1 \leq k-1$. Thus she knows $\boldsymbol{a}_1 = A$ if $\boldsymbol{t}_1 = \tilde{t}_2$. Thus by Lemma 10 (ii) , the best reply of Player2 at this type is the action A. The construction says that Player2 indeed does so. Thus Player2 is pointwise rational at $(t_2^\star, \tilde{t}_2)$ with $\tilde{t}_2 \leq k-1$.

(c) Assume $\tilde{t}_2 = k$. Since we have assumed $t_2^\star \geq 1$, at $(t_2^\star, \tilde{t}_2)$ Player2 knows that the true payoff matrix is β. And at this type, Player2 believes $\boldsymbol{t}_1 = k$ with probability $\frac{1}{1+(1-\varepsilon)}$, and $\boldsymbol{t}_1 = k+1$ with probability $\frac{1-\varepsilon}{1+(1-\varepsilon)}$. Also Player2 believes $\boldsymbol{t}_1^\star = k$ with probability $\rho^\star(k,k)/\gamma$, and $\boldsymbol{t}_1^\star = k+1$ with probability $\rho^\star(k+1,k)/\gamma$, where $\gamma = \rho^\star(k,k) + \rho^\star(k+1,k)$. The construction (2) implies $\boldsymbol{a}_1(t_1^\star, t_1) = B$ if $(t_1^\star, t_1) = (k,k), (k,k+1), (k+1,k+1)$, and $\boldsymbol{a}_1(t_1^\star, t_1) = A$ if $(t_1^\star, t_1) = (k+1, k)$. Thus at this type, Player2 believes $\boldsymbol{a}_1 = B$ with probability $\frac{1}{\gamma}\{\rho^\star(k,k) + \rho^\star(k+1,k)\frac{1-\varepsilon}{1+(1-\varepsilon)}\}$, and $\boldsymbol{a}_1 = A$ with probability $\frac{\rho^\star(k+1,k)}{\gamma}\frac{1}{1+(1-\varepsilon)}$. Then the similar argument to (i)(c) in the above applies to conclude that Player2, who indeed takes the action B, is pointwise rational at $(t_2^\star, \tilde{t}_2)$ with $\tilde{t}_2 = k$.

Now we conclude that Player2 is pointwise rational at each $(t_2^\star, \tilde{t}_2)$ with $t_2^\star \geq 1$ and $\tilde{t}_2 = 0, 1, 2, \cdots$.

(iii) Finally, we show that Player2 is not rational at $t_2^\star$ with $t_2^\star = 0$. Assume $t_2^\star = 0$. Let $\tilde{t}_2 = 0$. Then at $(t_2^\star, \tilde{t}_2)$, Player2 knows $\boldsymbol{t}_1^\star = 1$ and $\boldsymbol{t}_1 = 0$ or 1. By the construction (2) of $\mathcal{B}^K(\mathcal{G})$, she knows at this type that $\boldsymbol{a}_1 = A$ if $\boldsymbol{t}_1 = 0$. Then Lemma 10 (ii) implies Player2's best reply is the action A. But the construction (2) says that Player2 takes the action B. Thus she is not pointwise rational. This

implies that she is not rational at any $(t_2^\star, t_2)$ with $t_2^\star = 0$.
□

Lemma 12: *Let K be a positive integer. Let $\mathcal{B}^K(\mathcal{G})$ be given. Let $s \in \Omega^\star \cap R$. Then $\#MKR(s) = t_1^\star + t_2^\star - 2$.*

Proof: Following the construction (3) (ii) of $\mathcal{B}^K(\mathcal{G})$, index those elements of $T^\star$ that $\rho^\star$ gives positive probabilities. Let $s \in \Omega^\star \cap R$. Then say $t^\star$'s index is h i.e. $t^\star = t^{\star(h)}$ Note that $h = t_1^\star + t_2^\star$. Consider a path $(s^\nu)_{\nu=0}^{h-1}$ such that
$s^0 = s$,
$s^1 = (t_1^{\star(h-1)}, t_1, t_2^{\star(h-1)}, t_2)$,
$s^2 = (t_1^{\star(h-2)}, t_1, t_2^{\star(h-2)}, t_2)$,
$\vdots$
$s^{h-1} = (t_1^{\star(1)}, t_1, t_2^{\star(1)}, t_2)$.
Note that $t^{\star(1)} = (1, 0)$. Then by Lemma 11, $s^0, s^1, \cdots, s^{h-2} \in R$ and $s^{h-1} \notin R$. Clearly, this is the shortest path from s to some possible SOW outside R, and its length is $h - 1$. Then Lemma 7 implies $\#MKR(s) = h - 2 = t_1^\star + t_2^\star - 2$. □

Proof of Theorem 2: Let K satisfy $m = 2(K - 1)$. And let $\mathcal{B}^K(\mathcal{G})$ be given. Then in the following, we see that $\mathcal{B}^K(\mathcal{G})$ satisfies all of the properties (i)-(iv).

Proof of (i): By the construction of $\mathcal{B}^K(\mathcal{G})$, we have $\Omega = \tilde{E}$. Thus $\tilde{E}$ is common knowledge at any $s \in \Omega$.

Proof of (ii): **(if)** Let $s \in \Omega^\star \cap R$. Then by (i), $s \in CK\tilde{E}$. Then applying Theorem 1, we have the desired conclusion.

(only if) Let $s \in \Omega^\star \cap R$. Assume $\#MKP_\beta(s) > \#MKR(s)$. Then by Lemma 9, $\#MKP_\beta(s) = t_1 + t_2 - 1$. Also by Lemma 12, $\#MKR(s) = t_1^\star + t_2^\star - 2$. Thus we have $t_1 + t_2 - 1 > t_1^\star + t_2^\star - 2$. Now it remains to see $t_i \geq t_i^\star$ for each $i = 1, 2$ from the above inequality. Then we will conclude $\boldsymbol{a}_i(s_i) = B$ for each $i = 1, 2$ by the construction (2) of $\mathcal{B}^K(\mathcal{G})$. Recall that $[t_1 = t_2$ or $t_1 = t_2 + 1]$ and $[t_1^\star = t_2^\star$ or $t_1^\star = t_2^\star + 1]$. So there are four cases:

(a) Assume $t_1 = t_2$ and $t_1^\star = t_2^\star$. We plug these equalities into the above inequality. Then we have $2t_1 - 1 > 2t_1^\star - 2$. This implies $t_1 \geq t_1^\star$ and thus $t_2 \geq t_2^\star$.

(b) Assume $t_1 = t_2$ and $t_1^\star = t_2^\star + 1$. Then we have $2t_1 - 1 > 2t_1^\star - 3$. This implies $t_1 \geq t_1^\star$. Also we have $2t_2 - 1 > 2t_2^\star - 1$. Thus $t_2 > t_2^\star$.

(c) Assume $t_1 = t_2 + 1$ and $t_1^\star = t_2^\star$. Then we have $2t_1 - 2 > 2t_1^\star - 2$. This implies $t_1 > t_1^\star$. Also we have $2t_2 > 2t_2^\star - 2$. This implies $t_2 \geq t_2^\star$.

(d) Assume $t_1 = t_2 + 1$ and $t_1^\star = t_2^\star + 1$. Then we have $2t_1 - 2 > 2t_1^\star - 3$. This implies $t_1 \geq t_1^\star$. Also we have $2t_2 > 2t_2^\star - 1$. This implies $t_2 \geq t_2^\star$.

Proofs of (iii) and (iv): Immediate from Lemma 12.
□

3 Concluding Remarks

3.1 Degree of irrationality in the embedding belief systems $\mathcal{B}^K(\mathcal{G})$

For the embedding belief systems $\mathcal{B}^K(\mathcal{G})$ that we have constructed in Section 2.2, we want to establish that **the smaller the risk of the coordination behavior, the smaller the degree of irrationality that we need to introduce.**

$\frac{L}{M}$ represents the relative loss arising from the failure of the (B, B) coordination to the player who have chosen the action B. Also, ε is the probability with which a signal gets lost. So when the "e-mail communication" has stopped, from the viewpoint of one player, the smaller ε is, the larger the probability with which the last signal the player sent has arrived. Thus these two parameters are considered representing the risk of the coordination behavior: the larger these numbers are, the larger the risk.[9]

On the other hand, we can calculate the "minimum degree of irrationality" that we need in order to have $\mathcal{B}^K(\mathcal{G})$: Let θ denote $\frac{(2-\varepsilon)M}{L-(1-\varepsilon)M}$. Note that $0 < \theta < \infty$. Then by the formula in the construction (3) (ii) of $\mathcal{B}^K(\mathcal{G})$, θ represents the the upper bound of the ratio $\frac{\rho^\star(t^{\star(h+1)})}{\rho^\star(t^{\star(h)})}$. Let $Q^{NR}(\theta, K)$ denote the **minimum prior probability that at least one player is not rational**. Then $Q^{NR}(\theta, K) = 1/(\sum_{p=1}^{2K} \theta^{p-1})$. We regard that this number indicates the **minimum degree of irrationality** in need.[10]

We have two convergence results where the irrationality is vanishing: (i) For a fixed K with $K \geq 2$, if $\frac{L}{M}$ goes to 1 and ε goes to 0 at the same time, which means the risk of the coordination goes to the minimum, then the minimum degree of irrationality goes to 0. (ii) For fixed $\frac{L}{M}$ and ε that satisfy $\theta \geq 1$, the minimum degree of irrationality goes to 0 as K becomes larger.

We have another convergence result where the irrationality persists: (iii) For fixed $\frac{L}{M}$ and ε that satisfy $0 < \theta < 1$, the minimum degree of irrationality goes to $1 - \theta$ as K becomes larger.

We note that the effect of $\frac{L}{M}$ is significant as an indicator of the risk: Regardless of the value of ε, as $\frac{L}{M}$ becomes larger, that means the risk of the coordination becomes higher, θ goes to 0 thus Q^{NR} goes to 1. On the other hand, the effect of ε as an indicator of the risk is not as severe as that of $\frac{L}{M}$: When $\frac{L}{M}$ goes to 1 (that means a small risk) at the same time ε goes to $\frac{1}{2}$ (that means a high risk), the value of θ goes to 3, which is larger than 1 so that Q^{NR} goes to 0 as K becomes larger.

We also note that as K becomes larger, the maximum of the minimum numbers of signals that players should require for the coordination behavior becomes

[9] Recall $L > M > 0$ and $0 < \varepsilon < \frac{1}{2}$ by assumption.

[10] We should also look at "the minimum prior probability that no mutual knowledge of rationality obtains" as another reasonable indicator of the degree of irrationality. But expressing this number is complicated. So we do not adopt this number. However, instead we could look at "the prior probability that no mutual knowledge of rationality obtains when the degree of irrationality is at the minimum," which equals $(1 + \theta)/(\sum_{p=1}^{2K} \theta^{p-1})$. The behavior of this number is similar to that of Q^{NR}.

larger.[11] Thus for fixed $\frac{L}{M}$ and ε, as we demand the smaller degree of the minimum irrationality (that requires the larger K), the desired coordination becomes harder in the sense that it may require more communication.

In summing up, when the risk of the coordination behavior is small, in particular L is not very large relative to M, a small degree of irrationality suffices to have the embedding belief systems $\mathcal{B}^K(\mathcal{G})$. Also, as we accept the larger K, the necessary degree of irrationality becomes smaller but the coordination becomes harder. And when L is very large, the risk of coordination is very high and we need a significant degree of irrationality.

3.2 Interpretation of the embedding belief systems $\mathcal{B}^K(\mathcal{G})$

On the interpretation of the embedding belief systems $\mathcal{B}^K(\mathcal{G})$ we have constructed in Section 2.2, it is worth noting the following. Firstly, in $\mathcal{B}^K(\mathcal{G})$, at any SOW, it is common knowledge that each player switches the scheduled action from A to B before sending $K+1$ signals. This might be interpreted that "K signals is many enough to attempt the coordination behavior" is understood as a "common practice" in the environment in which these players play the game.

Secondly, the "zigzag" structure of the positive probability points in $T^\star$ is essentially similar to the knowledge structure in the e-mail game. (Compare Table 2 and Table 4.) Thus we might be able to interpret that knowledge about rationality is shared by the players by some communication process similar to the "e-mail communication." However, we have not specified this communication process in this paper. We just suggest here that studying communication processes of sharing knowledge about rationality in general be a research theme for future. We believe this direction of research is important for the following reason. A standard assumption in game theory says that the rationality of players is common knowledge.[12] But it is rarely discussed how players can obtain information about the rationality of players. If one says it is common knowledge, how have players commonized the knowledge of rationality? And if there is some "protocol" that commonizes the knowledge about rationality, one should also consider the cases where the "protocol" breaks down. Our analysis suggests that it does matter how players acquire the knowledge (or beliefs) of rationality in particular when it is not common knowledge. In fact, it has been argued in many places that some counterintuitive results about rational behavior (in particular those of backward induction such as the "centipede game" as in Rosenthal (1982) or the finitely repeated Prisoners' Dilemma) are much concerned with the fact that we cannot completely rule out irrationality in reality. Thus we believe that we should examine how players obtain information about rationality.

3.3 Paradoxical feature of the e-mail game

We refer to the literature that does not consider the e-mail game really paradoxical. In particular, we mention that Morris (2002) argues for that the e-mail game

[11] In $\mathcal{B}^K(\mathcal{G})$, if the rationality type profile is $t^\star$, then each player needs at least $t_i^\star$ signals sent to take the action B. Thus t_i is the minimum number of signals required for the coordination behavior. And the maximum value of $t_i^\star$ is K.

[12] putting aside the theories of evolutionary games and bounded rationality.

paradox is not really paradoxical in the sense that it calls for some "resolution" by considering bounded rationality. He says that he rather defends the view that the e-mail game and similar paradoxical examples are "a sensible starting point for modelling the role of higher order beliefs in applied settings," and he mentioned financial markets, bank runs and exchange rate crises as examples of areas for applications.

In this paper we have started with the view that the e-mail game poses a "paradox" that should need a resolution. However, our conclusion does not contradict the view of of Morris (2002) mentioned above. Our results suggest that the paradoxical behavior may be resolved when the order of mutual knowledge of rationality is suitably bounded. Thus our position is compatible with the view that the e-mail game does not need to be paradoxical when it is applied to situations in which assuming (approximate) common knowledge of rationality (mutual knowledge of sufficiently high orders or common belief of rationality with sufficiently high probabilities) is plausible. On the other hand, our results say that in environments where relatively small orders of mutual knowledge of rationality is expected, players' behavior does not need to be compatible with the paradoxical equilibrium, and the (B, B) coordination is likely for some cases. In those cases, what matters is the structures of players' information about rationality.

References

[1] Aumann RJ (1987): Correlated equilibrium as an expression of Bayesian rationality. *Econometrica* 55(1): pp1–18.

[2] Aumann RJ (1992): Irrationality in game theory. in Dasgupta P, Gale D, Hart O, Maskin E (eds) *Economic Analysis of Markets and Games, Essays in Honor of Frank Hahn.* MIT Press. pp214–27.

[3] Aumann R, Brandenburger A (1995): Epistemic conditions for Nash equilibrium. *Econometrica* 63(5): pp1161–80.

[4] Binmore K, Samuelson L (2001): Coordinated action in the electronic mail game. *Games Econ Behavior* 35: pp6–30.

[5] Dimitri N (2003): Coordination in an email game without "almost common knowledge". *J Logic, Language and Information* 12(1): pp127–51.

[6] Dimitri N (2004): Efficiency and equilibrium in the electronic mail game; the general case. *Theoretical Computer Science* 314(3): pp335–49.

[7] Dulleck U (2002): The e-mail game revisited - Modeling rough inductive reasoning. Working paper 0211. Department of Economics, University of Vienna.

[8] Harsanyi JC (1967-68): Games with incomplete information played by Bayesian players. parts I, II, III. *Management Science* 14: pp159–82, pp320–34, pp486–502.

[9] Megiddo N (1986): Remarks on bounded rationality. IBM research report RJ 54310.

[10] Monderer D, Samet D (1989): Approximating common knowledge with common beliefs. *Games Econ Behavior* 1: pp170–90.

[11] Morris S (2001): Faulty communication: Some variations on the electronic mail game. *Advances in Theoretical Economics* 1(1): article 5. (http://www.bepress.com/bejte/advances/vol1/iss1/art5/)

[12] Morris S (2002) Coordination, communication, and common Knowledge: A retrospective on the electronic-mail game. *Oxford Review of Economic Policy* 18(4): pp433–45.

[13] Rosenthal R (1982): Games of perfect information, predatory pricing and the chain store paradox. *J Econ Theory* 25: pp92–100.

[14] Rubinstein A (1989): The Electronic mail game: Strategic behavior under "almost common knowledge." *Amer Econ Review* 79: pp385–91.