Correlated Equilibrium via Hierarchies of Beliefs

Du, Songzi
Stanford Graduate School of Business

15. August 2009

Correlated Equilibrium via Hierarchies of Beliefs

Songzi Du *

Stanford GSB

First Draft: December 8, 2008

This Draft: August 15, 2009

Abstract

We study a model of correlated equilibrium where every player takes actions based on his hierarchies of beliefs (belief on what other players will do, on what other players believe about others will do, etc.) intrinsic to the game. Our model does away with messages from outside mediator that are usually assumed in the interpretation of correlated equilibrium. We characterize in every finite, complete information game the exact sets of correlated equilibria (both subjective and objective) that can be obtained conditioning on hierarchies of beliefs; the characterizations rely on a novel iterated deletion procedure. If the procedure ends after k rounds of deletion for a correlated equilibrium obtained from hierarchies of beliefs, then players in the equilibrium need to reason to at most k-th order beliefs. Further conceptual and geometric properties of the characterizations are studied.

1 Introduction

In this paper we study a model of correlated equilibrium (in any complete information game) where every player takes actions based on his hierarchies of beliefs (belief on what other players will do, on what other players believe about others will do, on what others believe others believe others will do, etc.) which are intrinsic to the game. Therefore, our model does

*songzidu@stanford.edu. I am very grateful to Yossi Feinberg and Andy Skrzypacz for advice and detailed comments on various drafts. And I thank Adam Brandenburger, Amanda Friedenberg, Qingmin Liu, Yair Livne, Robert Wilson, and seminar participants at Stanford GSB for comments and discussions.

away with messages from outside mediator that are usually assumed in the interpretation of correlated equilibrium.

Let us illustrate our kind of correlated equilibrium with a story (which goes back to Kohlberg and Mertens (1986)). Suppose players are recruited to play a complete information game. They are seated in separate rooms so that they cannot communicate with each other; each player inputs his strategy via a computer in his room. And they do not observe any signal or message while seating in the rooms; signals and messages are already incorporated in the payoffs and strategies of the game which are common knowledge among players. Then, can the players still play strategies that are part of a correlated equilibrium, even though they have no access to any correlation device? We will argue that they can, because the players may have intertwined hierarchies of beliefs (you believe that I believe that you believe that ...) about each other's strategies, which may come from the players' previous interactions (or the players may just be very imaginative people) and may serve as correlation devices.

More precisely, for a finite, complete information game, we have types that represent players' states of mind, and a pure strategy σ_i that maps types to actions for each player. We assume that these strategies satisfy the following condition:

$$\text{types } t_i \text{ and } t_i' \text{ induce the same hierarchy of beliefs} \Longrightarrow \sigma_i(t_i) = \sigma_i(t_i'), \tag{1}$$

in addition to the usual equilibrium (incentive compatibility) condition. Condition (1) simply says that players condition their actions on their hierarchies of beliefs in the game.

The types and strategies form a correlated equilibrium if the usual incentive compatibility condition is satisfied; it might be objective or subjective, depending on whether or not the beliefs associated with types come from a common prior. We work with *a posteriori equilibrium* (Aumann (1974, Section 8)) which is a refinement of subjective correlated equilibrium and where the incentive compatibility condition is satisfied in the a posteriori (or in other words, interim) stage; we call a posteriori equilibrium satisfying condition (1) *intrinsic a posteriori equilibrium.* And we refer to objective correlated equilibrium simply as correlated equilibrium; correlated equilibrium satisfying condition (1) is called *intrinsic correlated equilibrium.*

Notice that strategy σ_i in condition (1) must be a pure strategy. Therefore, correlated equilibrium distribution (over action profiles) consistent with condition (1) can be interpreted as being purifiable by hierarchies of beliefs.

Our main results characterize in strategic terms sets of action profiles played under intrinsic a posteriori equilibria and distributions of action profiles obtained from intrinsic correlated equilibria (for convenience, we call this distribution intrinsic correlated equilibrium as well, and likewise for correlated equilibrium), for every finite game. Our characterizations rely on a novel iterated deletion procedure. If the procedure terminates after k iterations of deletion for an intrinsic a posteriori equilibrium or an intrinsic correlated equilibria, then players in the equilibrium need to reason to (i.e. condition their actions on) at most k-th order beliefs.

We also show that in every finite game, the set of intrinsic correlated equilibria is convex, and any non-intrinsic correlated equilibrium can be broken down into irreducible sub-equilibria, one of which must be an extreme point in the set of correlated equilibria. This in particular implies that an irreducible and non-extreme correlated equilibrium must be intrinsic. Conceptually, higher order beliefs in a correlated equilibrium are analogous to the notions of "friend" of "friend" and higher-order "friendships" in a network, and this analogy leads to the notion of irreducibility (or connectedness) for correlated equilibrium.

On the other hand, we prove that in two-person games with generic payoffs[1], any non-degenerate mixed strategy Nash equilibrium (i.e. one that requires randomization for at least one player) is not an intrinsic correlated equilibrium (not intrinsic). The intuition is that a Nash equilibrium does not have any variation in belief about the other players' actions (for any given player), i.e. no variation in first order belief, which leads to the lack of variation in any higher order belief; on the other hand, condition (1) requires the presence of different hierarchies of beliefs to purify the mixed strategy — the source of mixing is the belief hierarchies. Thus, we have a contradiction. The payoff genericity assumption is needed: in Example 4.3 we construct a mixed Nash equilibrium that is intrinsic, i.e. it can be purified by hierarchies of beliefs.

Finally, our characterization reveals a connection between intrinsic a posteriori equilibrium and weakly dominated actions: a set of action profiles not played under any intrinsic a posteriori equilibrium will "typically" contain actions that are weakly dominated.

This paper is directly inspired by Brandenburger and Friedenberg (2008). Our characterization of intrinsic a posteriori equilibrium (Theorem 3.1) is a generalization of Brandenburger and Friedenberg's injectivity result for best-response set. The theorem contributes toward the open question of characterizing in strategic terms the solution concept studied in Brandenburger and Friedenberg; intrinsic a posteriori equilibrium forms a refinement of the solution of Brandenburger and Friedenberg.

[1]The class of generic games that we consider comes from the literature on the Lemke-Howson algorithm for computation of Nash equilibrium in two-person games; see von Stengel (2002).

An essential difference between our paper and Brandenburger and Friedenberg is that we are concerned with purification based on hierarchies of beliefs, while Brandenburger and Friedenberg are concerned with correlation based on hierarchies of beliefs. And Brandenburger and Friedenberg work with rationalizability, while we work with correlated equilibrium. Additionally, Brandenburger and Friedenberg do not work with common prior, while we do in the second half of the paper. We carefully compare our model and results to that of Brandenburger and Friedenberg in Section 5.

The paper proceeds as follows. In the next section we formally introduce our model. Section 3 studies intrinsic a posteriori equilibrium, and Section 4 studies intrinsic correlated equilibrium. We discuss related literature in Section 5. Section 6 concludes the paper.

2 The Model

2.1 Notations

We use the following standard notation: for product set $T = \prod_{i \in N} T_i$, let $T_{-j} = \prod_{i \neq j} T_i$. Likewise, for $t \in T$, let $t_{-i} = (t_j)_{j \neq i}$. And for $f_i : T_i \to X_i, i \in N$, we write $f_{-i}(t_{-i}) = (f_j(t_j))_{j \neq i}$.

Let $\Delta(X)$ be the set of Borel probability measures on topological space X; if X is finite or countable, we endow X with the discrete topology, so every subset is a Borel set.

For $\mu \in \Delta(T)$ where $T = \prod_{i \in N} T_i$ is finite or countable, let $\mu(t_i) = \mu(\{t_i\} \times T_{-i})$, $\mu(\cdot|t_i) \in \Delta(T_{-i})$ be μ conditional on the event $\{t_i\} \times T_{-i}$ if $\mu(t_i) > 0$, and let $\mu(t_j|t_i) = \mu(\{t_j\} \times \prod_{k \notin \{i,j\}} T_k|t_i)$ and likewise $\mu(t_j, t_i) = \mu(\{t_j\} \times \{t_i\} \times \prod_{k \notin \{i,j\}} T_k)$.

Finally, we write $x \neq y \in X$ to mean that $x \in X, y \in X$ and $x \neq y$

2.2 Set-up

We fix a finite, complete information game: (u, A, N), where N is a finite set of players ($|N| \geq 2$), $A = \prod_{i \in N} A_i$ a (non-empty) finite set of action profiles, and $u = (u_i)_{i \in N}$, $u_i : A \to \mathbb{R}$ for each $i \in N$, the payoffs.

We work with type space: $((\lambda_i)_{i \in N}, T)$, where $T = \prod_{i \in N} T_i$ is a (non-empty) finite or countably infinite[2] set of type profiles, and $\lambda_i : T_i \to \Delta(T_{-i})$ is player i's belief (i.e. probability measure), contingent on his type, about types of other players.

[2]This assumption is for the convenience of avoiding measurability issues. Since the game is finite, nothing significant changes when we let T_i be a general measurable space.

Every player i plays a pure action contingent on his type: $\sigma_i : T_i \to A_i$, which is his pure strategy. We write $\sigma = (\sigma_i)_{i \in N}$.

The equilibrium condition (incentive compatibility) is that for every $i \in N$, $t_i \in T_i$ and $a'_i \in A_i$:

$$\sum_{t_{-i} \in T_{-i}} u_i(\sigma_i(t_i), \sigma_{-i}(t_{-i}))\lambda_i(t_i)(t_{-i}) \geq \sum_{t_{-i} \in T_{-i}} u_i(a'_i, \sigma_{-i}(t_{-i}))\lambda_i(t_i)(t_{-i}) \tag{2}$$

Definition 2.1. $((\lambda_i)_{i \in N}, T, \sigma)$ *is an* a posteriori equilibrium *if (2) is satisfied.*

(λ, T, σ) *is a* correlated equilibrium *if* $\lambda \in \Delta(T)$ *is such that* $\lambda(t_i) > 0$ *for all* $i \in N$ *and* $t_i \in T_i$*, and (2) is satisfied for* $\lambda_i(t_i) := \lambda(\cdot|t_i)$*.*

Correlated equilibrium differ from a posteriori equilibrium only in that the beliefs of correlated equilibrium come from a common prior; the requirement that $\lambda(t_i) > 0$ is simply to get a well-defined conditional and is without loss of generality: we can throw aways type t_i such that $\lambda(t_i) = 0$.

For any $((\lambda_i)_{i \in N}, T, \sigma)$, we can define an extended type space (a *product structure*) that consolidates information contained in σ_i and λ_i. For each $i \in N$, let $\tilde{\lambda}_i : T_i \to \Delta(T_{-i} \times A_{-i})$ be such that

$$\tilde{\lambda}_i(t_i)(t_{-i}, a_{-i}) = \begin{cases} \lambda_i(t_i)(t_{-i}) & \text{if } \sigma_{-i}(t_{-i}) = a_{-i} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

for every $t_{-i} \in T_{-i}$ and $a_{-i} \in A_{-i}$.

Each type t_i induces through $\tilde{\lambda}_i$ a hierarchy of beliefs, of which the basic uncertainty for player i is A_{-i}, the actions of other players. The hierarchy of beliefs is player i's belief about other players' actions, his belief about their beliefs about others' actions, his belief about others' beliefs about others' beliefs, and so on. The following formulation of hierarchy of beliefs is standard: see Siniscalchi (2007) and Brandenburger and Friedenberg (2008). The set of all such hierarchies of beliefs forms an universal type space in which each player i has basic uncertainty A_{-i} [3].

For each $i \in N$, let $\mathcal{T}_i^1 = \Delta(A_{-i})$ be the set of player i's first order beliefs. And define $\delta_i^1 : T_i \to \mathcal{T}_i^1, t_i \mapsto \operatorname{marg}_{A_{-i}} \tilde{\lambda}_i(t_i)$. Therefore, the first order belief at type t_i is simply player i's belief on other players' actions. If player i is rational at type t_i, then his action $\sigma_i(t_i)$ must be a best response for this first order belief.

[3] In a "usual" universal type space (Mertens and Zamir (1985)), the basic uncertainty of every player is Θ, the set of "fundamentals" of the game that affect payoffs; in this paper the payoffs of the game are common knowledge among players (i.e. Θ is a singleton), so the only uncertainty is actions of players.

A second order belief is a joint probability over other players' actions and other players' first order beliefs. Notice that we can obtain first order belief from second order belief by "integrating" out in the second order belief other players' first order beliefs. And in general, a l-th order belief is a joint probability over other players' actions and other players' $(l-1)$-th order beliefs.

Formally, for $l \geq 2$ and $i \in N$, let $\mathcal{T}_i^l = \Delta(\mathcal{T}_{-i}^{l-1} \times A_{-i})$ be the set of player i's l-th order beliefs. Define $\delta_i^l : T_i \to \mathcal{T}_i^l$ such that $\delta_i^l(t_i)$ is the image measure of $\tilde{\lambda}_i(t_i)$ under map $(\delta_j^{l-1}, \mathrm{id}_{A_j})_{j \neq i}$. That is, for any Borel measurable $B \subseteq \mathcal{T}_{-i}^{l-1} \times A_{-i}$, $\delta_i^l(t_i)(B) = \tilde{\lambda}_i(t_i)((\delta_j^{l-1}, \mathrm{id}_{A_j})_{j \neq i}^{-1}(B))$, where $\mathrm{id}_{A_j} : A_j \to A_j$ is the identity function ($\mathrm{id}_{A_j}(a_j) = a_j$), and $(\delta_j^{l-1}, \mathrm{id}_{A_j})_{j \neq i} : T_{-i} \times A_{-i} \to \mathcal{T}_{-i}^{l-1} \times A_{-i}$ is the product map, i.e. $(\delta_j^{l-1}, \mathrm{id}_{A_j})_{j \neq i}(t_{-i}, a_{-i}) = (\delta_j^{l-1}(t_j), \mathrm{id}_{A_j}(a_j))_{j \neq i}$.

$(\delta_i^1(t_i), \delta_i^2(t_i), \delta_i^3(t_i), \ldots)$ is the *hierarchy of beliefs* (or belief hierarchy) of type t_i. The hierarchy of beliefs is a complete and canonical description of the state of mind of player i (regarding actions played in the game) at type t_i; it is canonical in the sense that it is independent of any type space.

Types with the same hierarchy of beliefs are called *redundant*.

Example 2.1. *Consider a symmetric (that is, $\lambda_1 = \lambda_2$) type space with two players: $i \in \{1, 2\}$, $T_i = \{\alpha, \alpha', \beta, \gamma\}$, $A_i = \{A, B\}$; and $\sigma_i(\alpha) = \sigma_i(\alpha') = A, \sigma_i(\beta) = \sigma_i(\gamma) = B$; and λ_i is as follows (each row is a probability distribution over the other player's types, e.g. $\lambda_i(\alpha) = 0.5\alpha + 0.5\gamma$, that is, with probability 0.5 the other player's type is α, and with probability 0.5 it is γ):*

	α	α'	β	γ
α	*0.5*	*0*	*0*	*0.5*
α'	*0.2*	*0.3*	*0*	*0.5*
β	*0.25*	*0.25*	*0.3*	*0.2*
γ	*0*	*0*	*0*	*1*

The first order beliefs of α, α' and β are the same: $0.5A + 0.5B$ (i.e. with probability 0.5 that the other player will do A, and with probability 0.5 that the other player will do B); the first-order belief of γ is B (i.e. with probability 1 that the other player will do B).

β is distinguished from α and α' by second-order belief ($\delta_i^2(\beta) \neq \delta_i^2(\alpha)$), because they have different beliefs about the other player's first order belief: α and α' believes that with probability 0.5 *the other player's first order belief is* $0.5A + 0.5B$*, and with probability* 0.5 *the other player's first order belief is B; while β believes that with probability* 0.8 *the other*

player's first order belief is $0.5A+0.5B$*, and with probability* 0.2 *the other player's first order belief is* B*.*

On the other hand, α *and* α' *are not distinguished by any order of belief, so they are redundant, having the same belief hierarchy.*

Definition 2.2. $((\lambda_i)_{i\in N}, T, \sigma)$ *is an* intrinsic a posteriori equilibrium *if it is an a posteriori equilibrium, and for every* $i \in N$*, for any two types* $t_i, t_i' \in T_i$ *with the same hierarchy of beliefs, i.e.* $\delta_i^l(t_i) = \delta_i^l(t_i')$ *for all* $l \geq 1$*, we have* $\sigma_i(t_i) = \sigma_i(t_i')$*.*

(λ, T, σ) *is an* intrinsic correlated equilibrium *if it is a correlated equilibrium, and for every* $i \in N$*, for any two types* $t_i, t_i' \in T_i$ *with the same hierarchy of beliefs, we have* $\sigma_i(t_i) = \sigma_i(t_i')$*; where the* δ_i^l*'s are defined with respect to* $\lambda_i(t_i) := \lambda(\cdot|t_i)$*.*

In other words, "intrinsicness" in the above definition requires the strategy of every player to be measurable on the partition generated by the player's hierarchy of beliefs; it rules out player i in an a posteriori equilibrium or correlated equilibrium playing different actions at types that have the same hierarchy of beliefs, i.e. types that player i himself cannot distinguish by thinking about other players' actions, about what others think about others' actions, about what others think about what others think, and so on.

Note that the redundant types α and α' in Example 2.1 will not cause any problem for the solution concepts in Definition 2.2, because σ_i assigns the same action at α and α'.

For intrinsic a posteriori equilibrium $((\lambda_i)_{i\in N}, T, \sigma)$, we are interested in action profiles played under this equilibrium, i.e. the product set $\prod_{i\in N} \sigma_i(T_i)$. And for intrinsic correlated equilibrium (λ, T, σ), we are interested in the distribution of action profiles *obtained* from the equilibrium, i.e. $\mu \in \Delta(A)$ such that $\mu(a) = \lambda(\{t \in T : \sigma(t) = a\})$ for every $a \in A$. We now briefly review the characterizations when the equilibrium is not required to be intrinsic.

Pearce (1984) and Bernheim (1984) in their studies of rationalizable actions introduce the concept of *best-response set* (BRS). A set of action profiles $Q = \prod_{i\in N} Q_i$ is a BRS if for each $i \in N$ and $a_i \in Q_i$, there exists a belief $\mu \in \Delta(Q_{-i})$ such that a_i is optimal for player i under μ (i.e. $u_i(a_i, \mu) \geq u_i(a_i', \mu)$ for all $a_i' \in A_i$, where as usual u_i is linearly extended to beliefs).

It is well-known (Brandenburger and Dekel, 1987) that for any non-empty set of action profiles $Q = \prod_{i\in N} Q_i$, there exists an a posteriori equilibrium $((\lambda_i)_{i\in N}, T, \sigma)$ under which Q is played ($Q_i = \sigma_i(T_i)$ for every $i \in N$) if and only if Q is a BRS.

It is also well-known that $\mu \in \Delta(A)$ is obtained from a correlated equilibrium (λ, T, σ) if and only if

$$\sum_{a_{-i} \in A_{-i}} u_i(a_i, a_{-i}) \mu(a_i, a_{-i}) \geq \sum_{a_{-i} \in A_{-i}} u_i(a'_i, a_{-i}) \mu(a_i, a_{-i}) \tag{4}$$

holds for every $i \in N$ and $a_i, a'_i \in A_i$.

In Section 3 and 4 we work out the exact strengthening in strategic terms that "intrinsicness" adds to the above characterizations.

We follow the convention in the literature to call $\mu \in \Delta(A)$ a correlated equilibrium (respectively, an intrinsic correlated equilibrium) if μ is obtained from a correlated equilibrium (respectively, an intrinsic correlated equilibrium) (λ, T, σ); that is, if $\mu(a) = \lambda(\{t \in T : \sigma(t) = a\})$ for all $a \in A$. Furthermore, we call a correlated equilibrium $\mu \in \Delta(A)$ intrinsic if it is an intrinsic correlated equilibrium.

3 Intrinsic A Posteriori Equilibrium

In this section we characterize the set of action profiles played under an intrinsic a posteriori equilibrium.

For a set of action profiles $Q = \prod_{i \in N} Q_i$, let

$$\beta_i^Q(a_i) = \{\mu \in \Delta(Q_{-i}) : a_i \text{ is optimal for player } i \text{ under } \mu\}, \tag{5}$$

for every $i \in N$ and $a_i \in Q_i$. For every μ in $\beta_i^Q(a_i)$, we say that μ is a *supporting belief* of action a_i in Q_{-i}.

It's easy to see that $\beta_i^Q(a_i)$ is a convex set (polytope, in fact); this simple property turns out to be crucial to our characterization theorems.

Clearly, $Q = \prod_{i \in N} Q_i$ is a best-response set (BRS) if and only if for every $i \in N$ and $a_i \in Q_i$ we have $\beta_i^Q(a_i) \neq \emptyset$.

If $\beta_i^Q(a_i) = \{\mu\}$, then we simply write $\beta_i^Q(a_i)$ for μ.

For each $i \in N$, let

$$\begin{aligned} W_i^1 &= \{a_i \in Q_i : |\beta_i^Q(a_i)| = 1\}, \\ W_i^l &= \{a_i \in W_i^1 : \beta_i^Q(a_i)(W_{-i}^{l-1}) = 1\}, l \geq 2, \\ W_i &= \bigcap_{l \geq 1} W_i^l. \end{aligned} \tag{6}$$

W_i^1 is the set of actions in Q_i that have a unique supporting belief in Q_{-i}. W_i^2 is the subset of W_i^1 for which the unique supporting belief has support contained in W_{-i}^1; in general, W_i^l is the subset of W_i^1 for which the unique supporting belief has support contained in W_{-i}^{l-1}.

Notice that $W = \prod_{i \in N} W_i$ is the largest BRS contained in $W^1 = \prod_{i \in N} W_i^1$.

We write $W_i(Q)$ and $W_i^l(Q)$ when it is necessary to emphasize the dependence on Q.

Definition 3.1. *A best-response set (BRS) $Q = \prod_{i \in N} Q_i$ is a* semi-injective BRS *if for every $i \in N$ and any two distinct actions a_i and a_i' in W_i, $\beta_i^Q(a_i) \neq \beta_i^Q(a_i')$.*

Brandenburger and Friedenberg define *injective BRS* as a BRS $Q = \prod_{i \in N} Q_i$ such that for every player i, we can find for every action in Q_i a distinct supporting belief in Q_{-i} (to which the action is optimal). This is equivalent to saying that for every player i and any two distinct actions a_i and a_i' in W_i^1, $\beta_i^Q(a_i) \neq \beta_i^Q(a_i')$. Semi-injectivity is weaker than injectivity, because semi-injectivity means that β_i^Q is injective over a smaller set — W_i, instead of W_i^1.

Brandenburger and Friedenberg (Proposition H.2) prove that (in our language) for any non-empty injective BRS $Q = \prod_{i \in N} Q_i$, there exists an intrinsic a posteriori equilibrium $((\lambda_i)_{i \in N}, T, \sigma)$ such that $Q_i = \sigma_i(T_i)$ for every $i \in N$. Here is our generalization:

Theorem 3.1. *For any non-empty set of action profiles $Q = \prod_{i \in N} Q_i$, there exist an intrinsic a posteriori equilibrium $((\lambda_i)_{i \in N}, T, \sigma)$ under which Q is played (i.e. $Q_i = \sigma_i(T_i)$ for every $i \in N$), if and only if Q is a semi-injective BRS.*

The proof of Theorem 3.1 implies the following corollary regarding the level of beliefs players need to reason in an intrinsic a posteriori equilibrium. When $l = 1$, (7) is Brandenburger and Friedenberg's injectivity condition; when $l = \infty$ (and let $W_i^\infty = W_i$), (7) is our semi-injectivity condition.

Corollary 3.2. *Fix a $l \geq 1$ and a non-empty $Q = \prod_{i \in N} Q_i$. If for every player i,*

$$a_i, a_i' \in W_i^l, a_i \neq a_i' \Longrightarrow \beta_i^Q(a_i) \neq \beta_i^Q(a_i'), \tag{7}$$

then there exists an a posteriori equilibrium in which players condition their actions on their l-th order beliefs, and under which Q is played. Conversely, if players only condition their actions on their l-th order beliefs in an a posteriori equilibrium, and Q is played under the equilibrium, then (7) holds.

The corollary implies that if the iterated deletions in (6) end in k rounds (i.e. $W_i^k = W_i$ for all $i \in N$) for a semi-injective BRS Q, then players need to reason to at most k-th order beliefs in a corresponding intrinsic a posteriori equilibrium.

Before moving on to the proof, we discuss the underlying idea. Notice that the W_i^l's constructed in Equation (6) partition Q_i into sets $Q_i \setminus W_i^1$, $W_i^1 \setminus W_i^2$, $W_i^2 \setminus W_i^3$, $W_i^3 \setminus W_i^4$, $\ldots$, and W_i. By construction, each action in $Q_i \setminus W_i^1$ is supported by an infinite number of first order beliefs, each action in $W_i^1 \setminus W_i^2$ is supported by an infinite number of second order beliefs and by a unique first order beliefs, each action in $W_i^2 \setminus W_i^3$ is supported by an infinite number of third order beliefs and by a unique second order beliefs, and so on. Note that if an action is supported by an infinite number of l-th order belief, then it is supported by an infinite number of hierarchies of beliefs. Since Q_i is finite, we will never have any trouble finding distinct hierarchies of beliefs to support actions in $Q_i \setminus W_i$.

On the other hand, each action a_i in W_i is supported by a unique l-th order belief, for every $l \geq 1$ (for $l = 1$, a_i is supported by the unique first order belief $\beta_i^Q(a_i)$); therefore a_i is supported by a unique hierarchy of beliefs. Therefore, the requirement that every player conditions his actions on his hierarchies of beliefs translate into the requirement that each action a_i in W_i has a distinct supporting belief $\beta_i^Q(a_i)$.

Proof of Theorem 3.1. **Only If:**

Fix an intrinsic a posteriori equilibrium $((\lambda_i)_{i\in N}, T, \sigma)$; let $Q_i = \sigma_i(T_i)$ for each $i \in N$, and let $\tilde{\lambda}_i$ be obtained from λ_i and σ by (3).

$Q = \prod_{i\in N} Q_i$ is clearly a BRS.

If $W_i = \emptyset$ for every $i \in N$, then there is nothing else to prove. Thus, suppose otherwise; note that this implies that $W_i \neq \emptyset$ for all $i \in N$.

The following lemma, which is essentially Proposition 11.1 in Brandenburger and Friedenberg (2008), demonstrates the connection between the set W_i^l and player i's l-th order beliefs.

Lemma 3.3. *For any $l \geq 1$, $i \in N$ and $a_i \in W_i^l$, there is exactly one l-th order belief in T_i mapped by σ_i to a_i; that is, if $\sigma_i(t_i) = \sigma_i(t_i') = a_i$, then $\delta_i^l(t_i) = \delta_i^l(t_i')$.*

Proof. If $\sigma_i(t_i) = a_i \in W_i^1, t_i \in T_i$, then clearly $\operatorname{marg}_{A_{-i}} \tilde{\lambda}_i(t_i) = \beta_i^Q(a_i)$. Thus the lemma is true when $l = 1$.

Now suppose $l \geq 2$, and that the lemma is true for $l-1$. Let $\sigma_i(t_i) = \sigma_i(t_i') = a_i \in W_i^2, t_i, t_i' \in T_i$. Then, $\operatorname{marg}_{A_{-i}} \tilde{\lambda}_i(t_i) = \operatorname{marg}_{A_{-i}} \tilde{\lambda}_i(t_i') = \beta_i^Q(a_i)$ because $a_i \in W_i^1$. If $\beta_i^Q(a_i)(a_{-i}) > 0$, $\tilde{\lambda}(t_i)(t_{-i}, a_{-i}) > 0$ and $\tilde{\lambda}(t_i')(t_{-i}', a_{-i}) > 0$, then we must have $\sigma_{-i}(t_{-i}) = \sigma_{-i}(t_{-i}') = a_{-i}$; and $a_{-i} \in W_{-i}^{l-1}$ by the construction of W_i^l. By the induction hypothesis, $\delta_j^{l-1}(t_j) = \delta_j^{l-1}(t_j')$ for every $j \neq i$. Thus, $\delta_i^l(t_i) = \delta_i^l(t_i')$. $\square$

Corollary 3.4. *For every $i \in N$ and $\mu \in \Delta(W_{-i})$, there can be at most belief hierarchy in T_i having first order belief μ, i.e. if $\mathrm{marg}_{A_{-i}} \tilde{\lambda}_i(t_i) = \mu = \mathrm{marg}_{A_{-i}} \tilde{\lambda}_i(t'_i)$, then $\delta^l_i(t_i) = \delta^l_i(t'_i)$ for every $l \geq 1$.*

Proof. Suppose $\mu \in \Delta(W_{-i})$ and $\mathrm{marg}_{A_{-i}} \tilde{\lambda}_i(t_i) = \mu = \tilde{\lambda}_i(t'_i), t_i, t'_i \in T_i$. If $\mu(a_{-i}) > 0$, $\tilde{\lambda}(t_i)(t_{-i}, a_{-i}) > 0$ and $\tilde{\lambda}(t'_i)(t'_{-i}, a_{-i}) > 0$, we must have $\sigma_{-i}(t_{-i}) = \sigma_{-i}(t'_{-i}) = a_{-i} \in W_{-i}$, and by the previous lemma $\delta^l_j(t_j) = \delta^l_j(t'_j)$ for every $j \neq i$ and $l \geq 1$. Thus, $\delta^l_i(t_i) = \delta^l_i(t'_i)$ for every $l \geq 1$. □

Now, for each $i \in N$ and $a_i \neq a'_i \in W_i$, by the assumption of $Q_i = \sigma_i(T_i)$, there exists $t_i, t'_i \in T_i$ such that $\sigma_i(t_i) = a_i$ and $\sigma_i(t'_i) = a'_i$; furthermore, t_i and t'_i have distinct belief hierarchies, by the "intrinsicness" of $((\lambda_i)_{i\in N}, T, \sigma)$. We have $\mathrm{marg}_{A_{-i}} \tilde{\lambda}_i(t_i) = \beta^Q_i(a_i)$ and $\mathrm{marg}_{A_{-i}} \tilde{\lambda}_i(t'_i) = \beta^Q_i(a'_i)$; and clearly $\beta^Q_i(a_i)(W_{-i}) = \beta^Q_i(a'_i)(W_{-i}) = 1$. Then $\beta^Q_i(a_i) \neq \beta^Q_i(a'_i)$, for otherwise the corollary above would imply that t_i and t'_i have the same hierarchy of beliefs.

If:

We prove this direction by construction.

Let $Q = \prod_{i\in N} Q_i$ be a non-empty semi-injective BRS. Let W^l_i and W_i be as defined in (6).

For each $i \in N$, let

$$T_i = \{a_i(k) : a_i \in Q_i \setminus W_i, k \in \{1,2\}\} \cup W_i$$

where $a_i(1)$ and $a_i(2)$ are two distinct copies of a_i.

We define the strategy $\sigma_i : T_i \to A_i$ as follows. For every $i \in N$, let $\sigma_i(a_i(1)) = \sigma_i(a_i(2)) = a_i$ for each $a_i \in Q_i \setminus W_i$; and let $\sigma_i(a_i) = a_i$, $a_i \in W_i$.

For every $i \in N$, let $t(a_i) = a_i(1)$ if $a_i \in Q_i \setminus W_i$; and let $t(a_i) = a_i$ if $a_i \in W_i$.

For every $i \in N$, define the belief $\lambda_i : T_i \to \Delta(T_{-i})$ as follows.

Step 1:

For each $a_i \in Q_i \setminus W^1_i$, fix $\nu(a_i, 1) \neq \nu(a_i, 2) \in \beta^Q_i(a_i) \setminus \beta^Q_i(W^1_i)$ such that

$$|\{\nu(a_i, k) : a_i \in Q_i \setminus W^1_i, k \in \{1,2\}\}| = 2|Q_i \setminus W^1_i|.$$

This is possible because $Q_i \setminus W^1_i$ and $\beta^Q_i(W^1_i)$ are finite sets, but $\beta^Q_i(a_i)$ is infinite for any $a_i \in Q_i \setminus W^1_i$ (recall that $\beta^Q_i(a_i)$ is a convex set).

For $a_i \in Q_i \setminus W_i^1$ and $k \in \{1, 2\}$, let

$$\lambda_i(a_i(k))(t_{-i}) = \begin{cases} \nu(a_i, k)(a_{-i}) & t_j = t(a_j) \text{ for every } j \neq i \\ 0 & \text{otherwise} \end{cases}$$

for every $t_{-i} \in T_{-i}$.

Clearly, each $a_i(k)$, $a_i \in Q_i \setminus W_i^1$ and $k \in \{1, 2\}$, induces through λ_i a distinct first order belief.

Step l: $(2 \leq l \leq L = \min\{l \geq 1 : W^l = W\})$

For each $a_i \in W_i^{l-1} \setminus W_i^l$, choose a $c(a_i) \in W_m^{l-2} \setminus W_m^{l-1}$, $m \neq i$, (where $W_m^0 = Q_m$) such that $\beta_i^Q(a_i)(c(a_i)) > 0$; such $c(a_i)$ exists by constructions of W_i^l's, and $c(a_i)$'s can be chosen so that $\beta_i^Q(a_i) = \beta_i^Q(a_i') \Rightarrow c(a_i) = c(a_i')$. And choose $\kappa(a_i, 1) \neq \kappa(a_i, 2) \in [0, 1]$ such that for any $a_i \neq a_i' \in W_i^{l-1} \setminus W_i^l$ with $\beta_i^Q(a_i) = \beta_i^Q(a_i')$, we have that $\kappa(a_i, 1)$, $\kappa(a_i', 1)$, $\kappa(a_i, 2)$ and $\kappa(a_i', 2)$ are all distinct.

For $a_i \in W_i^{l-1} \setminus W_i^l$ and $k \in \{1, 2\}$, let

$$\lambda_i(a_i(k))(t_{-i}) = \begin{cases} \beta_i^Q(a_i)(a_{-i}) & t_j = t(a_j), j \neq i, \text{ and } a_m \neq c(a_i) \\ \kappa(a_i, k)\beta_i^Q(a_i)(a_{-i}) & t_j = t(a_j), j \notin \{i, m\}, \text{ and } t_m = c(a_i)(1) \\ (1 - \kappa(a_i, k))\beta_i^Q(a_i)(a_{-i}) & t_j = t(a_j), j \notin \{i, m\}, \text{ and } t_m = c(a_i)(2) \\ 0 & \text{otherwise} \end{cases}$$

for every $t_{-i} \in T_{-i}$.

By induction on l, it's easy to see that each $a_i(k)$, $a_i \in W_i^{l-1} \setminus W_i^l$ and $k \in \{1, 2\}$, induces through λ_i a distinct l-th order belief.

Step $L + 1$:

Finally, for $a_i \in W_i$, let

$$\lambda_i(a_i)(t_{-i}) = \begin{cases} \beta_i^Q(a_i)(a_{-i}) & t_j = t(a_j) \text{ for every } j \neq i \\ 0 & \text{otherwise} \end{cases}$$

for every $t_{-i} \in T_{-i}$.

By assumption, each $a_i \in W_i$, has a distinct first order belief.

□

Example 3.1. *Consider the following symmetric two-person game:*

	A	B	C	D
A	1, 1	3, 3	0, 0	0, 4
B	3, 3	1, 1	0, 4	0, 0
C	0, 0	4, 0	1, 1	1, 1
D	4, 0	0, 0	1, 1	1, 1

First, note that $\{A, B, C, D\} \times \{A, B, C, D\}$ *is a BRS, so all actions can be played under a single a posteriori equilibrium.*

Let $Q_1 = Q_2 = \{A, B, C, D\}$. *Then* $\beta_1^Q(A) = \beta_1^Q(B) = \beta_2^Q(A) = \beta_2^Q(B) = \{1/2A + 1/2B\}$, *where* $1/2A + 1/2B$ *is the belief that assigns probability* $1/2$ *to* A *and* $1/2$ *to* B. *Thus,* $W_1 = W_2 = \{A, B\}$, *and* $Q = Q_1 \times Q_2$ *is not a semi-injective BRS. In fact, it's easy to see that for any* $C_1 \times C_2 \subseteq \{A, B, C, D\} \times \{A, B, C, D\}$, *if* $A \in C_i$ *or* $B \in C_i$ *for some* $i \in \{1, 2\}$, *then* $C_1 \times C_2$ *is either not a BRS, or not a semi-injective BRS.*

Thus, by Theorem 3.1, A or B cannot be played by either player under any intrinsic a posteriori equilibrium. In particular, intrinsic a posteriori equilibrium refines away the Nash equilibrium $(1/2A + 1/2B, 1/2A + 1/2B)$; *notice that both actions A and B are weakly dominated.*

3.1 Weak Domination

In this section we illustrate a connection between intrinsic a posteriori equilibrium and weakly dominated actions.

Recall the result of Brandenburger and Dekel (1987): for any non-empty set of action profiles $Q = \prod_{i \in N} Q_i$, there exists an a posteriori equilibrium $((\lambda_i)_{i \in N}, T, \sigma)$ under which Q is played (i.e. $Q_i = \sigma_i(T_i)$ for every $i \in N$) if and only if Q is a BRS.

Therefore, if action profiles $Q = \prod_{i \in N} Q_i$ is not played under any a posteriori equilibrium (i.e. is not a BRS), then there exist $i \in N$ and $a_i \in Q_i$ such that a_i is strictly dominated in Q_{-i}; that is, there exists $\alpha_i \in \Delta(A_i)$ such that $u_i(a_i, a_{-i}) < u_i(\alpha_i, a_{-i})$ for every $a_{-i} \in Q_{-i}$. This is because there must exist $i \in N$ and $a_i \in Q_i$ such that a_i is not a best response of player i to any $\mu \in \Delta(Q_{-i})$ (for otherwise Q would be a BRS), which is equivalent to the statement that a_i is strictly dominated in Q_i (Lemma 3 in Appendix B of Pearce (1984)).

We now show an analogous result with intrinsic a posteriori equilibrium and weak domination. Recall that $W_i \subseteq Q_i$ is defined from $Q = \prod_{i \in N} Q_i$ by Equation (6).

Proposition 3.5. *Suppose that a non-empty BRS* $Q = \prod_{i \in N} Q_i$ *is not played under any intrinsic a posteriori equilibrium (i.e. is not semi-injective), and that* $W_j \subsetneq Q_j$ *for some*

$j \in N$. Then, for every $i \neq j$, every $a_i \in W_i \neq \emptyset$ is weakly dominated in Q_{-i}; that is, there exists $\alpha_i \in \Delta(A_i)$ such that $u_i(a_i, a_{-i}) \leq u_i(\alpha_i, a_{-i})$ for every $a_{-i} \in Q_{-i}$, with strict inequality for some $a_{-i} \in Q_{-i}$.

Proof. We have $W = \prod_{i \in N} W_i \neq \emptyset$, for otherwise Q would be semi-injective. Take any $i \neq j$ and $a_i \in W_i$, the unique belief in Q_{-i} to which a_i is optimal has support contained in $W_{-i} \subsetneq Q_{-i}$. Thus, a_i is weakly dominated in Q_{-i}, because of the equivalence between being weakly dominated and not a best response to any belief with full support (Lemma 4 in Appendix B of Pearce (1984)). □

The next proposition shows that if Q is the set of correlated rationalizable action profiles (the maximum BRS), then we can dispense with the assumption of $W_j \subsetneq Q_j$. The proof is based on a geometric observation on the W_i^1 set.

Proposition 3.6. *Suppose that the set of correlated rationalizable action profiles $Q = \prod_{i \in N} Q_i$ is not played under any intrinsic a posteriori equilibrium (i.e. is not a semi-injective BRS). Then, for every $i \in N$, every $a_i \in W_i \neq \emptyset$ is weakly dominated in Q_{-i}. Furthermore, $a_i \in W_i$ cannot survive iterated deletion of weakly dominated actions in $A = \prod_{i \in N} A_i$.*

Proof. In light of the previous proposition, we will show that $W_i \subsetneq Q_i$ for all $i \in N$. This follows from the following claim:

Claim. *For any $i \in N$ and any $X_j \subseteq A_j$, $j \neq i$, such that $|X_{-i}| \geq 2$, there exists an $\bar{a}_i \in A_i$ such that $\bar{a}_i$ is player i's best response to two distinct beliefs on X_{-i}.*

First, notice that $|Q| > 1$, for otherwise Q would be a semi-injective BRS. Therefore, there exists $j \in N$ such that $|Q_j| > 1$.

For each $i \neq j$, apply the claim to get an $\bar{a}_i \in A_i$ that is player i's best response to two distinct beliefs on Q_{-i}. Clearly, $\bar{a}_i \in Q_i$ because Q is the set of correlated rationalizable action profiles. Therefore, $\bar{a}_i \notin W_i^1$. This implies that $W_i \subseteq W_i^1 \subsetneq Q_i$. Since $W_i \neq \emptyset$, this also means that $|Q_i| > 1$.

Now, apply the same reasoning to j to conclude that $W_j \subseteq W_j^1 \subsetneq Q_j$ as well.

Therefore, by the previous proposition, for every $i \in N$, every $a_i \in W_i \neq \emptyset$ is weakly dominated in Q_{-i}. Notice that any action $a_i \notin Q_i$ does not survive iterative deletion of strictly dominated actions in $A = \prod_{i \in N} A_i$. Therefore, $a_i \in W_i$ cannot survive iterative deletion of weakly dominated actions in $A = \prod_{i \in N} A_i$. □

Proof of the Claim. Let C be the convex hull,

$$C = \{(u_i(\mu_i, a_{-i}))_{a_{-i} \in X_{-i}} : \mu_i \in \Delta(A_i)\} \subseteq \mathbb{R}^{X_{-i}}.$$

Let $\bar{a}_i \in A_i$ be such that $x = (u_i(\bar{a}_i, a_{-i}))_{a_{-i} \in X_{-i}}$ is an extreme point of C that is not weakly dominated in C; clearly, such $\bar{a}_i$ exists . There must be multiple hyperplanes separating $C - x = \{y - x : y \in C\}$ from the positive orthant $\mathbb{R}_+^{X_{-i}}$, because the origin is an extreme point both of $C - x$ and of $\mathbb{R}_+^{X_{-i}}$, and $C - x \cap \mathbb{R}_+^{X_{-i}} = \{0\}$. Thus, $\bar{a}_i$ satisfies our desired conclusion. □

3.2 Iterated Deletion and Existence

In this section we work out an iterated deletion procedure that arrives at semi-injective BRS. We will show that this procedure always gives a non-empty set, thus there exists an intrinsic a posteriori equilibrium in every finite game.

Verbally, our iterated deletion works as follows: we start out with $\prod_{i \in N} R_i^1$, the set of all correlated rationalizable actions (the maximum BRS). Now, (1) delete a minimum number of actions from each R_i^1 so that the semi-injectivity condition in Definition 3.1 holds for the remaining actions; this gives $R_i^{2,1}$. But $\prod_{i \in N} R_i^{2,1}$ might not be a BRS, so (2) delete a minimum number of actions from each $R_i^{2,1}$ so that a BRS is obtained; this gives $\prod_{i \in N} R_i^2$. But now, the semi-injectivity condition might be lost for $\prod_{i \in N} R_i^2$, so we go back to (1) to get $\prod_{i \in N} R_i^{3,1}$, and then go to (2) to get a BRS $\prod_{i \in N} R_i^3$. We keep iterating this process until no more deletion is possible, i.e. until a semi-injective BRS $\prod_{i \in N} R_i$ emerges.

We now formally specify this iterated deletion procedure.

Step 1: For each $i \in N$, let R_i^1 be the the set of player i's correlated rationalizable actions, or equivalently, the set of player i's actions that survive iterated deletions of strictly dominated actions.

Step l ($l \geq 2$): Let a BRS $R^{l-1} = \prod_{i \in N} R_i^{l-1}$ be given from the previous step. Let $\beta_i^{l-1} = \beta_i^{R^{l-1}}$ (cf. Equation (5)), and let $W_i(l-1)$ be the $W_i(R^{l-1})$, i.e. the W_i obtained in Equation (6) when $Q = R^{l-1}$. And for each $i \in N$ and $\gamma \in \beta_i^{l-1}(W_i(l-1))$, fix an $a^{l-1}(\gamma) \in W_i(l-1)$ such that $\beta_i^{l-1}(a^{l-1}(\gamma)) = \gamma$; note that if β_i^{l-1} is injective on $W_i(l-1)$, there is a unique choice of $a^{l-1}(\gamma)$.

For each $i \in N$, let

$$R_i^{l,1} = \big(R_i^{l-1} \setminus W_i(l-1)\big) \cup \{a^{l-1}(\gamma) : \gamma \in \beta_i^{l-1}(W_i(l-1))\}, \tag{8}$$
$$R_i^{l,k} = \{a_i \in R_i^{l,1} : \exists \mu \in \Delta(R_{-i}^{l,k-1}) \text{ s.t. } a_i \text{ is optimal under } \mu\}, k \geq 2,$$
$$R_i^l = \bigcap_{k \geq 1} R_i^{l,k}.$$

Note that $R^l = \prod_{i \in N} R_i^l$ is the largest BRS contained in $R^{l,1} = \prod_{i \in N} R_i^{l,1}$.

Finally: Let $R_i = \bigcap_{l \geq 1} R_i^l$ for each $i \in N$.

By construction, for every $i \in N$ we have that

$$R_i^1 \supseteq R_i^2 \supseteq R_i^3 \supseteq \ldots \supseteq R_i.$$

Proposition 3.7. $R = \prod_{i \in N} R_i$ *is a non-empty, semi-injective BRS. And by some choice of* $a^{l-1}(\gamma)$ *for each* l *and* γ *in (8), we can obtain any maximal (in the set-inclusion partial order) semi-injective BRS as* R.

Proof. We will first show that each R_i is non-empty; it's clear that R is a semi-injective BRS.

It is well-known that each R_i^1 is non-empty: there always exist actions that are correlated rationalizable.

Now, fix a $l \geq 2$, and suppose that each R_i^{l-1} is non-empty. Then $R_i^{l,1}$ is non-empty because it contains $a^{l-1}(\gamma)$ where $\gamma \in \beta_i^{l-1}(W_i(l-1))$.

For any $k \geq 2$, suppose each $R_i^{l,k-1}$ is non-empty. Fix an $i \in N$ and any $\mu \in \Delta(R_{-i}^{l,k-1})$. Let $\mathrm{BR}_i(\mu) = \{a_i \in A_i : \ a_i \text{ is optimal for player } i \text{ under } \mu\}$.

Clearly, $\mathrm{BR}_i(\mu) \subseteq R_i^1$. And $\mathrm{BR}_i(\mu) \cap R_i^{2,1} \neq \emptyset$ because if there exists $a_i \in R_i^1 \setminus R_i^{2,1}$ such that $a_i \in \mathrm{BR}_i(\mu)$, then we must have $\beta_i^1(a_i) = \mu$, so by construction there exists an $a_i' \in \mathrm{BR}_i(\mu) \cap R_i^{2,1}$.

And we have $\mathrm{BR}_i(\mu) \cap R_i^{2,1} \subseteq R_i^{2,m}$ for any $m \geq 2$ (or $2 \leq m \leq k$ if $l = 2$) because $R_{-i}^{l,k-1} \subseteq R_{-i}^{2,m-1}$.

Repeating this argument, we conclude that $\emptyset \neq \mathrm{BR}_i(\mu) \cap R_i^{l,1} \subseteq R_i^{l,k}$, which implies that $R_i^{l,k}$ is non-empty.

Therefore, each R_i is non-empty.

For the second part of the proposition, fix a maximal semi-injective BRS $Q = \prod_{i \in N} Q_i$. Clearly, we have $Q_i \subseteq R_i^1$ for every $i \in N$. For any two distinct $a_i' \neq a_i \in W_i(Q)$, we have

$\beta_i^Q(a_i) \neq \beta_i^Q(a_i')$; and notice that $W_i(1) \cap Q_i \subseteq W_i(Q)$. Thus, so by some choices of $a^1(\gamma)$ in Equation (8), we have $Q_i \subseteq R_i^{2,1}$. And $Q_i \subseteq R_i^2$ because R^2 is the largest BRS contained in $R^{2,1}$.

Continuing on with this reasoning, we conclude that by some choice of $a^{l-1}(\gamma)$ for each l and γ in (8), we have $Q_i \subseteq R_i$. But this means that $Q_i = R_i$ since Q is a maximal semi-injective BRS. $\square$

Notice the proof does not use any property of W_i except that there is a unique supporting belief for each action in W_i. Therefore, the iterated deletion and the proposition also work if W_i is replaced by W_i^1; in this case we obtain Brandenburger and Friedenberg's injective BRS. Since an injective BRS is played under an intrinsic a posteriori equilibrium in which every player conditions his actions on his first order beliefs (Corollary 3.2), the proof of Proposition 3.7 implies that this more stringent kind of intrinsic a posteriori equilibrium also exists in every finite game.

4 Intrinsic Correlated Equilibrium

This section characterizes the distribution of actions profiles obtained from an intrinsic correlated equilibrium. Recall that we call this distribution (respectively, distribution obtained from a correlated equilibrium) intrinsic correlated equilibrium (respectively, correlated equilibrium) as well.

For a $\mu \in \Delta(A)$, let Q_i be the support of $\operatorname{marg}_{A_i} \mu$ for each $i \in N$, and let $Q = \prod_{i\in N} Q_i$.

Clearly, μ is a correlated equilibrium if and only if for every $i \in N$ and $a_i \in Q_i$ we have $\mu(\cdot|a_i) \in \beta_i^Q(a_i)$, where $\beta_i^Q(a_i)$, defined in (5), is the set of beliefs supporting a_i in Q_{-i}.

For each $i \in N$, define

$$\begin{aligned} Y_i^1 &= \{a_i \in Q_i :\ \mu(\cdot|a_i) \text{ is an extreme point of } \beta_i^Q(a_i)\}, \\ Y_i^l &= \{a_i \in Y_i^1 : \mu(Y_{-i}^{l-1}|a_i) = 1\}, l \geq 2, \\ Y_i &= \bigcap_{l\geq 1} Y_i^l, \end{aligned} \tag{9}$$

As before, we write $Y_i(\mu)$ and $Y_i^l(\mu)$ when it is necessary to emphasize the dependence on μ.

Theorem 4.1. *A correlated equilibrium $\mu \in \Delta(A)$ is an intrinsic correlated equilibrium if and only if for every $i \in N$, for any two distinct actions a_i and a_i' in Y_i, we have $\mu(\cdot|a_i) \neq \mu(\cdot|a_i')$.*

The theorem is completely analogous to Theorem 3.1 for intrinsic a posteriori equilibrium; see the discussion below Theorem 3.1 for some intuitions. To see why the iterated deletion Y_i^l's are of this form, we sketch the following lemma, which is analogous to Lemma 3.3:

Lemma 4.2. *Fix an intrinsic correlated equilibrium (λ, T, σ), and suppose that μ is obtained from (λ, T, σ) (i.e. $\mu(a) = \lambda(\{t \in T : \sigma(t) = a\})$). For any $l \geq 1$, $i \in N$ and $a_i \in Y_i^l$, there is exactly one l-th order belief in T_i mapped by σ_i to a_i; that is, if $\sigma_i(t_i) = \sigma_i(t_i') = a_i$, then $\delta_i^l(t_i) = \delta_i^l(t_i')$.*

Proof. Suppose $l = 1$. Fix $i \in N$ and $a_i \in Y_i^1$. If there exist $t_i, t_i' \in T_i$ such that $\delta_i^1(t_i) \neq \delta_i^1(t_i')$ but $\sigma_i(t_i) = \sigma_i(t_i') = a_i$ (and without loss of generality, assume that $\sigma_i^{-1}(a_i) = \{t_i, t_i'\}$), then because we have common prior, $\mu(\cdot|a_i)$ must be a strict convex combination of $\delta_i^1(t_i)$ and $\delta_i^1(t_i')$. This contradicts $\mu(\cdot|a_i)$ being an extreme point of $\beta_i^Q(a_i)$, because the incentive compatibility condition for correlated equilibrium (condition (2)) implies that $\delta_i^1(t_i)$ and $\delta_i^1(t_i')$ are in $\beta_i^Q(a_i)$.

The inductive step is same as that in Lemma 3.3 and does not use the common prior. □

The proof the only if of Theorem 4.1 then follows from the above lemma exactly as the proof of the only if in Theorem 3.1 follows from Lemma 3.3; it also does not use the common prior.

For the if direction of Theorem 4.1, we also follow the strategy of proof for Theorem 3.1. However, significant complications arise because we need to ensure that the belief hierarchies constructed come from a common prior, and that the common prior obtains μ, the correlated equilibrium under consideration; we leave details of the construction to the Appendix. In Example 4.3 we give a concrete example of the construction.

As with Theorem 3.1, we have the following corollary regarding the level of beliefs that players need to reason in an intrinsic correlated equilibrium.

Corollary 4.3. *Fix a $l \geq 1$ and a $\mu \in \Delta(A)$. If for every player i,*

$$a_i, a_i' \in Y_i^l, a_i \neq a_i' \Longrightarrow \mu(\cdot|a_i) \neq \mu(\cdot|a_i'), \tag{10}$$

then there exists an a correlated equilibrium that obtains μ in which players condition their actions on their l-th order beliefs. Conversely, if players only condition their actions on their l-th order beliefs in a correlated equilibrium that obtains μ, then (10) holds.

Before moving on to examples, we give an easy sufficient condition for a correlated equilibrium to be intrinsic. Brandenburger and Friedenberg in Appendix H observe that strict incentives imply injectivity in beliefs, which implies "intrinsicness". Here is an example of this implication for correlated equilibrium:

Correlated equilibrium $\mu \in \Delta(A)$ has *strict incentives* on the support if:

$$\sum_{a_{-i} \in A_i} u_i(a_i, a_{-i})\mu(a_i, a_{-i}) > \sum_{a_{-i} \in A_i} u_i(a_i', a_{-i})\mu(a_i, a_{-i}), \tag{11}$$

for every $i \in N$, $a_i \in Q_i = \mathrm{supp}(\mathrm{marg}_{A_{-i}} \mu)$ and $a_i' \in Q_i \setminus \{a_i\}$.

Myerson (1997) calls μ's incentives *elementary* if (11) is satisfied for every pair of distinct a_i and a_i' in A_i.

Proposition 4.4. *A correlated equilibrium with strict incentives on the support is intrinsic.*

The proof of the proposition is as follows: if incentives of a correlated equilibrium μ are strict on the support, then $\mu(\cdot|a_i)$ as a function of a_i must be injective on the support (but not vice versa), thus μ must be intrinsic.

Example 4.1 (Coordination game)**.**

	A	*B*
A	*10, 10*	*0, 0*
B	*0, 0*	*10, 10*

The Nash equilibrium $(1/2A + 1/2B, 1/2A + 1/2B)$ *is not an intrinsic correlated equilibrium:*

Let $Q_1 = Q_2 = \{A, B\}$*, then* $\beta_i^Q(A) = \{pA + (1-p)B : 1/2 \leq p \leq 1\}$ *and* $\beta_i^Q(B) = \{pA + (1-p)B : 0 \leq p \leq 1/2\}$ *for each* $i \in \{1, 2\}$*. Thus,* $1/2A + 1/2B$ *is an extreme point of both* $\beta_i^Q(A)$ *and* $\beta_i^Q(B)$*, and* $Y_i^1 = Y_i = \{A, B\}$*; but conditional beliefs of A and B in* $(1/2A + 1/2B, 1/2A + 1/2B)$ *are the same:* $1/2A + 1/2B$*.*

On the other hand, it's clear that (A, A) *and* (B, B) *are intrinsic correlated equilibria.*

More generally, a correlated equilibrium with full marginal support (i.e. the marginal distributions have full support, which includes all correlated equilibria except (A, A) *and* (B, B)*) can be represented as (where p is the probability of* (A, A) *being played, etc.)*

	A	B
A	p	q
B	r	s

with incentive inequalities $p/(p+r) \geq 1/2$, $p/(p+q) \geq 1/2$, $s/(s+q) \geq 1/2$, $s/(s+r) \geq 1/2$; *and* $p+q+s+r=1$. *Using previous characterizations of* $\beta_i^Q(A)$ *and* $\beta_i^Q(B)$, *we see that* $p=q=r=s=1/4$ *is the only correlated equilibrium that is not intrinsic; note that* $p=q=r=1/5$ *and* $s=2/5$ *is an intrinsic correlated equilibrium, with* $Y_1^1=Y_2^1=\{A\}$ *(we have* $Y_i^2=Y_i=\emptyset$ *for both* i*).*

Therefore, the set of intrinsic correlated equilibria in this game consists of all correlated equilibria except the fully mixed Nash equilibrium; note that this set is not closed.

Example 4.2 (Matching pennies, non-existence of intrinsic correlated equilibrium)**.**

	A	B
A	1, -1	-1, 1
B	-1, 1	1, -1

The Nash equilibrium $(1/2A+1/2B, 1/2A+1/2B)$ *here again is not an intrinsic correlated equilibrium; the same reasoning from the previous example applies.*

But $(1/2A+1/2B, 1/2A+1/2B)$ *is the unique correlated equilibrium of this game. Thus, this game has no intrinsic correlated equilibrium*

Notice that $\{A,B\} \times \{A,B\}$ *is a semi-injective BRS, so there certainly exists intrinsic a posteriori equilibrium in this game.*

Example 4.3 (Matching pennies with explicit randomization by one player, mixed Nash equilibrium being intrinsic)**.**

	A	B
A	1, -1	-1, 1
B	-1, 1	1, -1
C	0, 0	0, 0

The mixed Nash equilibrium $(1/4A+1/4B+1/2C, 1/2A+1/2B)$ *is an intrinsic correlated equilibrium:*

$Y_1^1=\{A,B,C\}$ *as before. But* $Y_2^1=\emptyset$ *because* $1/4A+1/4B+1/2C$ *can be written as a convex combination of* $1/6A+1/6B+2/3C$ *and* $1/2A+1/2B$, *to each of which* A *(respectively,* B*) is a best response of player 2. Thus,* $Y_i^2=Y_i=\emptyset$ *for any* $i \in \{1,2\}$.

Conceptually, $(1/4A+1/4B+1/2C, 1/2A+1/2B)$ *is an intrinsic correlated equilibrium because the presence of player 1's explicit randomization* C *introduces variations in player 2's supporting first order beliefs, which lead to variations in player 1's supporting second order beliefs that are used to purify player 1's mixed strategy.*

Here is an explicitly written intrinsic correlated equilibrium (λ, T, σ) *that obtains* $(1/4A+1/4B+1/2C, 1/2A+1/2B)$*; it illustrates the construction in the Appendix:*

$T_1 = \{A(1), A(2), B(1), B(2), C\}, T_2 = \{A(1), A(2), B\}$, $\sigma_1(A(1)) = \sigma_1(A(2)) = \sigma_2(A(1)) = \sigma_2(A(2)) = A$, $\sigma_1(B(1)) = \sigma_1(B(2)) = \sigma_2(B) = B$, $\sigma_1(C) = C$, *and* $\lambda \in \Delta(T_1 \times T_2)$ *is as follows:*

	A(1)	*A(2)*	*B*
A(1)	*1/128*	*7/128*	*1/16*
A(2)	*7/128*	*1/128*	*1/16*
B(1)	*2/128*	*6/128*	*1/16*
B(2)	*6/128*	*2/128*	*1/16*
C	*1/4*	*0*	*1/4*

Notice that the first order belief of player 2 at type $A(1)$ *is* $1/6A+1/6B+2/3C$*, at type* $A(2)$ *it is* $1/2A+1/2B$*, and at type* B *it is* $1/4A+1/4B+1/2C$*. Therefore, all types of player 2 are distinguished by first order beliefs. And clearly, all types of player 1 are distinguished by second order beliefs, while they all have first order belief* $1/2A+1/2B$*. Therefore,* (λ, T, σ) *is an intrinsic correlated equilibrium. And one can easily check that* (λ, T, σ) *obtains* $(1/4A+1/4B+1/2C, 1/2A+1/2B)$.

Example 4.4 (A non-intrinsic correlated equilibrium that is not Nash)**.**

The symmetric two-person game is as follows:

	A	*B*	*C*
A	*1, 1*	*0, 0*	*0, 0*
B	*0, 0*	*1, 1*	*0, 0*
C	*0, 0*	*0, 0*	*1, 1*

Consider the (asymmetric) correlated equilibrium of the game:

	A	*B*	*C*
A	*1/7*	*1/7*	*0*
B	*1/7*	*1/7*	*0*
C	*1/7*	*1/7*	*1/7*

$Q_1 = Q_2 = \{A, B, C\}$. For each $i \in \{1, 2\}$, $\beta_i^Q(A)$ is the convex hull spanned by extreme points A, $1/2A + 1/2B$, $1/2A + 1/2C$ and $1/3A + 1/3B + 1/3C$; and likewise for $\beta_i^Q(B)$ and $\beta_i^Q(C)$ (actions A, B and C are completely symmetric).

Therefore, we have that $Y_1 = Y_2 = \{A, B, C\}$, and $\mu(\cdot|a_i)$ is not injective on Y_i (for either i). Thus, this correlated equilibrium is not intrinsic. One can check that this correlated equilibrium is an extreme point in the set of correlated equilibria (see Proposition 4.6).

4.1 Geometric Properties

Proposition 4.5. *The set of intrinsic correlated equilibria is convex.*

Proof. Suppose that $\mu^1, \mu^2 \in \Delta(A)$ are two intrinsic correlated equilibria; for $\gamma \in (0, 1)$, let $\mu = \gamma\mu^1 + (1 - \gamma)\mu^2$.

For any $i \in N$, if $\mu^1(a_i) > 0$, $\mu^2(a_i) > 0$ and $\mu^1(\cdot|a_i) \neq \mu^2(\cdot|a_i)$, then $\mu(\cdot|a_i)$ is a strict convex combination of $\mu^1(\cdot|a_i)$ and $\mu^2(\cdot|a_i)$, so clearly $a_i \notin Y_i^1(\mu)$. Therefore, if $a_i \in Y_i^1(\mu)$, and $\mu^1(a_i) > 0$ (respectively, $\mu^2(a_i) > 0$), then we have that $\mu(\cdot|a_i) = \mu^1(\cdot|a_i)$ (respectively, $\mu(\cdot|a_i) = \mu^2(\cdot|a_i)$).

Let $Q_i^1 = \mathrm{supp}(\mathrm{marg}_{A_i}\,\mu^1)$ and $Q_i^2 = \mathrm{supp}(\mathrm{marg}_{A_i}\,\mu^2)$ for every $i \in N$. We thus have $Y_i^1(\mu) \cap Q_i^1 \subseteq Y_i^1(\mu^1)$ and $Y_i^1(\mu) \cap Q_i^2 \subseteq Y_i^1(\mu^2)$ for each $i \in N$. This implies that $Y_i(\mu) \cap Q_i^1 \subseteq Y_i(\mu^1)$ and $Y_i(\mu) \cap Q_i^2 \subseteq Y_i(\mu^2)$.

If $a_i \neq a_i' \in Y_i(\mu) \cap Q_i^1$, then $a_i \neq a_i' \in Y_i(\mu^1)$, and thus $\mu^1(\cdot|a_i) \neq \mu^1(\cdot|a_i')$. Therefore, we have $\mu(\cdot|a_i) \neq \mu(\cdot|a_i')$, since $\mu^1(\cdot|a_i) = \mu(\cdot|a_i)$ and $\mu^1(\cdot|a_i') = \mu(\cdot|a_i')$. And likewise for $a_i \neq a_i' \in Y_i(\mu) \cap Q_i^2$.

Now, suppose $a_i \neq a_i' \in Y_i^2(\mu)$ such that $a_i \in Q_i^1 \setminus Q_i^2$, $a_i' \in Q_i^2 \setminus Q_i^1$ and $\mu(\cdot|a_i) = \mu(\cdot|a_i')$. Then we have $\mu^1(\cdot|a_i) = \mu^2(\cdot|a_i')$. For any $a_j \in A_j$, $j \neq i$, such that $\mu^1(a_j|a_i) = \mu^2(a_j|a_i') > 0$, we have $a_j \in Y_j^1(\mu)$, which implies that $\mu(\cdot|a_j) = \mu^1(\cdot|a_j) = \mu^2(\cdot|a_j)$. But this implies that $\mu^1(a_i|a_j) = \mu(a_i|a_j) = \mu^2(a_i|a_j) > 0$, which contradicts $a_i \in Q_i^1 \setminus Q_i^2$.

Thus, we have that for any $i \in N$ and $a_i \neq a_i' \in Y_i(\mu)$, $\mu(\cdot|a_i) \neq \mu(\cdot|a_i')$; i.e. μ is an intrinsic correlated equilibrium. □

The following proposition shows that intrinsic correlated equilibrium is related to the notion of irreducibility and to extreme point in the set of correlated equilibria.

For a fixed correlated equilibrium $\mu \in \Delta(A)$, with $Q_i = \mathrm{supp}(\mathrm{marg}_{A_i}\,\mu)$ for $i \in N$, let $S = \bigcup_{i \in N} Q_i$. Two actions a^1 and a^k in S *communicate* (with each other) if $a^1 \in Q_{i_1}$, $a^k \in Q_{i_k}$, and there exists $a^m \in Q_{i_m}$, $2 \leq m \leq k - 1$, such that $i_m \neq i_{m-1} \in N$ and $\mu(a^m|a^{m-1}) > 0$ for each $2 \leq m \leq k$. Verbally, two actions communicate if they are connected by a sequence of intermediate actions in which μ places positive probability for every consecutive pair of actions. One can think of such consecutive pair of actions as a link; then two actions communicate if they are connected by a series of intermediate links.

It is readily checked that communication is an equivalence relation. Therefore, communication partitions S into equivalence classes (communication classes): $S = \bigcup_{1\leq k\leq n} S^k$, where every $S^k = \bigcup_{i\in N} Q_i^k$ and $\emptyset \neq Q_i^k \subseteq Q_i$. We say that the correlated equilibrium μ is *irreducible* if $n = 1$. For each $1 \leq k \leq n$, let $\mu^k(a) = \mu(a)/\mu(\prod_{i\in N} Q_i^k)$ for each $a \in \prod_{i\in N} Q_i^k$. It is clear that each μ^k is an irreducible correlated equilibrium, and μ can be written uniquely as convex combination of μ^k's. We say that μ^k is an *irreducible sub-equilibrium* of μ.

μ can be thought of as obtained from a public randomization over correlated equilibria $\mu^k, 1 \leq k \leq n$.

As a concrete illustration, the correlated equilibrium below (where $\{A, B, C, D\}$ are actions for each of the two players) has three irreducible sub-equilibria: AA, BB, and $1/4CC + 3/8CD + 1/8DC + 1/4DD$.

	A	B	C	D
A	1/4	0	0	0
B	0	1/4	0	0
C	0	0	1/8	3/16
D	0	0	1/16	1/8

Proposition 4.6. *Suppose that a correlated equilibrium μ has irreducible sub-equilibria μ^k, $1 \leq k \leq n$, and let $Q_i^k = \operatorname{supp}(\operatorname{marg}_{A_i} \mu^k)$ for each $i \in N$ and $1 \leq k \leq n$. Then,*

1. *For each $1 \leq k \leq n$, we either have $Y_i(\mu^k) = Q_i^k$ for all $i \in N$, or $Y_i(\mu^k) = \emptyset$ for all $i \in N$. And for each $i \in N$, $Y_i(\mu) = \bigcup_{1\leq k\leq n} Y_i(\mu^k)$.*

2. *If $Y_i(\mu^k) = Q_i^k$ for all $i \in N$ (e.g., when μ^k is not intrinsic), then μ^k is an extreme point in the polytope of correlated equilibria.*

3. *μ is intrinsic if and only if μ^k is intrinsic for every $1 \leq k \leq n$.*

Proof. 1 and 3 are immediate.

For 2, suppose μ is an irreducible correlated equilibrium, and $Y_i(\mu) = Q_i = \operatorname{supp}(\operatorname{marg}_{A_i} \mu)$ for each $i \in N$. We will show that μ is an extreme point in the set of correlated equilibria.

Suppose μ^1 and μ^2 are two correlated equilibrium such that $\mu = \mu^1/2+\mu^2/2$ and $\operatorname{supp}\mu^1 = \operatorname{supp}\mu^2 = \operatorname{supp}\mu$. Because $Y_i(\mu) = Q_i$, we must have $\mu^1(\cdot|a_i) = \mu^2(\cdot|a_i) = \mu(\cdot|a_i)$ for every $i \in N$ and $a_i \in Q_i$.

Suppose that $\mu^1 \neq \mu^2$, then there exists $a \in Q = \prod_{i \in N} Q_i$ such that $\mu^1(a) \neq \mu^2(a)$. Without loss of generality, suppose $\mu^1(a) < \mu^2(a)$. Because $\mu^1(\cdot|a_i) = \mu^1(\cdot|a_i)$ for every $i \in N$, we have that $\mu^1(b_{-i}, a_i) > 0 \Rightarrow \mu^1(b_{-i}, a_i) < \mu^2(b_{-i}, a_i)$ for every $i \in N$ and $b_{-i} \in Q_{-i}$. Because μ is irreducible, so are μ^1 and μ^2, and this together with the last sentence imply that $\mu^1(b) > 0 \Rightarrow \mu^1(b) < \mu^2(b)$ for every $b \in Q$, which clearly cannot be. Thus, we must have $\mu^1 = \mu^2$.

Therefore, μ is an extreme point in the set of correlated equilibria. $\square$

Thus, if two distinct irreducible correlated equilibria μ^1 and μ^2 (not necessarily themselves intrinsic) are such that there exist $i \in N$ and $a_i \in A_i$ such that $\mu^1(a_i) > 0$ and $\mu^2(a_i) > 0$, then $\gamma\mu^1 + (1-\gamma)\mu^2$ is an intrinsic correlated equilibrium for any $\gamma \in (0, 1)$.

4.2 Mixed Strategy Nash Equilibrium

Previous examples suggest that non-degenerate mixed Nash equilibrium (i.e. one that requires randomization for at least one player) is "typically" not an intrinsic correlated equilibrium. This is indeed the case, for generic two-person finite games. The class of generic games that we consider is usually associated with the Lemke-Howson algorithm, which is a simplex-like algorithm that computes Nash equilibrium in two-person games; a good reference is von Stengel (2002).

We say that a two-person game $(u, A = A_1 \times A_2, N = \{1, 2\})$ is *generic* if for any $i \in \{1, 2\}$ and $x \in \Delta(A_i)$, we have $|\operatorname{BR}_j(x)| \leq |\operatorname{supp}(x)|$, where $j \neq i$, $\operatorname{supp}(x) = \{a_i \in A_i : x(a_i) > 0\}$ and $\operatorname{BR}_j(x) = \{a_j \in A_j : u_j(a_j, x) \geq u_j(a'_j, x) \text{ for all } a'_j \in A_j\}$.

Proposition 4.7. *Fix a generic two-person game. Suppose $(x, y) \in \Delta(A_1) \times \Delta(A_2)$ is a non-degenerate mixed Nash equilibrium. Then (x, y) is not an intrinsic correlated equilibrium.*

Proof. Since (x, y) is a Nash equilibrium, we have $\operatorname{supp}(x) \subseteq BR_1(y)$ and $\operatorname{supp}(y) \subseteq BR_2(x)$. Thus, $|\operatorname{supp}(x)| + |\operatorname{supp}(y)| \leq |\operatorname{BR}_1(y)| + |\operatorname{BR}_2(x)|$. By the genericity of the game, we have $|\operatorname{supp}(x)| = |BR_2(x)|$ and $|\operatorname{supp}(y)| = |BR_1(y)|$.

Theorem 2.10 of von Stengel (2002) (which again uses the genericity condition) implies that the convex set $C = \{z \in \Delta(A_1) : \operatorname{supp}(z) = \operatorname{supp}(x) \text{ and } \operatorname{BR}_2(z) = \operatorname{BR}_2(x)\}$ is of dimension 0, i.e. $C = \{x\}$ [4]. Fix any $a_2 \in \operatorname{supp}(y)$, we claim that x is an extreme point of $\beta_2^A(a_2)$.

[4] More generally, Theorem 2.10 of von Stengel (2002) says that the convex set $\{z \in \Delta(A_1) : \operatorname{supp}(z) = \operatorname{supp}(x) \text{ and } \operatorname{BR}_2(z) = \operatorname{BR}_2(x)\}$ is of dimension $m - n$ for any $x \in \Delta(A_1)$, where $m = |\operatorname{supp}(x)|$ and $n = |\operatorname{BR}_2(x)|$.

Suppose otherwise, i.e. there exist $z_1 \neq z_2 \in \beta_2^A(a_2)$ such that $z_1/2 + z_2/2 = x$; we can choose z_1 and z_2 such that $\operatorname{supp}(z_1) = \operatorname{supp}(x) = \operatorname{supp}(z_2)$. And we have that $x \in \beta_2^A(a_2')$ implies that $z_1, z_2 \in \beta_2^A(a_2')$: if $z_1 \notin \beta_2^A(a_2')$, then we have

$$u_2(x, a_2') = u_2(z_1, a_2')/2 + u_2(z_2, a_2')/2 < u_2(z_1, a_2)/2 + u_2(z_2, a_2)/2 = u_2(x, a_2)$$

which means $x \notin \beta_2^A(a_2')$.

Thus, we have $\mathrm{BR}_2(x) \subseteq \mathrm{BR}_2(z_1) \cap \mathrm{BR}_2(z_2)$; this means that $BR_2(x) = \mathrm{BR}_2(z_1) = \mathrm{BR}_2(z_2)$, because $|\mathrm{BR}_2(z_1)| \leq |\operatorname{supp}(z_1)| = |\operatorname{supp}(x)| = |BR_1(x)|$ and likewise for z_2. Thus we have $z_1 \in C$ and $z_2 \in C$, which contradicts C being a singleton.

Likewise, y is an extreme point of $\beta_1^A(a_1)$ for every $a_1 \in \operatorname{supp}(x)$. Our desired conclusion then follows from the characterization of intrinsic correlated equilibrium in Theorem 4.1. □

Finally, for any finite game, it's easy to show that the iterated deletion procedure for Y_i^l, Equation (9), always ends in two rounds (i.e. $Y_i^2 = Y_i$ for all $i \in N$) if μ is a Nash equilibrium. Therefore, *if* a mixed Nash equilibrium is an intrinsic correlated equilibrium (e.g., Example 4.3), then we can be sure that in the equilibrium players need only to condition their actions on their second order beliefs, i.e. the equilibrium can be purified by second order beliefs. This suggests an inherent simplicity of Nash equilibrium, if it is an intrinsic correlated equilibrium.

5 Related Literature

Our paper is most related to Brandenburger and Friedenberg (2008). Brandenburger and Friedenberg study rationalizability in complete information game with correlation resulting from hierarchies of beliefs (intrinsic correlation).

They work with type space[5] $((\tilde{\lambda}_i)_{i \in N}, T)$, where $\tilde{\lambda}_i : T_i \to \Delta(T_{-i} \times A_{-i})$ for each $i \in N$, that is not necessarily obtained from Equation (3). Let l-th order belief map $\delta_i^l : T_i \to \mathcal{T}_i^l$ be defined as before, and let $\delta_i(t_i) = (\delta_i^1(t_i), \delta_i^2(t_i), \ldots)$ be the whole hierarchy of beliefs induced at type t_i.

Brandenburger and Friedenberg define intrinsic correlation of players' actions in a type space with the following notions of conditional independence and sufficiency.

[5] As before, we assume that each T_i is (non-empty) finite or countably infinite to avoid measurability issues.

In a type space $((\tilde{\lambda}_i)_{i\in N}, T)$, type $t_i \in T_i$ satisfies *conditional independence* (CI) if his belief about actions of other players is independent conditional on their hierarchies of beliefs; that is,

$$\tilde{\lambda}_i(t_i)(a_{-i}|\{\delta_{-i}(t_{-i}) = x_{-i}\}) = \prod_{j\neq i} \tilde{\lambda}_i(t_i)(a_j|\{\delta_{-i}(t_{-i}) = x_{-i}\})$$

for every actions $a_{-i} \in A_{-i}$ and hierarchies of beliefs $x_{-i} \in \prod_{j\neq i} \delta_j(T_j)$ such that $\tilde{\lambda}_i(t_i)(\{\delta_{-i}(t_{-i}) = x_{-i}\}) > 0$. Note that we abbreviate $\{t_{-i} \in T_{-i} : \delta_{-i}(t_{-i}) = x_{-i}\}$ as $\{\delta_{-i}(t_{-i}) = x_{-i}\}$.

And $t_i \in T_i$ satisfies *sufficiency* (SUFF) if he believes that player j's action ($j \neq i$) is influenced only by player j's belief hierarchy (and not influenced by belief hierarchies of other players); that is,

$$\tilde{\lambda}_i(t_i)(a_j|\{\delta_j(t_j) = x_j\}) = \tilde{\lambda}_i(t_i)(a_j|\{\delta_{-i}(t_{-i}) = x_{-i}\})$$

for every actions $a_j \in A_j$ and hierarchies of beliefs $x_{-i} \in \prod_{k\neq i} \delta_k(T_k)$ such that $\tilde{\lambda}_i(t_i)(\{\delta_{-i}(t_{-i}) = x_{-i}\}) > 0$.

Therefore, if both CI and SUFF hold at $t_i \in T_i$, then we have

$$\tilde{\lambda}_i(t_i)(a_{-i}|\{\delta_{-i}(t_{-i}) = x_{-i}\}) = \prod_{j\neq i} \tilde{\lambda}_i(t_i)(a_j|\{\delta_j(t_j) = x_j\})$$

for every actions $a_{-i} \in A_{-i}$ and hierarchies of beliefs $x_{-i} \in \prod_{j\neq i} \delta_j(T_j)$ such that $\tilde{\lambda}_i(t_i)(\{\delta_{-i}(t_{-i}) = x_{-i}\}) > 0$.

Going back to our model: $((\lambda_i)_{i\in N}, T, \sigma)$, where $\lambda_i : T_i \to \Delta(T_{-i})$ and $\sigma_i : T_i \to A_i$ for each $i \in N$, it's clear that if $\tilde{\lambda}_i$ is defined from λ_i and $(\sigma_j)_{j\neq i}$ via (3), and if condition (1) holds, then at every $t_i \in T_i$ of every player i, CI and SUFF hold. In particular, we have

$$\tilde{\lambda}_i(t_i)(a_{-i}|\{\delta_{-i}(t_{-i}) = x_{-i}\}) = \prod_{j\neq i} \mathbf{1}(a_j = \sigma_j(x_j))$$

for every actions $a_{-i} \in A_{-i}$ and hierarchies of beliefs $x_{-i} \in \prod_{j\neq i} \delta_j(T_j)$ such that $\lambda_i(t_i)(\{\delta_{-i}(t_{-i}) = x_{-i}\}) > 0$, where $\mathbf{1}(\cdot)$ is the indicator function, and $\sigma_j(x_j) := \sigma_j(t_j)$ where $\delta_j(t_j) = x_j$.

Following Tan and Werlang (1988), one defines the set of states (types and actions) of player i at which rationality and l-th order belief of rationality hold:

$$
\begin{aligned}
Rat_i^1(\tilde{\lambda}) &= \{(t_i, a_i) \in T_i \times A_i : a_i \text{ is optimal for player } i \text{ under } \operatorname{marg}_{A_{-i}} \tilde{\lambda}_i(t_i)\}, \\
Rat_i^l(\tilde{\lambda}) &= \{(t_i, a_i) \in Rat_i^1(\tilde{\lambda}) : \tilde{\lambda}_i(t_i)(Rat_{-i}^{l-1}(\tilde{\lambda})) = 1\}, l \geq 2, \\
Rat_i(\tilde{\lambda}) &= \bigcap_{l \geq 1} Rat_i^l(\tilde{\lambda})
\end{aligned}
$$

$Rat_i(\tilde{\lambda})$ is the set of states of player i at which rationality and common belief of rationality (RCBR) hold. Notice that $Rat_i^l(\tilde{\lambda})$ and $Rat_i(\tilde{\lambda})$ are defined with respect to a type space $((\tilde{\lambda}_i)_{i\in N}, T)$.

Brandenburger and Friedenberg are interested in the set of actions that are consistent with RCBR and intrinsic correlation (i.e. CI and SUFF):

$$
\begin{aligned}
C_i = \{a_i \in A_i : {} & \text{there exist } ((\tilde{\lambda}_i)_{i\in N}, T) \text{ such that at every type of every player, CI and SUFF hold,} \\
& \text{and } t_i \in T_i \text{ such that } (a_i, t_i) \in Rat_i(\tilde{\lambda})\}
\end{aligned}
$$

It is easy to check that if $((\lambda_i)_{i\in N}, T, \sigma)$ is an intrinsic a posteriori equilibrium, then $\sigma_i(T_i) \subseteq C_i$ for every $i \in N$.

Brandenburger and Friedenberg prove that $C = \prod_{i\in N} C_i$ is contained in the set of correlated rationalizable action profiles, and C contains the set of independent rationalizable action profiles. Furthermore, they show that there exist games in which C is strictly contained in the set of correlated rationalizable action profiles.

A precise characterization of the set C, in terms of payoffs and strategies of the game and without mentioning type space, is (and remains) an open question raised in Brandenburger and Friedenberg. Our Theorem 3.1 provides a partial answer: if $Q = \prod_{i\in N} Q_i$ is a semi-injective best-response set, then $Q \subseteq C$.

A contemporaneous paper by Peysakhovich (2009) provides another partial answer: if $\mu \in \Delta(A)$ is a correlated equilibrium, then actions of player i with positive probability by μ must be in C_i, i.e. $\operatorname{supp}(\operatorname{marg}_{A_i} \mu) \subseteq C_i$ for every $i \in N$.

Peysakhovich's result can be interpreted in our model as follows. Suppose as before that we have type space $((\lambda_i)_{i\in N}, T)$, where $\lambda_i : T_i \to \Delta(T_{-i})$ for each $i \in N$. We now allow mixed (or more accurately, behavioral) strategy: $\sigma_i : T_i \to \Delta(A_i)$. On the other hand, we insist on a more stringent "intrinsicness" condition: player can condition his randomized action only on his first order belief; that is,

$$\text{types } t_i \text{ and } t_i' \text{ have the same first order belief} \Longrightarrow \sigma_i(t_i) = \sigma_i(t_i'),$$

Peysakhovich proves that every correlated equilibrium $\mu \in \Delta(A)$ can be obtained from a (λ, T, σ) (where $\lambda \in \Delta(T), \lambda_i(t_i) := \lambda(\cdot|t_i)$), such that the above condition and the incentive compatibility condition (2) are both satisfied.

Therefore, we have an interesting trade-off between mixed strategy and higher order beliefs. On the one hand, every correlated equilibrium can be obtained from an incentive compatible type space (with common prior) in which every player plays a randomized action contingent on his first-order belief. On the other hand, "most" correlated equilibria (e.g., correlated equilibrium whose irreducible sub-equilibria are non-extreme) can be obtained from an incentive compatible type space (with common prior) in which every player plays a pure action contingent on his whole hierarchy of beliefs; that is, the player does not randomize, but he might have to rely on more refined information, i.e. his higher order beliefs.

6 Conclusion

Even if players sit in separate rooms and do not communicate or observe any signal, they might still display correlated equilibrium behaviors, because of their entangled beliefs of you believe that I believe that you believe that .... This paper analyzes the theory of such kind of correlated equilibrium.

APPENDIX

A Proof of If in Theorem 4.1

The proof of the if direction extensively uses the following lemma, whose proof we defer until the end of this section.

Lemma A.1. *Fix a finite and non-empty $X = \prod_{i \in N} X_i$ and a $\mu \in \Delta(X)$ such that $\mu(x_i) = \mu(\{x_i\} \times X_{-i}) > 0$ for every $i \in N$ and $x_i \in X_i$. And fix $(Z_i)_{i \in N}$, where each $Z_i \subseteq X_i$, and $\{(\nu(x_i, 1), \nu(x_i, 2))\}_{x_i \in Z_i, i \in N}$ such that for each $i \in N$ and $x_i \in Z_i$, $\nu(x_i, 1), \nu(x_i, 2) \in \Delta(X_{-i})$, and $\mu(\cdot|x_i) = \kappa\nu(x_i, 1) + (1 - \kappa)\nu(x_i, 2)$ for some $\kappa \in (0, 1)$.*

Let $\tilde{X} = \prod_{i \in N} \tilde{X}_i$, $\tilde{X}_i = \{x_i(k) : x_i \in Z_i, k \in \{1, 2\}\} \cup (X_i \setminus Z_i)$ (where $x_i(1)$ and $x_i(2)$ are two distinct copies of x_i). Define $f_i : \tilde{X}_i \to X_i$ such that $f_i(x_i) = x_i$ for $x_i \notin Z_i$, and

$f_i(x_i(1)) = f_i(x_i(2)) = x_i$ for $x_i \in Z_i$; define $f : \tilde{X} \to X$ and $f_{-i} : \tilde{X}_{-i} \to X_{-i}$ in the obvious way.

Then, there exists a $\tilde{\mu} \in \Delta(\tilde{X})$ such that $\tilde{\mu}(f^{-1}(x)) = \mu(x)$ for each $x \in X$, and $\tilde{\mu}(f_{-i}^{-1}(x_{-i})|x_i(k)) = \nu(x_i, k)(x_{-i})$ for every $i \in N$, $x_i \in Z_i$, $k \in \{1, 2\}$ and $x_{-i} \in X_{-i}$. Furthermore, if for every $i \in N$ and $x_i \in Z_i$, $\nu(x_i, 1)$ and $\nu(x_i, 2)$ have the same support as $\mu(\cdot|x_i)$, then for every $i \in N$, $x_i \in Z_i$ and $x_{-i} \in \tilde{X}_{-i}$, $\tilde{\mu}(x_i(1), x_{-i}) > 0$ if and only if $\tilde{\mu}(x_i(2), x_{-i}) > 0$ (if and only if $\mu(x_i, f_{-i}(x_{-i})) > 0$).

Suppose a correlated equilibrium $\mu \in \Delta(A)$ is given such that for every $i \in N$ and for any two distinct $a_i \neq a'_i \in Y_i$, we have that $\mu(\cdot|a_i) \neq \mu(\cdot|a'_i)$. We will construct an intrinsic correlated equilibrium (λ, T, σ) that obtains μ. For each $i \in N$ let Q_i be the support of $\mathrm{marg}_{A_i} \mu$. Our construction is to split each action $a_i \in Q_i \setminus Y_i$ into two copies (and making each copy a type with distinct belief hierarchy) using Lemma A.1; it works in opposite direction to the "amalgamation" construction in Aumann and Dreze (2008).

Step 1:

For each $i \in N$ and $a_i \in Q_i \setminus Y_i^1$, choose $\nu(a_i, 1) \neq \nu(a_i, 2) \in \beta_i^Q(a_i)$ such that $\mu(\cdot|a_i) = \nu(a_i, 1)/2 + \nu(a_i, 2)/2$ and that $\nu(a_i, 1)$ and $\nu(a_i, 2)$ have the same support as $\mu(\cdot|a_i)$. This is possible by construction of Y_i^1. Furthermore, we can choose $\nu(a_i, k)$'s in a way such that for every $i \in N$:

$$|\{\nu(a_i, k) : a_i \in Q_i \setminus Y_i^1, k \in \{1, 2\}\}| = 2|Q_i \setminus Y_i^1|$$

and

$$\{\nu(a_i, k) : a_i \in Q_i \setminus Y_i^1, k \in \{1, 2\}\} \cap \{\mu(\cdot|a_i) : a_i \in Y_i^1\} = \emptyset.$$

Now, apply Lemma A.1 to μ, Q, $(Q_i \setminus Y_i^1)_{i \in N}$ and $\{(\nu(a_i, 1), \nu(a_i, 2))\}_{a_i \in Q_i \setminus Y_i^1, i \in N}$ to obtain $T^1 = \prod_{i \in N} T_i^1$ (where $T_i^1 = \{a_i(k) : a_i \in Q_i \setminus Y_i^1, k \in \{1, 2\}\} \cup Y_i^1$), $\lambda^1 \in \Delta(T^1)$ and $f_i^1 : T_i^1 \to Q_i, i \in N$, with properties stated in the lemma. These properties implies that (λ^1, T^1, f^1) is a correlated equilibrium that obtains μ, and that each $a_i(j)$, $a_i \in Q_i \setminus Y_i^1$ and $j \in \{1, 2\}$, has a distinct first order belief through λ^1.

Step l: $(2 \leq l \leq L = \min\{l \geq 1 : Y^l = Y\})$

Suppose that $T^{l-1} = \prod_{i \in N} T_i^{l-1}$ (where $T_i^{l-1} = \{a_i(k) : a_i \in Q_i \setminus Y_i^{l-1}, k \in \{1, 2\}\} \cup Y_i^{l-1}$), $\lambda^{l-1} \in \Delta(T^{l-1})$ and $f_i^{l-1} : T_i^{l-1} \to T_i^{l-2}, i \in N$, (let $T_i^0 = Q_i$) are obtained from Lemma A.1 in the previous step.

For each $i \in N$ and $a_i \in Y_i^{l-1} \setminus Y_i^l$, choose a $c(a_i) \in Y_j^{l-2} \setminus Y_j^{l-1}$, $j \neq i$, (let $Y_j^0 = Q_j$) such that $\mu(c(a_i)|a_i) > 0$; such $c(a_i)$ exists by construction of Y_i^l's, and $c(a_i)$'s can be chosen so that $\mu(\cdot|a_i) = \mu(\cdot|a'_i) \Rightarrow c(a_i) = c(a'_i)$. For each $t_{-(i,j)} \in T_{-(i,j)}^{l-1} = \prod_{k \notin \{i,j\}} T_k^{l-1}$, we have

$\lambda^{l-1}(t_{-(i,j)}, c(a_i)(1), a_i) > 0$ if and only if $\lambda^{l-1}(t_{-(i,j)}, c(a_i)(2), a_i) > 0$ (by Lemma A.1); and $\lambda^{l-1}(\{c(a_i)(1), c(a_i)(2)\} \times \{a_i\} \times T^{l-1}_{-(i,j)}) = \mu(c(a_i), a_i) > 0$. Let

$$\nu(a_i,1)(t_{-i}) = \begin{cases} \lambda^{l-1}(t_{-i}|a_i) & \lambda^{l-1}(t_{-i}|a_i) = 0 \text{ or } t_j \notin \{c(a_i)(1), c(a_i)(2)\} \\ \lambda^{l-1}(t_{-(i,j)}, c(a_i)(1)|a_i) - \kappa(a_i) & \lambda^{l-1}(t_{-i}|a_i) > 0 \text{ and } t_j = c(a_i)(1) \\ \lambda^{l-1}(t_{-(i,j)}, c(a_i)(2)|a_i) + \kappa(a_i) & \lambda^{l-1}(t_{-i}|a_i) > 0 \text{ and } t_j = c(a_i)(2) \end{cases},$$

and

$$\nu(a_i,2)(t_{-i}) = \begin{cases} \lambda^{l-1}(t_{-i}|a_i) & \lambda^{l-1}(t_{-i}|a_i) = 0 \text{ or } t_j \notin \{c(a_i)(1), c(a_i)(2)\} \\ \lambda^{l-1}(t_{-(i,j)}, c(a_i)(1)|a_i) + \kappa(a_i) & \lambda^{l-1}(t_{-i}|a_i) > 0 \text{ and } t_j = c(a_i)(1) \\ \lambda^{l-1}(t_{-(i,j)}, c(a_i)(2)|a_i) - \kappa(a_i) & \lambda^{l-1}(t_{-i}|a_i) > 0 \text{ and } t_j = c(a_i)(2) \end{cases},$$

for every $t_{-i} \in T^{l-1}_{-i}$, where $\kappa(a_i) > 0$ is sufficiently small so that $\nu(a_i, 1)$ and $\nu(a_i, 2)$ has the same support as $\mu^{l-1}(\cdot|a_i)$. Notice that $\nu(a_i, 1)/2 + \nu(a_i, 2)/2 = \lambda^{l-1}(\cdot|a_i)$. Furthermore, we can choose the $\kappa(a_i)$'s so that for any $a_i \neq a_i' \in Y_i^{l-1} \setminus Y_i^l$ such that $\mu(\cdot|a_i) = \mu(\cdot|a_i')$, we have that $\nu(a_i, 1), \nu(a_i, 2), \nu(a_i', 1)$ and $\nu(a_i', 2)$ all differ from each other in their probabilities on $c(a_1)(1)$.

Now, apply Lemma A.1 to λ^{l-1}, T^{l-1}, $(Y_i^{l-1} \backslash Y_i^l)_{i \in N}$ and $\{(\nu(a_i, 1), \nu(a_i, 2))\}_{a_i \in Y_i^{l-1} \backslash Y_i^l, i \in N}$ to obtain $T^l = \prod_{i \in N} T_i^l$ (where $T_i^l = \{a_i(k) : a_i \in Q_i \setminus Y_i^l, k \in \{1, 2\}\} \cup Y_i^l$), $\lambda^l \in \Delta(T^l)$ and $f_i^l : T_i^l \to T_i^{l-1}, i \in N$, with properties stated in the lemma. These properties imply that $(\lambda^l, T^2, f^1 \circ \cdots \circ f^l)$ is a correlated equilibrium that obtains μ, and that each $a_i(k)$, $a_i \in Y_i^{l-1} \setminus Y_i^l$ and $k \in \{1, 2\}$, induces a distinct l-th order belief through λ^l.

Finally:

Let $T = T^L$ ($T_i = T_i^L = \{a_i(k) : a_i \in Q_i \backslash Y_i, k \in \{1, 2\}\} \cup Y_i$), $\lambda = \lambda^L$, and $\sigma_i = f_i^1 \circ \ldots \circ f_i^L$. It's easy to see that that (λ, T, σ) is an intrinsic correlated equilibrium that obtains μ.

Proof of Lemma A.1. Without loss of generality suppose that $N = \{1, \ldots, n\}$.

Let $\mu^1 \in \Delta(\tilde{X}_1 \times \prod_{2 \leq i \leq n} X_i)$ be such that

$$\mu^1(x_1(1), x_{-1}) = \mu(x_1)\kappa(x_1)\nu(x_1, 1)(x_{-1})$$

and

$$\mu^1(x_1(2), x_{-1}) = \mu(x_1)(1 - \kappa(x_1))\nu(x_1, 2)(x_{-1}),$$

where $\mu(\cdot|x_1) = \kappa(x_1)\nu(x_1, 1) + (1 - \kappa(x_1))\nu(x_1, 2)$, for each $x_1 \in Z_1$ and $x_{-1} \in X_{-1}$.

And let $\mu^1(x_1, x_{-1}) = \mu(x_1, x_{-1})$ for every $x_1 \notin Z_1$ and $x_{-1} \in X_{-1}$.

In general, for $2 \leq l \leq n$, let $\mu^l \in \Delta(\prod_{1 \leq j \leq l} \tilde{X}_j \times \prod_{l+1 \leq i \leq n} X_i)$ be such that for every $x_l \in Z_l$, $(x_1, \dots, x_{l-1}) \in \prod_{1 \leq i \leq l-1} \tilde{X}_i$ and $(x_{l+1}, \dots, x_n) \in \prod_{l+1 \leq i \leq n} X_i$:

$$\begin{aligned}\mu^l(x_1, \dots, x_{l-1}, x_l(1), x_{1+1}, \dots, x_n) =& \mu(x_l)\kappa(x_l)\frac{\mu^{l-1}(x_1, \dots, x_{l-1}, x_l, \dots, x_n)}{\mu(f_1(x_1), \dots, f_{l-1}(x_{l-1}), x_l, \dots, x_n)} \\ &\times \nu(x_l, 1)(\mu(f_1(x_1), \dots, f_{l-1}(x_{l-1}), x_{l+1}, \dots, x_n)\end{aligned}$$

and

$$\begin{aligned}\mu^l(x_1, \dots, x_{l-1}, x_l(2), x_{1+1}, \dots, x_n) =& \mu(x_l)(1 - \kappa(x_l))\frac{\mu^{l-1}(x_1, \dots, x_{l-1}, x_l, \dots, x_n)}{\mu(f_1(x_1), \dots, f_{l-1}(x_{l-1}), x_l, \dots, x_n)} \\ &\times \nu(x_l, 2)(\mu(f_1(x_1), \dots, f_{l-1}(x_{l-1}), x_{l+1}, \dots, x_n),\end{aligned}$$

if $\mu(f_1(x_1), \dots, f_{l-1}(x_{l-1}), x_l, \dots, x_n) > 0$, and

$$\mu^l(x_1, \dots, x_{l-1}, x_l(1), x_{1+1}, \dots, x_n) = \mu^l(x_1, \dots, x_{l-1}, x_l(2), x_{1+1}, \dots, x_n) = 0$$

otherwise, where $\mu(\cdot|x_l) = \kappa(x_l)\nu(x_l, 1) + (1 - \kappa(x_l))\nu(x_l, 2)$.

And let

$$\mu^l(x_1, \dots, x_{l-1}, x_l, x_{1+1}, \dots, x_n) = \mu^{l-1}(x_1, \dots, x_{l-1}, x_l, x_{1+1}, \dots, x_n)$$

for every $x_l \notin Z_l$, $(x_1, \dots, x_{l-1}) \in \prod_{1 \leq i \leq l-1} \tilde{X}_i$ and $(x_{l+1}, \dots, x_n) \in \prod_{l+1 \leq i \leq n} X_i$.

It is easy to verify that $\tilde{\mu} = \mu^n$ satisfies the desired properties.

□

References

[1] Aumann, R. J. (1974) Subjectivity and Correlation in Randomized Strategies. Journal of Mathematical Economics, Volume 1, Issue 1, pp. 67-96.

[2] Aumann, R. J., and J. H. Dreze (2008) Rational Expectations in Games. American Economic Review, 98(1), pp. 72-86.

[3] Bernheim, B. D. (1984) Rationalizable Strategic Behavior. Econometrica, Volume 52, No. 4, pp. 1007-1028.

[4] Brandenburger, A. and E. Dekel (1987) Rationalizability and Correlated Equilibria. Econometrica, Volume 55, No. 6, pp. 1391-1402.

[5] Brandenburger, A. and A. Friedenberg (2008) Intrinsic Correlation in Games. Journal of Economic Theory, Volume 141, Issue 1, pp. 28-67.

[6] Kohlberg, E. and J. F. Mertens (1986) On the Strategic Stability of Equilibria. Econometrica, Volume 54, No. 5, pp. 1003-1037.

[7] Mertens, J. F., and S. Zamir (1985) Formulation of Bayesian Analysis for Fames with Incomplete Information. International Journal of Game Theory, Volume 14, No. 1, pp. 1-29.

[8] Myerson, R. B. (1997) Dual Reduction and Elementary Games. Games and Economic Behavior, Volume 21, Issue 1-2, pp. 183-202.

[9] Pearce, D. G. (1984) Rationalizable Strategic Behavior and the Problem of Perfection. Econometrica, Volume 52, No. 4, pp. 1029-1050.

[10] Peysakhovich, A. (2009) Correlation Without Signals. Memeo.

[11] Siniscalchi, M. (2007) Epistemic Game Theory: Beliefs and Types. The New Palgrave Dictionary of Economics, Second Edition.

[12] Tan, T. and S. Werlang (1988) The Bayesian Foundations of Solution Concepts of Games. Journal of Economic Theory, Volume 45, Issue 2, pp. 370-391.

[13] von Stengel, B. (2002) Computing equilibria for two-person games. Chapter 45, Handbook of Game Theory, Vol. 3, eds. R. J. Aumann and S. Hart, North-Holland, Amsterdam, pp. 1723-1759.