14.126 Lecture Notes on Rationalizability

Muhamet Yildiz

April 9, 2010

When we define a game we implicitly assume that the structure (i.e. the set of players, their strategy sets and the fact that they try to maximize the expected value of the von-Neumann and Morgenstern utility functions) described by the game is common knowledge. The exact implications of this implicit assumption is captured by rationalizability. In this lecture, I will formally demonstrate this fact.

I will further extend rationalizability to incomplete information games. Of course, every incomplete-information game can be represented as a complete information game, and the rationalizability is already defined for the latter game. That solution concept is called *ex-ante rationalizability.* It turns out that that notion is more restrictive and imposes some stronger assumptions than what is intended in incomplete information game. To capture the exact implications of the assumptions in the incomplete-information game, I will introduce another solution concept, called *interim correlated rationalizability*, which is related to the rationalizability applied to the interim representation of the game, in which types are considered as players.

Along the way, I will introduce a formulation of the Bayesian games that will be used in the remainder of the course.

1 Rationalizability in Complete-Information Games

Consider a complete-information game (N, S, u), where N is the set of players, with generic elements $i, j \in N$, $S = \prod_{i \in N} S_i$ is the set of strategy profiles, and $u : S \to \mathbb{R}^N$ is the profile of payoff functions $u_i : S \to \mathbb{R}$. A game (N, S, u) is said to be *finite* if N and S are finite. Implicit in the definition of the game game that player i maximizes the expected value of u_i with respect to a belief about the other players' strategies. I will next formalize this idea.

1.1 Belief, Rationality, and Dominance

Definition 1 *For any player i, a (correlated)* belief *of i about the other players' strategies is a probability distribution μ_{-i} on $S_{-i} = \prod_{j \neq i} S_j$.*

Definition 2 *The* expected payoff *from a strategy s_i against a belief μ_{-i} is*

$$u_i\left(s_i, \mu_{-i}\right) = E_{\mu_i}\left[u_i\left(s_i', s_{-i}\right)\right] \equiv \int u_i\left(s_i', s_{-i}\right) d\mu_{-i}\left(s_{-i}\right)$$

Note that in a finite game $u_i\left(s_i, \mu_{-i}\right) = \sum_{s_{-i} \in S_{-i}} u_i\left(s_i, s_{-i}\right) \mu_{-i}\left(s_{-i}\right)$.

Definition 3 *For any player i, a strategy s_i^* is a* best response *to a belief μ_{-i} if and only if*

$$u_i(s_i^*, \mu_{-i}) \geq u_i(s_i, \mu_{-i}) \qquad (\forall s_i \in S_i)\,.$$

Here I use the notion of a *weak best reply*, requiring that there is no other strategy that yields a strictly higher payoff against the belief. A notion of strict best reply would require that s^* yields a strictly higher expected payoff than any other strategy.

Definition 4 *Playing s_i is said to be* rational *(with respect to μ_{-i}) if s_i is a best response to a correlated belief μ_{-i} on S_{-i}.*

Remark 1 (Correlation) *The essential part in the definition of a belief is that the belief μ_{-i} of player i allows correlation between the other players' strategies. For example, in a game of three players in which each player is to choose between Left and Right, Player 1 may believe that with probability 1/2 both of the other players will play Left and with probability 1/2 both players will play Right. Hence, viewed as mixed strategies, it may appear as though Players 2 and 3 use a common randomization device, contradicting the fact that Players 2 and 3 make their decisions independently. One may then find such a correlated belief unreasonable. This line of reasoning is based on mistakenly identifying a player's belief with other players' conscious randomization. For Player 1 to have such a correlated belief, he does not need to believe that the other players make their decisions* *together. Indeed, he does not think that the other players are using randomization device. He thinks that each of the other players play a pure strategy that he does not know. He may assign correlated probabilities on the other players' strategies because he may assign positive probability to various theories and each of these theories may lead to a prediction about how the players play. For example, he may think that players play Left (as in the cars in England) or players play Right (as in the cars in France) without knowing which of the theories is correct.*

Depending on whether one allows correlated beliefs, there are two versions of rationalizability. Because of the above reasoning, in this course, I will focus on correlated version of rationalizability. Note that the original definitions of Bernheim (1985) and Pearce (1985) impose independence, and these concepts are identical in two player games.

Rationality is closely related to the following notion of dominance.

Definition 5 *A strategy* s_i^* strictly dominates s_i *if and only if*

$$u_i(s_i^*, s_{-i}) > u_i(s_i, s_{-i}), \forall s_{-i} \in S_{-i}.$$

Similarly, a mixed strategy σ_i strictly dominates s_i *if and only if* $u_i(\sigma_i, s_{-i}) \equiv \sum_{s_i' \in S_i} \sigma_i(s_i') u_i(s_i', s_{-i}) > u_i(s_i, s_{-i}), \forall s_{-i} \in S_{-i}$. *A strategy* s_i *is said to be* strictly dominated *if and only if there exists a pure or mixed strategy that strictly dominates* s_i.

That is, no matter what the other players play, playing s_i^* is strictly better than playing s_i for player i. In that case, if i is rational, he would never play the strictly dominated strategy s_i. That is, there is no belief under which he would play s_i, for s_i^* would always yield a higher expected payoff than s_i no matter what player i believes about the other players. The converse of this statement is also true, as you have seen in the earlier lectures.

Theorem 1 *Playing a strategy* s_i *is not rational for* i *(i.e.* s_i *is never a weak best response to a belief* μ_{-i}*) if and only if* s_i *is strictly dominated.*

Theorem 1 states that *if we assume that players are rational (and that the game is as described), then we conclude that no player plays a strategy that is strictly dominated (by some mixed or pure strategy), and this is all we can conclude.*

1.2 Iterated Dominance and Rationalizability

Let us write

$$S_i^1 = \{s_i \in S_i |\ s_i \text{ is not strictly dominated}\}.$$

By Theorem 1, S_i^1 is the set of all strategies that are best response to some belief.

Let us now explore the implications of the assumption that player i is rational and knows that the other players are rational. To this end, we consider the strategies s_i that are best response to a belief μ_{-i} of i on S_{-i} such that for each $s_{-i} = (s_j)_{j \neq i}$ with $\mu_{-i}(s_{-i}) > 0$ and for each j, there exists a belief μ_j of j on S_{-j} such that s_j is a best response to μ_j. Here, the first part (i.e. s_i is a best response to a belief μ_{-i}) corresponds to rationality of i and the second part (i.e. if $\mu_{-i}(s_{-i}) > 0$, then s_j is a best response to a belief μ_j) corresponds to the assumption that i knows that j is rational. By Theorem 1, each such s_j is not strictly dominated, i.e., $s_j \in S_j^1$. Hence, by another application of Theorem 1, s_i is not *strictly dominated given* S_{-i}^1, i.e., there does not exist a (possibly mixed) strategy σ_i such that

$$u_i(\sigma_i, s_{-i}) > u_i(s_i, s_{-i}) \qquad \forall s_{-i} \in S_{-i}^1.$$

Of course, by Theorem 1, the converse of the last statement is also true. Therefore, the set of strategies that are rationally played by player i knowing that the other players is also rational is

$$S_i^2 = \left\{s_i \in S_i |\ s_i \text{ is not strictly dominated given } S_{-i}^1\right\}.$$

By iterating this logic, one obtains the following iterative elimination procedure, called *iterative elimination of strictly-dominated strategies.*

Definition 6 (Iterative Elimination of Strictly-Dominated Strategies) *Set* $S^0 = S$*, and for any* $m > 0$ *and set*

$$S_i^m = \left\{s_i \in S_i |\ s_i \textit{ is not strictly dominated given } S_{-i}^{m-1}\right\},$$

i.e., $s_i \in S_i^m$ *iff there does not exist any* σ_i *such that*

$$u_i(\sigma_i, s_{-i}) > u_i(s_i, s_{-i}) \qquad \forall s_{-i} \in S_{-i}^{m-1}.$$

Rationalizability corresponds to the limit of the iterative elimination of strictly-dominated strategies.

Definition 7 (Rationalizability) *For any player i, a strategy is said to be* rationalizable *if and only if $s_i \in S_i^\infty$ where*

$$S_i^\infty = \bigcap_{m \geq 0} S_i^m.$$

Rationalizability corresponds to the set of strategies that are rationally played in situations in which it is common knowledge that everybody is rational, as defined at the beginning of the lecture. When a strategy s_i is rationalizable it can be justified/rationalized by an indefinite chain of beliefs μ_{-i} as above. On the other hand, if a strategy is not rationalizable, it must have been eliminated at some stage m, and such a strategy cannot be rationalized by a chain of beliefs longer than m.

We call the elimination process that keeps iteratively eliminating all strictly dominated strategies until there is no strictly dominated strategy *Iterated Elimination of Strictly Dominated Strategies*; we eliminate indefinitely if the process does not stop. We call a strategy *rationalizable* if and only if it survives iterated elimination of strictly dominated strategies.

I will next recall a fixed-point property of rationalizability from the earlier lectures.

Definition 8 *A set $Z = \prod_{i \in N} Z_i \subseteq S$ is said to be* closed under rational behavior *(CURB) iff for every $z_i \in Z_i$, there exists a belief μ^{z_i} on Z_{-i} such that z_i is a best response to μ^{z_i}.*

It is a useful finger exercise to establish some basic properties of CURB sets.

Exercise 1 *Show the following facts for a finite game (N, S, u).*

1. *For any CURB sets Z and Z', $Z \vee Z' \equiv (Z_i \cup Z_i')_{i \in N}$ is closed under rational behavior.*

2. *There exists the largest CURB set.*

The next result establishes a fixed-point definition for rationalizability.

Proposition 1 *For any finite game (N, S, u), S^∞ is the largest product set $Z \subseteq S$ that is closed under rational behavior.*

Proof. See Lectures 1 and 2. ■

1.3 Common Knowledge of Rationality and Rationalizability

I will now formalize the idea of common knowledge and show that rationalizability captures the idea of common knowledge of rationality precisely. I first introduce the notion of an incomplete-information epistemic model.

Definition 9 (Information Structure) *An* information (or belief) structure *is a list* $\left(\Omega, (I_i)_{i \in N}, (p_{i,\omega})_{i \in N, \omega \in \Omega}\right)$ *where*

- Ω *is a (finite) state space,*
- I_i *is a partition of* Ω *for each* $i \in N$*, called information partition of* i*,*
- $p_{i,\omega}$ *is a probability distribution on* $I_i(\omega)$*, which is the cell of* I_i *that contains* ω*, representing* belief *of* i*.*

Here, state summarizes all relevant facts of the world. Note that only one of the state is the true state of the world; all the other states are hypothetical states needed to encode the players' beliefs. If the true state is ω, player i is informed that the true state is in $I_i(\omega)$, and he does not get any other information. Such an information structure arises if each player observes a state-dependent signal, where $I_i(\omega)$ is the set of states in which the value of the signal of player i is identical to the value of the signal at state ω. The next definition formalizes the idea that I_i summarizes all of the information of i.

Definition 10 *Given any event* $F \subseteq \Omega$*, player* i *is said to* know *at* ω *that event* F *obtains iff* $I_i(\omega) \subseteq F$*. The event that* i *knows* F *is*

$$K_i(F) = \{\omega | I_i(\omega) \subseteq F\}.$$

In an information structure one can also check whether a player i knows at ω that another player j knows an event F, i.e., whether event $K_j(F)$ obtains. Clearly, this is the case if $I_i(\omega) \subseteq K_i(F)$. That is, for each $\omega' \in I_i(\omega)$, $I_j(\omega') \subseteq F$. Similarly, we can check whether a player i knows that another player j knows that k knows that ... event F obtains. An event F is *common knowledge* at ω iff one can make such statements ad infinitum, i.e., everybody knows that everybody knows that ... ad infinitum that F obtains. There is also a simple formulation of common knowledge.

Definition 11 *An event F' is a public event iff*

$$F' = \bigcup_{\omega' \in F'} I_i(\omega') \qquad \forall i \in N. \tag{1}$$

An event F is said to be common knowledge *at ω iff there exists a public event F' with $\omega \in F' \subseteq F$.*

Exercise 2 *Show the following facts.*

1. *The two definitions of common knowledge are equivalent, i.e., (1) holds for some F' with $\omega \in F' \subseteq F$ iff at ω, everybody knows F, everybody knows that everybody knows F, everybody knows that everybody knows that everybody knows F, ad infinitum.*

2. *If F is a public event, then F is common knowledge at each $\omega \in F$.*

3. *If F is a public event, then $\Omega \backslash F$ is common knowledge at each $\omega \in \Omega \backslash F$.*

4. *Ω is common knowledge at each $\omega \in \Omega$.*

5. *Let F be the set of all states at which some proposition q holds and suppose that F is common knowledge at ω. Then, there exists an information structure $\left(\Omega', (I'_i)_{i \in N}, (p_{i,\omega})_{i \in N, \omega \in \Omega'}\right)$ with $I'_i(\omega) = I_i(\omega)$ at each $i \in N$ and $\omega \in \Omega'$, such that proposition q holds throughout Ω'.*

Remark 2 *Note that when it is common knowledge that a proposition is true at a state, then the information structure $M = \left(\Omega, (I_i)_{i \in N}, (p_{i,\omega})_{i \in N, \omega \in \Omega}\right)$ is just a collage of an information structure $M' = \left(\Omega', (I'_i)_{i \in N}, (p_{i,\omega})_{i \in N, \omega \in \Omega'}\right)$ on which the proposition holds throughout and another information structure $\left(\Omega \backslash \Omega', (I''_i)_{i \in N}, (p_{i,\omega})_{i \in N, \omega \in \Omega \backslash \Omega'}\right)$. Therefore, it is without loss of generality to define common knowledge of a proposition for it being true throughout the state space.*

I have so far considered an abstract information structure for players N. In order to give a strategic meaning to the states, we also need to describe what players play at each state by introducing a strategy profile $\mathbf{s} : \Omega \to S$.

Definition 12 *A strategy profile* $\mathbf{s} : \Omega \to S$ *with respect to* $\left(\Omega, (I_i)_{i\in N}, (p_{i,\omega})_{i\in N,\omega\in\Omega}\right)$ *is said to be* adapted *if* $\mathbf{s}_i(\omega) = \mathbf{s}_i(\omega')$ *whenever* $I_i(\omega) = I_i(\omega')$.

The last condition on the strategy profile ensures that each player knows what he is playing. The possibility that $\mathbf{s}_i(\omega) \neq \mathbf{s}_i(\omega')$ for some $I_i(\omega) = I_i(\omega')$ would contradict the fact $\mathbf{s}_i(\omega)$ is what player i plays at state ω and that he cannot distinguish the states ω and ω' when $I_i(\omega) = I_i(\omega')$.

Definition 13 *An* epistemic model *is a pair* $M = \left(\Omega, (I_i)_{i\in N}, (p_{i,\omega})_{i\in N,\omega\in\Omega}, \mathbf{s}\right)$ *of an information structure and an adapted strategy profile with respect to the information structure.*

The ideas of rationality and common knowledge of it can be formalized as follows.

Definition 14 *For any epistemic model* $M = \left(\Omega, (I_i)_{i\in N}, (p_{i,\omega})_{i\in N,\omega\in\Omega}, \mathbf{s}\right)$ *and any* $\omega \in \Omega$, *a player* i *is said to be* rational at ω *iff*

$$\mathbf{s}_i(\omega) \in \arg\max_{s_i\in S_i} \sum_{\omega'\in I_i(\omega)} u_i(s_i, \mathbf{s}_{-i}(\omega'))\, p_{i,\omega}(\omega').$$

That is, $\mathbf{s}_i(\omega)$ is a best response to $\mathbf{s}_{-i}$ under player i's belief at ω. (Since $\mathbf{s}$ gives the strategic meaning to the states, player i's beliefs about s_{-i} at ω is given by $p_{i,\omega}$ and the mapping $\mathbf{s}_{-i}$, restricted to $I_i(\omega)$.)

Let's write

$$R_i = \{\omega | \text{player } i \text{ is rational at } \omega\}$$

for the event that corresponds to the rationality of player i. It is common knowledge that player i is rational at ω iff event R_i is common knowledge at ω.

Definition 15 *For any* $i \in N$, *a strategy* $s_i \in S_i$ *is said to be* consistent with common knowledge of rationality *iff there exists a model* $M = \left(\Omega, (I_j)_{j\in N}, (p_{j,\omega})_{j\in N,\omega\in\Omega}, \mathbf{s}\right)$ *with state* ω^* *at which it is common knowledge that all players are rational and* $\mathbf{s}_i(\omega^*) = s_i$.

By Remark 2, this is equivalent to saying that there exists a model M such that $\mathbf{s}_j(\omega')$ is a best response to $\mathbf{s}_{-j}$ at each $\omega' \in \Omega$ for each player $j \in N$. The next result states that rationalizability is equivalent to common knowledge of rationality in the sense that S_i^∞ is precisely the set of strategies that are consistent with common knowledge of rationality.

Theorem 2 *For any player $i \in N$ and any $s_i \in S_i$, s_i is consistent with common knowledge of rationality if and only if s_i is rationalizable (i.e. $s_i \in S_i^{\infty}$).*

Proof. ($\Longrightarrow$) First, take any s_i that is consistent with common knowledge of rationality. Then, there exists a model $M = \left(\Omega, (I_j)_{j \in N}, (p_{j,\omega})_{j \in N, \omega \in \Omega}, \mathbf{s}\right)$ with a state $\omega^* \in \Omega$ such that $\mathbf{s}_i(\omega^*) = s_i$ and for each j and ω,

$$\mathbf{s}_j(\omega) \in \arg\max_{s_j \in S_j} \sum_{\omega' \in I_j(\omega)} u_j(s_j, \mathbf{s}_{-j}(\omega')) p_{j,\omega}(\omega'). \tag{2}$$

Define Z by setting $Z_j = \mathbf{s}_j(\Omega)$. By Proposition 1, in order to show that $s_i \in S_i^{\infty}$, it suffices to show that $s_i \in Z_i$ and Z is closed under rational behavior. First part is immediate, as $s_i = \mathbf{s}_i(\omega^*) \in \mathbf{s}_i(\Omega) = Z_i$. To see the second part, for each $z_j \in Z_j$, noting that $z_j = \mathbf{s}_j(\omega)$ for some $\omega \in \Omega$, define belief $\mu_{-j,\omega}$ on $Z_{-j} = \mathbf{s}_j(\Omega)$ by setting

$$\mu_{-j,\omega}(s_{-j}) = \sum_{\omega' \in I_j(\omega), \mathbf{s}_{-j}(\omega') = s_{-j}} p_{j,\omega}(\omega'). \tag{3}$$

(By definition $\mu_{-j,\omega}$ is a probability distribution on Z_{-j}.) Then, by (2),

$$\begin{aligned} z_j &= \mathbf{s}_j(\omega) \in \arg\max_{s_j \in S_j} \sum_{\omega' \in I_j(\omega)} u_i(s_i, \mathbf{s}_{-j}(\omega')) p_{i,\omega}(\omega') \\ &= \arg\max_{s_j \in S_j} \sum_{s_{-j} \in Z_{-j}} \sum_{\omega' \in I_j(\omega), \mathbf{s}_{-j}(\omega') = s_{-j}} u_i(s_i, s_{-j}) p_{i,\omega}(\omega') \\ &= \arg\max_{s_j \in S_j} \sum_{s_{-j} \in Z_{-j}} u_i(s_i, s_{-j}) \mu_{-j,\omega}(s_{-j}), \end{aligned}$$

showing that Z is closed under rational behavior. (Here, the first line is by (2); the second equality is by the fact that $\mathbf{s}$ is adapted, and the last equality is by definition of $\mu_{-j,\omega}$.) Therefore, $s_i \in S_i^{\infty}$.

($\Longleftarrow$) Conversely, since S^{∞} is closed under rational behavior, for every $s_j \in S_j^{\infty}$, there exists a probability distribution μ_{-j,s_j} on S_{-j}^{∞} against which s_j is a best response. Define model

$$M^* = \left(S^{\infty}, (I_j, p_{j,s})_{j \in N, s \in S^{\infty}}, \mathbf{s}\right)$$

with

$$\begin{aligned} I_j(s) &= \{s_j\} \times S_{-j}^{\infty} \qquad \forall s \in S^{\infty} \\ p_{j,s}(s') &= \mu_{-j,s_j}\left(s'_{-j}\right) \qquad \forall s' \in I_j(s) \\ \mathbf{s}(s) &= s \qquad \forall s \in S^{\infty} \end{aligned}$$

In model M^* it is common knowledge that each player j is rational. Indeed, for each $s \in S^\infty$,

$$\mathbf{s}_j(s) = s_j \in \arg\max_{s'_j \in S_j} \sum_{s_{-j} \in S^\infty_{-j}} u_i\left(s'_j, s_{-j}\right) \mu_{-j,s}\left(s'_{-j}\right) = \arg\max_{s'_j \in S_j} \sum_{s' \in I_j(s)} u_i\left(s'_j, s_{-j}\right) p_{j,s}\left(s'\right),$$

where the equalities are by definition M^* and the inclusion is by definition of $\mu_{-j,s}$. Of course for every $s_i \in S_i^\infty$, there exists $s = (s_i, s_{-i}) \in S^\infty$ such that $\mathbf{s}_i(s) = s_i$, showing that s_i is consistent with common knowledge of rationality. ■

2 Games of Incomplete Information

Now, I will introduce a slightly different formulation of Bayesian games that we will use in the rest of the course.

Definition 16 *A Bayesian game is a list (N, A, Θ, T, u, p) where*

- *N is the set of players (with generic members i, j)*
- *$A = (A_i)_{i \in N}$ is the set of action profiles (with generic member $a = (a_i)_{i \in N}$)*
- *Θ is a set of payoff parameters θ*
- *$T = (T_i)_{i \in N}$ is the set of action profiles (with generic member $t = (t_i)_{i \in N}$)*
- *$u_i : \Theta \times A \to R$ is the payoff function of player i, and*
- *$p_i(\cdot|t_i) \in \Delta(\Theta \times T_{-i})$ is the belief of type t_i about (θ, t_{-i}).*

Here, each player i knows his own type t_i does not necessarily know θ or the other players' types, about which he has a belief $p_i(\cdot|t_i)$. The game is defined in terms of players' *interim* beliefs $p_i(\cdot|t_i)$, which they obtain after they observe their own type but before taking their action. The game can also be defined by ex-ante beliefs $p_i \in \Delta(\Theta \times T)$ for some belief p_i. The game has a *common prior* if there exists $\pi \in \Delta(\Theta \times T)$ such that

$$p_i(\cdot|t_i) = \pi(\cdot|t_i) \qquad \forall t_i \in T_i, \forall i \in N.$$

In that case, the game is simply denoted by $(N, A, \Theta, T, u, \pi)$.

Remark 3 *Recall that in Lecture 2 a Bayesian game was defined by a list* (N, A, T, u, π) *where utility function* $u_i : T \times A \to \mathbb{R}$ *depended on type profile* t *and the action profile.*[1] *Here, the utility function depends explicitly on payoff parameters but not on type profiles. The formulation here is slightly more general. Given a game* (N, A, T, u, π) *in the earlier formulation, one can simply introduce the set* $\Theta = T$ *of parameters and a new prior* $\hat{\pi}$ *on* $\Theta \times T$ *with support on the diagonal* $\{(t,t)\,|t \in T\}$*. Conversely, given a game* $(N, A, \Theta, T, u, \pi)$ *in our formulation with* $u_i : \Theta \times A \to \mathbb{R}$*, one can define a new utility function* $v_i : T \times A \to \mathbb{R}$ *by* $v_i(t,a) = E[u_i(\theta,a)\,|t] = \int_{\theta \in \Theta} u_i(\theta,a)\, d\pi(\theta|t)$*. Note that, however, by suppressing the dependence on the payoff parameter* θ*, the earlier formulation loses some information. Such information is not needed for Bayesian Nash equilibrium, but that information* is *used by interim correlated rationalizability, the main concept we will introduce in this lecture. Also, the interim formulation here reflects the idea behind the idea of incomplete information better.*

When a researcher models an incomplete information, there is often no ex-ante stage or an explicit information structure in which players observe values of some signals. In the modeling stage, each player i has

- some belief $\tau_i^1 \in \Delta(\Theta)$ about the payoffs (and the other aspects of the physical world), a belief that is referred to as the first order belief of i,
- some belief $\tau_i^2 \in \Delta\left(\Theta \times \Delta(\Theta)^{N\setminus\{i\}}\right)$ about the payoffs and the other players' first order beliefs (i.e. (θ, τ_{-i}^1)),
- some belief τ_i^3 about the payoffs and the other players' first and second order beliefs (i.e. $(\theta, \tau_{-i}^1, \tau_{-i}^2)$),
- ... up to infinity.

(It is an understatement that some of these beliefs may not be fully articulated even in players' own minds.)

Modeling incomplete information directly in this form is considered to be quite difficult. Harsanyi (1967) has proposed a tractable way to model incomplete information through a type space. In this formalization, one models the infinite hierarchy of beliefs

[1] The notation there was also different, using $\theta = (\theta_i)_{i \in N}$ for the type profile $t = (t_i)_{i \in N}$.

above through a type space (Θ, T, p) and a type $t_i \in T_i$ as follows. Given a type t_i and a type space (Θ, T, p), one can compute the first-order belief of type t_i by $h_i^1(\cdot|t_i) = \text{marg}_\Theta p(\cdot|t_i)$, so that $h_i^1(\theta|t_i) = \sum_{t_{-i}} p(\theta, t_{-i}|t_i)$, the second-order belief $h_i^2(\cdot|t_i)$ of type t_i by $h_i^2\left(\theta, \hat{h}_{-i}^1|t_i\right) = \sum_{\left\{t_{-i}|h_{-i}^1(\cdot|t_{-i})=\hat{h}_{-i}^1\right\}} p(\theta, t_{-i}|t_i)$, and so on. A type space (Θ, T, p) and a type $t_i \in T_i$ model a belief hierarchy $(\tau_i^1, \tau_i^2, \ldots)$ iff $h_i^k(\cdot|t_i) = \tau_i^k$ for each k.

It is important to keep in mind that in a type space only one type profile corresponds to the actual incomplete-information situation that is meant to be modeled. All the remaining type profiles are hypothetical situations that are introduced in order to model the players' beliefs.

Recall that given any Bayesian game $\mathcal{B} = (N, A, \Theta, T, u, \pi)$ with common prior π, one can define *ex-ante game* $G(\mathcal{B}) = (N, S, U)$ where $S_i = A_i^{T_i}$ and $U_i(s) = E_\pi[u_i(\theta, s(t))]$ for each $i \in N$ and $s \in S$. For any Bayesian game $\mathcal{B} = (N, A, \Theta, T, u, p)$, one can also define *interim game* $AG(\mathcal{B}) = \left(\hat{N}, \hat{S}, \hat{U}\right)$ where $\hat{N} = \bigcup_{i \in N} T_i$, $\hat{S}_{t_i} = A_i$ for each $t_i \in \hat{N}$ and $\hat{U}_{t_i}(\hat{s}) = E\left[u_i\left(\theta, \hat{s}_{t_{-i}}\right)|p_i(\cdot|t_i)\right] \equiv \sum_{(\theta, t_{-i})} u_i\left(\theta, \hat{s}_{t_{-i}}\right) p_i(\theta, t_{-i}|t_i)$.

3 Rationalizability in Games of Incomplete Information

There are many notions of rationalizability in Bayesian games, each reflecting a different view of these games.

3.1 Ex-ante Rationalizability

If one takes an ex-ante view of Bayesian games, the above analysis readily gives a definition for rationalizability as follows.

Definition 17 *Given any Bayesian game $\mathcal{B} = (N, A, \Theta, T, u, \pi)$ and any player $i \in N$, a strategy $s_i : T_i \to A_i$ is said to be* ex-ante rationalizable *iff s_i is rationalizable in ex-ante game $G(\mathcal{B})$.*

Ex-ante rationalizability makes sense if there is an ex-ante stage. In that case, ex-ante rationalizability captures precisely the implications of common knowledge of rationality as perceived in the ex-ante planning stage of the game. It does impose unnecessary restrictions on players' beliefs from an interim perspective, however. In order to illustrate the idea of ex-ante rationalizability and its limitations consider the following example.

Example 1 *Take* $N = \{1, 2\}$, $\Theta = \{\theta, \theta'\}$, $T = \{t_1, t_1'\} \times \{t_2\}$, $p(\theta, t_1, t_2) = p(\theta', t_1', t_2) = 1/2$. *Take also the action spaces and the payoff functions as*

θ	L	R
U	$1, \varepsilon$	$-2, 0$
D	$0, 0$	$0, 1$

θ'	L	R
U	$-2, \varepsilon$	$1, 0$
D	$0, 0$	$0, 1$

for some $\varepsilon \in (0, 1)$. *The ex-ante representation of this game is as follows:*

	L	R
UU	$-1/2, \varepsilon$	$-1/2, \varepsilon$
UD	$1/2, \varepsilon/2$	$-1, 1/2$
DU	$-1, \varepsilon/2$	$1/2, 1/2$
DD	$0, 0$	$0, 1$

Here, for the strategies of player 1, the first entry is the action for t_1 *and the second entry is the action for* t_1', *e.g.,* $UD(t_1) = U$ *and* $UD(t_1') = D$. *This game has a unique rationalizable strategy profile*

$$S^\infty(G(\mathcal{B})) = \{(DU, R)\}.$$

In computing the rationalizable strategies, one first eliminates UU, noting that it is dominated by DD, and then eliminates L and finally eliminates UD and DD. Note however that elimination of UU crucially relies on the assumption that player 1's belief about the other player's action is independent of player 1's type. Otherwise, we could not eliminate DD. For example, if type t_1 believed player 2 plays L, he could play U as a best response, and if type t_1' believed player 2 plays R, he could play U as a best response. The assumption that the beliefs of types t_1 and t_1' are embedded in the definition of ex-ante game. Moreover, the conclusion that (DU, R) is the only rationalizable strategy profile crucially relies on the assumption that player 1 knows that player 2 knows that player 1's belief about player 2's action is independent of player 1's belief about the state.

From an interim perspective, such invariance assumption for beliefs (and common knowledge of it) is unwarranted. Because distinct types of a player correspond to distinct hypothetical situations that are used in order to encode players' beliefs. There is no reason to assume that in those hypothetical situations a player's belief about the other player's action is independent of his beliefs about the payoffs. (Of course, if it were actually the case that player 1 observes a signal about the state without observing a signal about the other player's action and player 2 does not observe anything, then it would have been plausible to assume that player 1's belief about player 2's action does depend on his signal. This is what ex-ante rationalizability captures. This is not the story however in a genuine incomplete information.)

3.2 Interim Independent Rationalizability

In order to capture the implication of common knowledge of rationality from an interim perspective without imposing any restriction on the beliefs of distinct types, one then needs to eliminate actions for type in the interim stage. While most contemporary game theorists would agree on the relevant notion of ex-ante rationalizability and the relevant notion of rationalizability in complete-information games, there is a disagreement about the relevant notion of interim rationalizability in incomplete information games.

One straightforward notion of interim rationalizability is to apply rationalizability to interim game $AG(\mathcal{B})$. An embedded assumption on the interim game is however that it is common knowledge that the belief of a player i about (θ, t_{-i}), which is given by $p_i(\cdot|t_i)$ is independent of his belief about the other players actions. That is, his belief about (θ, t_{-i}, a_i) is derived from some belief $p_i(\cdot|t_i) \times \mu_{t_i}$ for some $\mu_{t_i} \in \Delta\left(A_{-i}^{T_{-i}}\right)$. This is because we have taken the expectations with respect to $p_i(\cdot|t_i)$ in defining $AG(\mathcal{B})$, before considering his beliefs about the other players' actions. Because of this independence assumption, such rationalizability notion is called interim *independent* rationalizability.

Definition 18 *Given any Bayesian game $\mathcal{B} = (N, A, \Theta, T, u, p)$ and any type t_i of player $i \in N$, an action $a_i \in A_i$ is said to be* interim independent rationalizable *(IIR) for t_i iff a_i is rationalizable for t_i in interim game $AG(\mathcal{B})$.*

As an illustration, I will next apply interim independent rationalizability to the previous Bayesian game.

Example 2 *Consider the Bayesian game in the previous example. The interim game $AG(\mathcal{B})$ is 3-player game with player set $\hat{N} = \{t_1, t_1', t_2\}$ with the following payoff table,* *where t_1 chooses rows, t_2 chooses columns, and t_1' chooses the matrices:*

$$U \quad : \quad \begin{array}{r|c|c|} \multicolumn{1}{r}{} & \multicolumn{1}{c}{L} & \multicolumn{1}{c}{R} \\ \cline{2-3} U & 1, \varepsilon, -2 & -2, 0, 1 \\ \cline{2-3} D & 0, \varepsilon/2, -2 & 0, 1/2, 1 \\ \cline{2-3} \end{array}$$

$$D \quad : \quad \begin{array}{r|c|c|} \multicolumn{1}{r}{} & \multicolumn{1}{c}{L} & \multicolumn{1}{c}{R} \\ \cline{2-3} U & 1, \varepsilon/2, 0 & -2, 1/2, 0 \\ \cline{2-3} D & 0, 0, 0 & 0, 1, 0 \\ \cline{2-3} \end{array}$$

(The first entry is the payoff of t_1, the second entry is the payoff of t_2, and the last entry is the payoff of t_1'.) In $AG(\mathcal{B})$, no strategy eliminated, and all actions are rationalizable for all types, i.e., $S_{t_1}^{\infty}(AG(\mathcal{B})) = S_{t_1'}^{\infty}(AG(\mathcal{B})) = \{U, D\}$ and $S_{t_2}^{\infty}(AG(\mathcal{B})) = \{L, R\}$. For example, for type t_1, who is a player in $AG(\mathcal{B})$, U is a best response to t_2 playing L (regardless of what t_1' would have played), and U is a best response to t_2 playing R. For t_2, L is a best response to (U, U) and R is a best response to (D, D).

3.3 Interim Correlated Rationalizability

As discussed in Remark 1, the fact that two players choose their actions independently or does not mean that a third player's belief about their actions will have a product form. In particular, just because all of player j's information about θ, which is the action of the nature, is summarized by t_j does not mean the belief of i about the state θ and the action of j does not have any correlation once one conditions on t_j. Once again i might find it possible that the factors that affect the payoffs may also affect how other players will behave given their beliefs (regarding the payoffs). This leads to the following notion of rationalizability, called interim *correlated* rationalizability.

Iterated Elimination of Strictly Dominated Actions Consider a Bayesian game $\mathcal{B} = (N, A, \Theta, T, u, p)$. For each $i \in N$ and $t_i \in T_i$, set $S_i^0[t_i] = A_i$, and define sets $S_i^k[t_i]$ for $k > 0$ iteratively, by letting $a_i \in S_i^k[t_i]$ if and only if

$$a_i \in \arg\max_{a_i'} \int u_i(\theta, a_i', a_{-i})\, d\pi(\theta, t_{-i}, a_{-i})$$

for some $\pi \in \Delta(\Theta \times T_{-i} \times A_{-i})$ such that

$$\text{marg}_{\Theta \times T_{-i}} \pi = p_i(\cdot|t_i) \text{ and } \pi\left(a_{-i} \in S_{-i}^{k-1}[t_{-i}]\right) = 1.$$

That is, a_i is a best response to a belief of t_i that puts positive probability only on the actions that survive the elimination in round $k-1$. We write $S_{-i}^{k-1}[t_{-i}] = \prod_{j\neq i} S_j^{k-1}[t_j]$ and $S^k[t] = \prod_{i\in N} S_i^k[t_i]$.

Definition 19 *The set of all* interim correlated rationalizable *(ICR) actions for player i with type t_i is*

$$S_i^{\infty}[t_i] = \bigcap_{k=0}^{\infty} S_i^k[t_i].$$

Since interim correlated rationalizability allows more beliefs, interim correlated rationalizability is a weaker concept than interim independent rationalizability, i.e., if an action is interim independent rationalizable for a type, then it is also interim correlated rationalizable for that type. When all types have positive probability, ex-ante rationalizability is stronger than both of these concepts because it imposes not only independence but also the assumption that a player's conjecture about the other actions is independent of his type. Since all of the equilibrium concepts are refinements of ex-ante rationalizability, interim correlated rationalizability emerges as the weakest solution concept we have seen so far, i.e., all of them are refinements of interim correlated rationalizability.

I will present three justifications for using interim correlated rationalizability in genuine cases of incomplete information, which I described in the previous section. First, interim correlated rationalizability captures the implications of common knowledge of rationality precisely. Second, interim independent rationalizability depends on the way the hierarchies are modeled, in that there can be multiple representations of the same hierarchy with distinct set of interim independent rationalizable actions. Finally, and most importantly, one cannot have any extra robust prediction from refining interim correlated rationalizability. Any prediction that does not follow from interim correlated rationalizability alone crucially relies on the assumptions about the infinite hierarchy of beliefs. A researcher cannot verify such a prediction in the modeling stage without the knowledge of infinite hierarchy of beliefs.

The following example illustrates the second justification.

Example 3 *Take $\Theta = \{-1, 1\}$, $N = \{1, 2\}$, and the payoff functions as follows*

$\theta = 1$	a_2	b_2	c_2
a_1	$1, 1$	$-10, -10$	$-10, 0$
b_1	$-10, -10$	$1, 1$	$-10, 0$
c_1	$0, -10$	$0, -10$	$0, 0$

$\theta = -1$	a_2	b_2	c_2
a_1	$-10, -10$	$1, 1$	$-10, 0$
b_1	$1, 1$	$-10, -10$	$-10, 0$
c_1	$0, -10$	$0, -10$	$0, 0$

Note that this is a coordination game with outside option (c_i) where the labels may be mismatched ($\theta = -1$). First consider the type space $\hat{T} = \left\{\left(\hat{t}_1, \hat{t}_2\right)\right\}$ with $\hat{p}\left(\theta = 1, \hat{t}\right) = \hat{p}\left(\theta = -1, \hat{t}\right) = -1/2$. It is common knowledge that both payoff parameters are equally likely. The interim game is the complete information game with the following expected payoff vector

	a_2	b_2	c_2
a_1	$-9/2, -9/2$	$-9/2, -9/2$	$-10, 0$
b_1	$-9/2, -9/2$	$-9/2, -9/2$	$-10, 0$
c_1	$0, -10$	$0, -10$	$0, 0$

Here, c_i strictly dominates the other two actions. Therefore, c_i is the only interim independent rationalizable action. On the other hand, a_i is a best respond to belief that puts probability 1/2 on $(1, a_j)$ and probability 1/2 on $(-1, b_j)$; b_i is a best respond to belief that puts probability 1/2 on $(-1, a_j)$ and probability 1/2 on $(1, b_j)$, and c_i is a best respond to belief that puts probability 1/2 on $(-1, a_j)$ and probability 1/2 on $(1, a_j)$. Therefore, no action is strictly dominated, showing that $S^\infty = A$. That is, while IIR has a unique solution (c_1, c_2), every outcome is allowed in ICR.

Now, consider the type space $T = \{1, -1\}^2$ with common prior

$$\pi\left(\theta, t_1, t_2\right) = \begin{cases} 1/4 & \text{if } \theta = t_1 \cdot t_2 \\ 0 & \text{otherwise.} \end{cases}$$

The strategy profile s^* is a Bayesian Nash equilibrium in the new Bayesian game where

$$s_i^*\left(1\right) = a_i \text{ and } s_i^*\left(-1\right) = b_i,$$

in addition to the Bayesian Nash equilibrium in which each player plays c_i regardless of his type. Therefore, all actions are IIR in the new game. This of course implies that all actions are ICR in the new game. Note, however, that in this game too each type t_i assigns 1/2 on $\theta = 1$ and 1/2 on $\theta = -1$, and therefore it is common knowledge that *both states are equally likely. That is, the two Bayesian games model the same belief hierarchy, while Bayesian Nash equilibria and IIR are distinct in two games. In contrast ICR is the same in both games, which is a general fact.*

3.4 Structure Theorem for ICR

I will now present a structure theorem that establishes that without knowledge of infinite hierarchies one cannot refine interim correlated rationalizability. In that in order to verify any predictions that relies on a refinement, the researcher has to have the knowledge of infinite hierarchy of beliefs. Along the way, I will also discuss upper-hemicontinuity of ICR.

Fix a finite set $N = \{1, \ldots, n\}$ of players and a finite set A of action profiles. Let $\Theta^* = \left([0,1]^A\right)^N$ be the space of all possible payoff functions. For any $\theta = (\theta_1, \ldots, \theta_n) \in \Theta^*$, the payoff of player i from any $a \in A$ is $u_i(\theta, a) = \theta_i(a)$. Consider the Bayesian games with varying finite type spaces (Θ, T, p) with $\Theta \subset \Theta^*$. Recall that for each t_i in T_i, we can compute the first-order belief $h_i^1(t_i)$ about θ, the second-order belief $h_i^2(t_i)$ about $\left(\theta, h_{-i}^1\right)$, and so on, where I suppress the dependence of h_i on (Θ, T, p) for simplicity of notation. The type t_i and (Θ, T, p) are meant to model the infinite belief hierarchy

$$h_i(t_i) = \left(h_i^1(t_i), h_i^2(t_i), \ldots\right).$$

We assume that in the modeling stage the researcher can have information only on finite orders of beliefs $\left(h_i^1(t_i), h_i^2(t_i), \ldots, h_i^k(t_i)\right)$, where k can be arbitrarily high but finite and the information can about these finite orders can be arbitrarily precise (without knowing $\left(h_i^1(t_i), h_i^2(t_i), \ldots, h_i^k(t_i)\right)$). If we consider the open sets generated the sets of hierarchies such a researcher can find possible, then we obtain the following (point-wise) convergence notion: For any sequence $t_i(m)$, $m \in \mathbb{N}$, coming from finite type spaces and any type t_i,

$$t_i(m) \to t_i \iff h_i^k(t_i(m)) \to h_i^k(t_i) \qquad \forall k, \tag{4}$$

where $h_i^k(t_i(m)) \to h_i^k(t_i)$ in the usual sense of convergence in distribution (i.e. for every bounded, continuous function f, $\int f dh_i^k(t_i(m)) \to \int f dh_i^k(t_i)$).

3.4.1 Upper-hemicontinuity

Proposition 2 *S^{∞} is upper-hemicontinuous in t. That is, for any sequence $t_i(m)$ and any type t_i with $t_i(m) \to t_i$ as in (4), if $a_i \in S_i^{\infty}[t_i(m)]$ for all large m, then $a_i \in S_i^{\infty}[t_i]$.*

Note that since A is finite, a sequence $a(m)$ convergence to a if and only if $a(m) = m$ for all large m. Hence, the last statement in the proposition states that if $a_i(m) \to a_i$ for some $a_i(m) \in S_i^{\infty}[t_i(m)]$, then $a_i \in S_i^{\infty}[t_i]$. To appreciate the result, consider the following two implications.

Fact 1 *For any upper-hemicontinuous solution concept $F : t \mapsto F[t] \subseteq A$,*

1. *F is invariant to the way hierarchies of beliefs are modeled, i.e., $F_i(t_i) = F_i$ for any two types t_i and t'_i with $h_i(t_i) = h_i(t'_i)$;*
2. *F is locally constant when the solution is unique, i.e., if $F[t] = \{a\}$, then for any sequence $t(m) \to t$, $F[t(m)] = \{a\}$ for all large m.*

Exercise 3 *Prove these facts.*

While upper-hemicontinuity of solution concepts with respect to payoff parameters within a simple model is usual, upper-hemicontinuity with respect to beliefs in the above sense is unusual because we allow types $t_i(m)$ come from different type spaces. For example, Example 3 implies that IIR is not upper-hemicontinuous because it is not invariant to the way hierarchies are modeled. Indeed, the structure theorem below will imply that there is no strict refinement of S^{∞} that is upper-hemicontinuous.

The meaning of upper-hemicontinuity to economic modeling is as follows. Consider the researcher above who has noisy information about finite orders of beliefs $\left(h_i^1(t_i), h_i^2(t_i), \ldots, h_i^k(t_i)\right)$. Suppose that a type $\hat{t}_i$ from some type space $\hat{T}$ is consistent with her information. Upper-hemicontinuity states that if k is sufficiently high and the noise is sufficiently small, then the researcher will be sure that all of the rationalizable actions of the actual type is in $S_i^{\infty}\left[\hat{t}_i\right]$. That is, the predictions of the ICR for $\hat{t}_i$ (i.e. the propositions that are true for all actions in $S_i^{\infty}\left[\hat{t}_i\right]$) remain true even if there is a small misspecification of interim beliefs due to lack of information,and the researcher can validate these predictions. I will call such predictions robust to misspecification of interim beliefs. The structure theorem implies the converse of the above statement, showing that the only robust predictions are those that follow rationalizability alone.

3.4.2 Structure Theorem

Theorem 3 (Structure Theorem) *For any finite $\hat{t}_i$ and any $a_i \in S_i^{\infty}\left[\hat{t}_i\right]$, there exists a sequence of types $t_i(m)$ from finite models converging to $\hat{t}_i$ such that $S_i^{\infty}\left[t_i(m)\right] = \{a_i\}$. Moreover, every open neighborhood of t_i contains an open neighborhood on which a_i is the only rationalizable action.*

The first statement states that any rationalizable action a_i can be made *uniquely* rationalizable by perturbing the interim beliefs of the type. Since ICR is upper-hemicontinuous, the previous fact implies that a_i remains the unique rationalizable action under further small perturbations. That is, a_i remains as the unique rationalizable action over an open neighborhood of the perturbed type. This leads to the last statement of the structure theorem.

In order to spell out the implications of the structure theorem for economic modeling, consider the researcher above, who can observe arbitrarily precise noisy signal about arbitrarily high but finite orders of beliefs. There are infinitely many types from various type spaces that are consistent with information. Suppose that she chooses to model the situation by one of these types, denoted by $\hat{t}_i$. Note that the set of possible types that is consistent with her information leads to an open neighborhood of $\hat{t}_i$. Consider any a_i that is rationalizable for $\hat{t}_i$. The structure theorem states that the set of alternatives types has an open subset on which a_i is uniquely rationalizable. Hence, she cannot rule out the possibility that a_i is the unique solution in the actual situation or in the alternative models that are consistent with her information. Moreover, if a_i is uniquely rationalizable in the actual situation, she could have learned that the actual situation is in the open set on which a_i is uniquely rationalizable by obtaining a more precise information about higher orders of beliefs. Therefore, she could not rule out the possibility that she could have actually verify that a_i is the unique ICR action.

Now suppose that the researcher uses a particular non-empty refinement Σ of ICR as her solution concept. Since Σ has to prescribe a_i to t_i when a_i is uniquely rationalizable for t_i, and since she cannot rule out the possibility that a_i is uniquely rationalizable, she cannot rule out the possibility that her solution concept prescribes a_i as the unique solution. Hence, in order to verify a prediction of her refinement, it must be the case that her prediction holds for a_i. Since a_i is an arbitrary ICR action, this implies that the only predictions of her solution concept that she can verify are those that she could have made without refining ICR.

Exercise 4 *Using the structure theorem, show that S^∞ does not have an upper-hemicontinuous non-empty strict refinement, i.e., if F is non-empty, upper-hemicontinuous and $F[t] \subseteq S^\infty[t]$ for all t, then $F = S^\infty$.*

The proof of the structure theorem is based on a generalized "contagion" argument. I will illustrate the idea of contagion on a well-known example, called the *E-mail Game*. We will see another application of contagion in global games.

Example 4 *Consider a two-player game with the following payoff matrix*

	Attack	*No Attack*
Attack	θ, θ	$\theta - 1, 0$
No Attack	$0, \theta - 1$	$0, 0$

where $\theta \in \Theta = \{-2/5, 2/5, 6/5\}$. Write $T = \left\{t^{CK}(2/5)\right\}$ for the model in which it is common knowledge that $\theta = 2/5$. This is a typical coordination game, which is called the Coordinated-Attack Game. In this game there are two pure-strategy equilibria, one in which each attack and obtain the payoff of 2/3, and one in which nobody attacks, each receiving zero. Pareto-dominant Nash equilibrium selects the former equilibrium. Now imagine an incomplete information game in which the players may find it possible that $\theta = -2/5$. Ex ante, players assign probability 1/2 to each of the values $-2/5$ and $2/5$. Player 1 observes the value of θ and automatically sends a message if $\theta = 2/5$. Each player automatically sends a message back whenever he receives one, and each message is lost with probability 1/2. When a message is lost the process automatically stops, and each player is to take one of the actions of Attack or No Attack. This game can be modeled by the type space $\tilde{T} = \{-1, 1, 3, 5, \ldots\} \times \{0, 2, 4, 6, \ldots\}$, where the type t_i is the total number of messages sent or received by player i (except for type $t_1 = -1$ who knows that $\theta = -2/5$), and the common prior p on $\Theta \times \tilde{T}$ where $p(\theta = -2/5, t_1 = -1, t_2 = 0) = 1/2$ and for each integer $m \geq 1$, $p(\theta = 2/5, t_1 = 2m-1, t_2 = 2m-2) = 1/2^{2m}$ and $p(\theta = 2/5, t_1 = 2m-1, t_2 = 2m) = 1/2^{2m+1}$. Here, for $k \geq 1$, type k knows that $\theta = 2/5$, knows that the other player knows $\theta = 2/5$, and so on through k orders. Now, type $t_1 = -1$ knows that $\theta = -2/5$, and hence his unique rationalizable action is No Attack. Type $t_2 = 0$ does not know θ but puts probability $2/3$ on type $t_1 = -1$, thus believing *that player 1 will play No Attack with at least probability $2/3$, so that No Attack is the only best reply and hence the only rationalizable action. More interestingly, type $t_1 = 1$ knows that $\theta = 2/5$, but his unique rationalizable action is still No Attack. Although he knows that $\theta = 2/5$, he does not know that player 2 it. He assigns probability $2/3$ to type 0, who does not know that $\theta = 2/5$, and probability $1/3$ to type 2, who knows that $\theta = 2/5$. Since type 0 plays No Attack in his unique rationalizable action, under rationalizability, type 1 assigns at least probability $2/3$ that player 2 plays No Attack. As a unique best reply, he plays No Attack. Applying this argument inductively for each type k, one concludes that the new incomplete-information game is dominance-solvable, and the unique rationalizable action for all types is No Attack.*

If we replace $\theta = -2/5$ with $\theta = 6/5$, we obtain another model, for which Attack is the unique rationalizable action. We consider type space $\check{T} = \{-1, 1, 3, 5, \ldots\} \times \{0, 2, 4, 6, \ldots\}$ and the common prior q on $\Theta \times \check{T}$ where $q(\theta = 6/5, t_1 = -1, t_2 = 0) = 1/2$ and for each integer $m \geq 1$, $q(\theta = 2/5, t_1 = 2m-1, t_2 = 2m-2) = 1/2^{2m}$ and $q(\theta = 2/5, t_1 = 2m-1, t_2 = 2m) = 1/2^{2m+1}$. One can easily check that this game is dominance-solvable, and all types play Attack.

Note that for $k > 0$, type k knows that it is kth-order mutual knowledge that $\theta = 2/5$, but he does not know if the other player knows this, assigning probability $2/3$ to the type who only knows that it is $k-1$th-order mutual knowledge that $\theta = 2/5$. While the interim beliefs of the types with low k differ substantially from those of the common knowledge type, the beliefs of the types with sufficiently high k are indistinguishable from those of the common knowledge type according to the researcher above. But it is the behavior of those far away types that determines the behavior of the indistinguishable types; the unique behavior of $k = -1$, determines a unique behavior for $k = 0$, which in turn determines a unique behavior for $k = 1$, which in turn determines a unique behavior for $k = 2 \ldots$ up to arbitrarily high orders. This is called contagion. The proof of Structure Theorem is based on a very general contagion argument. We will see another application of contagion in global games.

4 Notes on Literature

For complete information games, rationalizability has been introduced by Bernheim (1985) and Pearce (1985) in their dissertations. They have in addition assumed that the beliefs do not put correlation between different players' strategies. Aumann (1987) introduced the formulation of epistemic model for strategies and characterization of the solution concept in terms of rationality assumptions within the context of correlated equilibrium (under the common-prior assumption). The analysis of epistemic foundations of solution concepts in the more general set up is due to Tan and Werlang (1988), who have also formally proved that rationalizability captures the strategic implications of common-knowledge of rationality. (The arguments of Bernheim (1985) and Pearce (1985) were less formal; see also Brandenburger and Dekel (1987)).

Modeling hierarchies of beliefs through type spaces is proposed by Harsanyi (1967). The formalization of hierarchies is due to Mertens and Zamir (1985) and Brandenburger and Dekel (1993).

Battigalli (1998) has an extensive discussion of rationalizability concepts in incomplete-information games. The formulation of interim-correlated rationalizability is due to Dekel, Fudenberg, and Morris (2007), who also proved the upper-hemicontinuity of ICR. This paper also contains a characterization of common knowledge of rationalizability in terms of ICR, extending the characterization in the complete information games to Bayesian games. Example 3 is taken from Ely and Peski (2006).

The e-mail game is due to Rubinstein (1989). In this example Rubinstein demonstrated that efficient equilibrium of (Attack, Attack) is sensitive to the specification of higher order beliefs. This was the first application of contagion argument to the best of my knowledge. Kajii and Morris (1987) contain some more general applications of contagion.

The Structure Theorem is due to Weinstein and Yildiz (2007). Chen (2008), Penta (2008), and Weinstein and Yildiz (2009) extend the structure theorem to dynamic games. Penta (2008) also characterizes the robust predictions under arbitrary common-knowledge restriction on who knows which parameter. Weinstein and Yildiz (2008) characterizes the robust predictions of equilibrium in nice games (with convex action spaces, continuous utility functions and unique best replies) under arbitrary common-knowledge restrictions on payoffs. See also Oury and Tercieux (2007) for an interesting mechanism design application with small payoff perturbations.

In response to Rubinstein (1989), Monderer and Samet (1989) explored the notion of closeness to the common knowledge in the sense that the equilibrium behavior remains similar. They have shown that common knowledge is approximated in this strategic sense by their concept of *common p-belief*: every player assigns at least probability p to the event F, every player assigns at least probability p to the event that every player assigns at least probability p to the event F ... up to infinity. Given any strict equilibrium of a common knowledge game, if the game is common p-belief for sufficiently high p under a perturbation, then the equilibrium remains approximate equilibrium in the perturbed game. The idea of common p-belief has been very useful. Several recent papers explored the idea of strategic closeness using this concept (e.g., Dekel, Fudenberg, and Morris (2006), Ely and Peski (2008), Chen et al (2010)).

References

[1] Aumann, Robert (1987): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1-18.

[2] Bernheim, D. (1984): "Rationalizable Strategic Behavior," *Econometrica*, 52, 1007-1028.

[3] Brandenburger, A. and E. Dekel (1987): "Rationalizability and Correlated Equilibria," *Econometrica*, 55, 1391-1402.

[4] Brandenburger, A. and E. Dekel (1993): "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, 59, 189-198.

[5] Chen, Y. (2008): "A Structure Theorem for Rationalizability in Dynamic Games", Northwestern University Working Paper.

[6] Dekel, E. D. Fudenberg, S. Morris (2006): "Topologies on Types," *Theoretical Economics*, 1, 275-309.

[7] Dekel, E. D. Fudenberg, S. Morris (2007): "Interim Correlated Rationalizability," *Theoretical Economics*, 2, 15-40.

[8] Ely J. and M. Peski (2006): Theoretical Economics

[9] Ely J. and M. Peski (2008): "Critical Types".

[10] Harsanyi, J. (1967): “Games with Incomplete Information played by Bayesian Players. Part I: the Basic Model,” *Management Science* 14, 159-182.

[11] Kajii, A. and S. Morris (1997): “The Robustness of Equilibria to Incomplete Information,” *Econometrica*, 65, 1283-1309.

[12] Mertens, J. and S. Zamir (1985): “Formulation of Bayesian Analysis for Games with Incomplete Information,” *International Journal of Game Theory*, 10, 619-632.

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

[14] Oury, M. and O. Tercieux (2007): “Continuous Implementation,” PSE Working Paper.

[15] Pearce, D. (1984): “Rationalizable Strategic Behavior and the Problem of Perfection,” *Econometrica*, 52, 1029-1050.

[16] Penta, A. (2008): “Higher Order Beliefs in Dynamic Environments,” University of Pennsylvania Working Paper.

[17] Rubinstein, A. (1989): “The Electronic Mail Game: Strategic Behavior Under ‘Almost Common Knowledge’,” *The American Economic Review*, Vol. 79, No. 3, 385-391.

[18] Tan, T. and S. Werlang (1988): “The Bayesian foundations of solution concepts of games,” *Journal of Economic Theory* 45, 370-391.

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

[20] Weinstein, J. and M. Yildiz (2008): “Sensitivity of Equilibrium Behavior to Higher-order Beliefs in Nice Games,” MIT Working Paper.

[21] Weinstein, J. and M. Yildiz (2009): "A Structure Theorem for Rationalizability in Infinite-horizon Games," Working Paper.