Correlated Equilibrium via Hierarchies of Beliefs

Songzi Du *

Stanford GSB

First Draft: December 8, 2008
This Draft: February 8, 2010

Abstract

We study a formulation of correlated equilibrium in which every player conditions his actions on his hierarchies of beliefs about the play of the game (belief on what other players will do, on what other players believe others will do, etc.). Our formulation can be thought of as a purification requirement based on hierarchies of beliefs. For any finite, complete information game, we are able to exactly characterize the strategic implications of correlated equilibria, both subjective and objective, in which players condition their actions on their hierarchies of beliefs. The characterizations are independent of type space and rely on a novel iterated deletion procedure. We show that "most" (objective) correlated equilibrium distributions can be obtained conditioning on hierarchies of beliefs; but interestingly, for generic two-person games, any non-degenerate mixed-strategy Nash equilibrium cannot be obtained. Therefore, we can purify "most" public randomizations, but not private randomizations, via hierarchies of beliefs.

JEL classification: C72, D83

Keywords: Correlated Equilibrium, Hierarchy of Beliefs, Purification, Common Prior Assumption

*songzidu@stanford.edu. I am indebted to Adam Brandenburger and Amanda Friedenberg for the suggestion to rephrase the formulation to players taking the same action on redundant types. And I thank Yossi Feinberg, Qingmin Liu, Yair Livne, Andy Skrzypacz, and Robert Wilson for valuable advices, comments and discussions.

1 Introduction

There are two views on correlation in non-cooperative game theory. The classical view introduced by Aumann (1974) relies on external, payoff irrelevant signals: players display correlated behavior because they condition their strategies on correlated signals.

Recently, Brandenburger and Friedenberg (2008) introduced an *intrinsic* view of correlation based on players' hierarchies of beliefs (belief, belief on others' belief, and so on) about the play of the game. According to this view, a player has correlated belief about strategies of the other players because he has a hierarchy of belief which expresses correlation over other players' hierarchies of beliefs. This view of correlation is intrinsic, because according to epistemic game theory, hierarchies of beliefs are an intrinsic part of the description of the game.

It seems quite natural that correlation can arise out of players' intertwined hierarchies of beliefs (I believe that you believe that I believe that ...). The hierarchies of beliefs may come from the players' previous interactions, or the players may simply be very imaginative people. Moreover, intrinsic correlation is in a sense more primitive than extrinsic correlation: extrinsic correlation can be "embedded" in the game by expanding strategy sets and adjusting the payoffs to explicitly account for players' conditioning on the relevant signals; but then players might display further correlated behaviors in the expanded strategies because of their correlated hierarchies of beliefs (about the play in the expanded strategies).

Brandenburger and Friedenberg formalized intrinsic correlation with two epistemic conditions (sufficiency and conditional independence) in type spaces; their solution concept is the combined implications of these two conditions together with the condition rationality and common belief of rationality. A drawback of their approach is that it is dependent on the type space. Since a type space specifies every player's (potentially infinite) hierarchy of beliefs about the play of the game, it might be cumbersome or difficult to construct in applications and thus to apply Brandenburger and Friedenberg's solution concept. In fact, Brandenburger and Friedenberg posed a type-free characterization of their solution concept (i.e., a characterization solely in terms of strategies and payoffs of the game) as an open question.

In this paper we introduce a new formulation to study intrinsic correlation and related phenomena, and obtain an exact characterization that is independent of type space. Brandenburger and Friedenberg rely on rationality and common belief of rationality, which lead to a *rationalizability*-like solution concept. In contrast, we propose an *equilibrium* solution concept, consistent with Aumann (1974)'s correlated equilibrium, that players condition their (pure) actions on their hierarchies of beliefs about the play of the game, i.e., for player i's pure strategy σ_i that maps types to actions,

$$\begin{array}{c}\text{types } t_i \text{ and } t_i' \text{ have the same hierarchy of} \\ \text{beliefs about the play of the game}\end{array} \implies \sigma_i(t_i) = \sigma_i(t_i'). \tag{$*$}$$

We study the combined implication of condition $(*)$ and the rationality condition that every player chooses the optimal action given his belief about the play of the game. A conventional correlated equilibrium is one in which players condition their actions on types while satisfying the rationality condition, but condition $(*)$ does not apply; there, types are external signals or sunspots. By condition $(*)$, types on which players condition their actions are exactly hierarchies of beliefs (on the play of the game), which are variables "inside" the game. Thus, we call our solution concept *intrinsic correlated equilibrium*.

Hierarchies of beliefs in condition $(*)$ may or may not be consistent with a common prior. We analyze both cases; the characterizations in the two cases turn out to be closely related. When players' hierarchies of beliefs come from some previous interaction, they are likely to be consistent with a common prior. For example, if Ann stole money from Bob and Clare in their previous business venture, then for the present interaction Bob's beliefs about Ann's play (say cooperate or defect) would probably be very consistent with Clare's beliefs about Ann's play, and so on. On the other hand, if players' hierarchies of beliefs come from mere introspection, then they are likely to be subjective, not necessarily consistent with a common prior.

Since we insist on every player playing a pure action conditional on his hierarchy of beliefs, condition $(*)$ in fact goes beyond requiring intrinsic correlation. It says that for a fixed equilibrium, conditional on a hierarchy of beliefs, a deterministic action of the player will be played, and this is commonly believed by the players and is reflected in their hierarchies of beliefs. In particular, players do no use private randomization, unless the randomization is written explicitly as an action in the game. Thus, not only correlation, but every strategic uncertainty in the game is traced back to players' hierarchies of beliefs about the play of the game. Player i is uncertain about the other players' actions, because he is uncertain about their hierarchies of beliefs; each uncertainty (i.e., a non-degenerate belief) about other players' hierarchies of beliefs is exactly one of player i's hierarchies of beliefs. This is obviously a strong restriction, yet we show that "most" of the conventional (objective) correlated equilibrium distributions are consistent with this restriction. And while players not using private randomization is a reasonable assumption in many settings, e.g. when playing one-shot games, we do examine what happens when private randomization is allowed in Section 5.2.

Moreover, condition $(*)$ can be seen as a purification condition. The purification literature (following Harsanyi (1973)) assumes that there are vanishing payoff shocks, and each player conditions his pure actions on his shocks (i.e., on his realized payoffs). In contrast, we assume no payoff shocks and assume that each player conditions his pure actions on his hierarchies of beliefs about the play of the game. While we are able to obtain "most" correlated equilibrium distributions in this fashion, for generic two-person games we cannot obtain non-degenerate mixed-strategy Nash equilibria, i.e. they are non-purifiable. In contrast, Harsanyi (1973) proved that for this generic class of games all mixed-strategy Nash equilibria are purifiable via payoff shocks. We interpret this difference as showing that to purify mixed-strategy Nash equilibrium one must go "outside" of the game, either by introducing private randomization or via vanishing payoff shocks.

In relation to Brandenburger and Friedenberg (2008), our solution concept, in the subjective case, is a refinement of theirs. Therefore, we obtain a new sufficient condition (semi-injective best-response set; cf. Definition 3.1) that is independent of type space for intrinsic correlation. A contemporaneous and independent paper by Peysakhovich (2009) provides another type-free sufficient condition: any action in the support of a (objective) correlated equilibrium distribution is consistent with intrinsic correlation. Incidentally, Peysakhovich's result has a natural interpretation in our formulation when we allow for private randomization.

The paper proceeds as follows. In the next section we formally specify our set-up. Section 3 studies the strategic implications of condition $(*)$ when the hierarchies of beliefs of players are subjective, i.e., not necessarily consistent with a common prior. Section 4 studies the analogous implications when the hierarchies of beliefs are required to be consistent with a common prior. Section 5.1 discusses the relationship to Brandenburger and Friedenberg (2008) and shows a private-randomization extension of our result based on Peysakhovich (2009). Section 6 concludes the paper.

2 Set-up

2.1 Notations

We use the following standard notations:

For product set $X = \prod_{i\in N} X_i$, let $X_{-j} = \prod_{i\neq j} X_i$. Likewise, for $x \in X$, let $x_{-i} = (x_j)_{j\neq i}$. And for $f_i : X_i \to Y_i, i \in N$, we write $f_{-i}(x_{-i}) = (f_j(x_j))_{j\neq i}$.

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

For $\mu \in \Delta(X)$ where $X = \prod_{i\in N} X_i$ is countable, let $\mu(x_i) = \mu(\{x_i\} \times X_{-i})$. For $\mu(x_i) > 0$, let $\mu(\cdot|x_i) \in \Delta(X_{-i})$ be μ conditional on the event $\{x_i\} \times X_{-i}$, and let $\mu(x_j|x_i) = \mu(\{x_j\} \times \prod_{k\notin\{i,j\}} X_k\,|x_i)$; likewise, let $\mu(x_j, x_i) = \mu(\{x_j\} \times \{x_i\} \times \prod_{k\notin\{i,j\}} X_k)$.

We write $x \neq y \in X$ to mean that $x \in X, y \in X$ and $x \neq y$.

Throughout this paper we identify a Nash equilibrium with its uniquely induced (correlated equilibrium) distribution on action profiles. And we abbreviate *best-response set* (cf. Definition 2.3) to BRS and *correlated equilibrium distribution* (cf. Equation (3)) to CED.

2.2 Set-up

We fix a (arbitrary) 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 that captures players' strategic uncertainty in (u, A, N). Formally, let $((\lambda_i)_{i\in N}, T)$, where $T = \prod_{i\in N} T_i$ is a (non-empty) finite or countably infinite[1] set of type profiles, and $\lambda_i : T_i \to \Delta(T_{-i})$ is player i's belief (i.e., a probability measure), contingent on his type, about types of other players.

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 $a_i \in A_i$:

$$\sum_{t_{-i}\in T_{-i}} u_i(\sigma_i(t_i), \sigma_{-i}(t_{-i}))\lambda_i(t_i)(t_{-i}) \geq \sum_{t_{-i}\in T_{-i}} u_i(a_i, \sigma_{-i}(t_{-i}))\lambda_i(t_i)(t_{-i}) \tag{1}$$

[1] 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 and require λ_i and σ_i to be measurable.

Definition 2.1. $((\lambda_i)_{i\in N}, T, \sigma)$ *is an* a posteriori equilibrium *if (1) 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 (1) is satisfied for* $\lambda_i(t_i) := \lambda(\cdot|t_i)$.

Correlated equilibrium differs 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 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

$$\tilde{\lambda}_i(t_i)(t_{-i}, a_{-i}) = \begin{cases} \lambda_i(t_i)(t_{-i}) & \text{if } \sigma_{-i}(t_{-i}) = a_{-i} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

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 (first order belief), his belief about their beliefs about others' actions (second order belief), his belief about others' beliefs about others' beliefs (third order belief), and so on. The following formulation of hierarchy of beliefs is standard: see for example Siniscalchi (2007) and Brandenburger and Friedenberg (2008). The set of all hierarchies of beliefs forms an universal type space in which every player i has basic uncertainty A_{-i} [2].

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.

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; thus, a second order belief contains strictly more information than first order belief. And in general, a l-th order belief is a joint probability over other players' actions and other players' $(l-1)$-th order beliefs.

[2] 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.

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 $(\delta_j^{l-1}, \mathrm{id}_{A_j})_{j\neq i}$, 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., $(\delta_j^{l-1}, \mathrm{id}_{A_j})_{j\neq i}(t_{-i}, a_{-i}) = (\delta_j^{l-1}(t_j), \mathrm{id}_{A_j}(a_j))_{j\neq i}$. That is, for any Borel measurable $B \subseteq \mathcal{T}_{-i}^{l-1} \times A_{-i}$, $\delta_i^l(t_i)(B) = \tilde{\lambda}_i(t_i)((\delta_j^{l-1}, \mathrm{id}_{A_j})_{j\neq i}^{-1}(B))$.

$(\delta_i^1(t_i), \delta_i^2(t_i), \delta_i^3(t_i), \ldots)$ is the *hierarchy of beliefs* (or *belief hierarchy*) of type t_i. The hierarchy of beliefs is a complete and canonical description of player i's state of mind (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*.

Before moving on, let us illustrate redundant types with an example.

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.

We now come to the objectives of the paper:

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. $\delta^l_i(t_i) = \delta^l_i(t'_i)$ 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^l_i's are defined with respect to $\lambda_i(t_i) := \lambda(\cdot|t_i)$.*

In other words, "intrinsicness" in the above definition rules out player i in an equilibrium playing different actions at types that have the same hierarchy of beliefs; such types must be extrinsic, because player i cannot distinguish them by thinking "inside" the game, i.e., thinking about other players' actions, about what others think about others' actions, about what others think about what others think, and so on.

Notice 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 the 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 call $\mu \in \Delta(A)$ an intrinsic correlated equilibrium distribution (respectively, a correlated equilibrium distribution) if μ is obtained from an intrinsic correlated equilibrium (respectively, a correlated equilibrium) (λ, T, σ). We abbreviate correlated equilibrium distribution to CED.

Our goal is to understand the strategic implications of the "intrinsicness" in Definition 2.2. Therefore, we need to review the implications of the equilibrium when it is not required to be intrinsic. In the next two sections we will work out in strategic terms the exact strengthening added by "intrinsicness."

Bernheim (1984) and Pearce (1984) in their studies of rationalizability introduced the concept of *best-response set* (BRS):

Definition 2.3. *A set of action profiles $Q = \prod_{i\in N} Q_i$ is a* best response set (BRS) *if it is non-empty, and for every $i \in N$ and $a_i \in Q_i$, there exists a belief $\mu \in \Delta(Q_{-i})$ such that a_i is optimal in A_i for player i under μ.*

It is well-known (Brandenburger and Dekel, 1987) that for any 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 a *correlated equilibrium distribution* (CED) if and only if

$$\sum_{a_{-i} \in A_{-i}} u_i(a_i, a_{-i}) \mu(a_i, a_{-i}) \geq \sum_{a_{-i} \in A_{-i}} u_i(a_i', a_{-i}) \mu(a_i, a_{-i}) \tag{3}$$

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

3 Intrinsic A Posteriori Equilibrium

3.1 Characterization

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

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

$$\beta_i^Q(a_i) = \{\mu \in \Delta(Q_{-i}) : a_i \text{ is optimal in } A_i \text{ for player } i \text{ under } \mu\}, \tag{4}$$

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

$$\begin{aligned} W_i^1 &= \{a_i \in Q_i : |\beta_i^Q(a_i)| = 1\}, \\ W_i^l &= \{a_i \in W_i^1 : \beta_i^Q(a_i)(W_{-i}^{l-1}) = 1\}, l \geq 2, \\ W_i &= \bigcap_{l \geq 1} W_i^l. \end{aligned} \tag{5}$$

Notice that W_i^l's are defined with respect to a fixed Q. We write $W_i(Q)$ and $W_i^l(Q)$ when it is necessary to emphasize the dependence on Q.

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 $W^1 = \prod_{i \in N} W_i^1$.

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, we have $\beta_i^Q(a_i) \neq \beta_i^Q(a_i')$. It is an* injective BRS *if for every $i \in N$ and any two distinct actions a_i and a_i' in W_i^1, we have $\beta_i^Q(a_i) \neq \beta_i^Q(a_i')$.*

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.

The concept of injective BRS is due to Brandenburger and Friedenberg; their original definition defines *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). Clearly, this is equivalent to our definition.

Brandenburger and Friedenberg (Proposition H.2) proved that (in our language) for any 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, which is the main result of Section 3:

Theorem 1. *For any 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.*

Theorem 1 has the following finite-level version. When $l = 1$, (6) is Brandenburger and Friedenberg's injectivity condition; when $l = \infty$ (and let $W_i^\infty = W_i$), (6) is our semi-injectivity condition.

Theorem 1 (Finite-level version)**.** *Fix a $l \geq 1$ and a BRS $Q = \prod_{i \in N} Q_i$. If for every player i, we have*

$$a_i, a_i' \in W_i^l, a_i \neq a_i' \Longrightarrow \beta_i^Q(a_i) \neq \beta_i^Q(a_i'), \tag{6}$$

then there exists an a posteriori equilibrium $((\lambda_i)_{i \in N}, T, \sigma)$ in which players condition their actions on their l-th order beliefs (i.e., $\delta_i^l(t_i) = \delta_i^l(t_i') \Rightarrow \sigma_i(t_i) = \sigma_i(t_i')$)), and under which Q is played (i.e., $Q_i = \sigma_i(T_i)$ for every player i).

Conversely, if players condition their actions on their l-th order beliefs in an a posteriori equilibrium, and Q is played under the equilibrium, then (6) holds for every player i.

The theorem implies that if the iterated deletions in (5) 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 (5) 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 translates into the requirement that each action a_i in W_i has a distinct supporting belief $\beta_i^Q(a_i)$.

Proof of Theorem 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 (2).

$Q = \prod_{i\in N} Q_i$ 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 3.1. *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 $\sigma_i(t_i) = \sigma_i(t_i') = a_i \in W_i^l, t_i, t_i' \in T_i$. 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 $a_{-i} \in W^{l-1}_{-i}$ by the construction of W^l_i. By the induction hypothesis, $\delta^{l-1}_j(t_j) = \delta^{l-1}_j(t'_j)$ for every $j \neq i$. Thus, $\delta^l_i(t_i) = \delta^l_i(t'_i)$. $\square$

Corollary 3.2. *For every $i \in N$ and $\mu \in \Delta(W_{-i})$, there can be at most one belief hierarchy in T_i having first order belief μ, i.e. if $\operatorname{marg}_{A_{-i}} \tilde{\lambda}_i(t_i) = \mu = \operatorname{marg}_{A_{-i}} \tilde{\lambda}_i(t'_i)$, then $\delta^l_i(t_i) = \delta^l_i(t'_i)$ for every $l \geq 1$.*

Proof. Suppose $\mu \in \Delta(W_{-i})$ and $\operatorname{marg}_{A_{-i}} \tilde{\lambda}_i(t_i) = \mu = \operatorname{marg}_{A_{-i}} \tilde{\lambda}_i(t'_i), t_i, t'_i \in T_i$. 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 $l \geq 1$. $\square$

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 $\operatorname{marg}_{A_{-i}} \tilde{\lambda}_i(t_i) = \beta^Q_i(a_i)$ and $\operatorname{marg}_{A_{-i}} \tilde{\lambda}_i(t'_i) = \beta^Q_i(a'_i)$; 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 semi-injective BRS. Let W^l_i and W_i be as defined in (5).

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 $Q_i \setminus W_i^1$ and $\beta_i^Q(W_i^1)$ are finite sets, but $\beta_i^Q(a_i)$ is infinite for any $a_i \in Q_i \setminus W_i^1$ (recall that $\beta_i^Q(a_i)$ 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, where player 1 chooses the row and player 2 chooses the column:*

	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*. Clearly,* $W_1 = W_2 = \{A, B\}$*, 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 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.*

Example 3.2 (A semi-injective BRS that is not injective)**.** *Now consider the following modification of the previous game:*

	X	Y	Z	W
A	1, 0	3, 0	0, 1	0, 1
B	3, 0	1, 0	0, 1	0, 1
C	0, 1	4, 3	1, 4	1, 0
D	4, 3	0, 1	1, 0	1, 4

The payoffs of player 1 are unchanged while the payoffs of player 2 are "permuted."

Let $Q_1 = \{A, B, C, D\}$ *and* $Q_2 = \{X, Y, Z, W\}$*. As before,* $Q = Q_1 \times Q_2$ *is a BRS, but it's not injective because* A *and* B *are uniquely supported by the same* $1/2X + 1/2Y$*; likewise,* X *and* Y *are uniquely supported by the same* $1/2C + 1/2D$*. However,* $W_2^2 = \emptyset$*, because the* *unique supporting belief of X and Y places positive probability on C which is not in W_1^1; and $W_2^2 = \emptyset$ implies that $W_2^3 = \emptyset$. Thus, Q is a semi-injective BRS.*

Finally, notice that if player 1 has the same payoff as here, while player 2 has generic payoffs, then Q is a semi-injective BRS that is not injective.

3.2 Weak Domination

Example 3.1 alludes to a connection, of which we will now show, between intrinsic a posteriori equilibrium and weakly dominated actions.

Recall the result of Brandenburger and Dekel (1987): for any 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. For a player i and $B \subseteq A_{-i}$, we say that i's action a_i is weaked dominated in B if there exists $\alpha_i \in \Delta(A_i)$ such that $u_i(a_i, a_{-i}) \le u_i(\alpha_i, a_{-i})$ for every $a_{-i} \in B$, with strict inequality for some $a_{-i} \in B$

Recall that $W_i \subseteq Q_i$ is defined with respect to $Q = \prod_{i\in N} Q_i$ by Equation (5).

Proposition 3.3. *Suppose that a 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}.*

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)). $\square$

The next proposition shows that if Q is the set of correlated rationalizable action profiles (the largest BRS), then we can dispense with the assumption of $W_j \subsetneq Q_j$.

Proposition 3.4. *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}.*

Proof. By the definition of $\beta_i^{A_i \times X_{-i}}$ (cf. Equation (4)), we have $\delta(X_{-i}) = \bigcup_{a_i \in A_i} \beta_i^{A_i \times X_{-i}}(a_i)$. Since $\delta(X_{-i})$ is infinite, there exists $\bar{a}_i \in A_i$ such that $\beta_i^{A_i \times X_{-i}}(a_i)$ is infinite. □

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 $A = \prod_{i \in N} A_i$. □

3.3 Existence

A natural question is whether semi-injective BRS exists in every finite game. Recall that an injective BRS is semi-injective; Brandenburger and Friedenberg (2008) proved in Proposition H.3 that for generic games, the set of correlated rationalizable action profiles, which is the largest BRS and is always non-empty, is an injective BRS. It's not clear if their method of proof can be extended to handle non-generic games. Moreover, the conventional argument to pass from generic existence of a solution to everywhere existence relies on the upper-hemi continuity of the solution correspondence (e.g., Cho (1987)); it's tricky to apply here, because the solution concept, injective BRS, is set-valued, so the solution correspondence is a set of sets for every game.

Nevertheless, when we think about injective BRS in terms of its W_i^1 (cf. Equation 5), there is a natural constructive proof of existence for injective BRS (thus for semi-injective BRS as well).

Proposition 3.5. *In every finite game, an injective BRS exists.*

The ideas of the construction are as follows: start out with the set of correlated rationalizable action profiles (call it $R^1 = \prod_{i \in N} R_i^1$), (1) find its $W_i^1(R^1)$ for each player i (cf. Equation 5). And (2) delete from R^1 (in an arbitrary manner) actions in $W_i^1(R^1)$ so that every action remaining in $W_i^1(R^1)$ has a distinct supporting belief. Then, (3) find the largest BRS contained in the remaining actions of R^1 (call it R^2) and go back to (1) with R^1 being replaced by R^2, and so on.

Recall that a BRS Q is injective if and only if each action in $W_i^1(Q)$ has a distinct (unique) supporting belief. The deletions in step (2) force this condition. But the remaining actions might not be a BRS, so we need to delete further actions so that it is a BRS; this is step (3). Now the BRS might not be injective, so we go back to step (1).

Clearly, step (2) always gives a non-empty set (some actions in W_i^1 will survive the deletions). Furthermore, it can be seen that the largest BRS given in step (3) is always non-empty as well. Therefore, we have a sequence of decreasing and non-empty sets; it then follows that their intersection, the limit of the sequence, is a (non-empty) injective BRS.

The details of the proof can be found in Appendix A.

4 Intrinsic Correlated Equilibrium

4.1 Characterization

We now turn our attention to characterizing the set of intrinsic correlated equilibrium distributions (CED).

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 CED 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 (4), is the set of beliefs supporting a_i in Q_{-i}.

For each $i \in N$, define

$$
\begin{aligned}
Y_i^1 &= \{a_i \in Q_i : \ \mu(\cdot|a_i) \text{ is an extreme point of } \beta_i^Q(a_i)\}, \qquad (7)\\
Y_i^l &= \{a_i \in Y_i^1 : \mu(Y_{-i}^{l-1}|a_i) = 1\}, l \geq 2,\\
Y_i &= \bigcap_{l \geq 1} Y_i^l,
\end{aligned}
$$

Notice that Y_i^l's are defined with respect to a fixed μ. We write $Y_i(\mu)$ and $Y_i^l(\mu)$ when it is necessary to emphasize the dependence on μ.

We now state the main result of Section 4.

Theorem 2. *A CED $\mu \in \Delta(A)$ is intrinsic 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 1 for intrinsic a posteriori equilibrium. To see why Y_i^1 makes reference to extreme points of $\beta_i^Q(a_i)$, we sketch the proof of the following Lemma 4.1, which is the analogue of Lemma 3.1. In essence, posterior $\mu(\cdot|a_i)$ that is a non-extreme point of $\beta_i^Q(a_i)$ can be "split" into two distinct beliefs in $\beta_i^Q(a_i)$, so action a_i can be supported by these two beliefs; this is analogous to the a posteriori equilibrium case where when $\beta_i^Q(a_i)$ is not a singleton, a_i can be supported by any two distinct beliefs in $\beta_i^Q(a_i)$. The difference is of course that for correlated equilibrium beliefs must come from a common prior, so not any two beliefs in $\beta_i^Q(a_i)$ can support a_i—they must be two beliefs whose convex combination is equal to $\mu(\cdot|a_i)$.

Lemma 4.1. *Fix an intrinsic correlated equilibrium (λ, T, σ), and suppose that $\mu \in \Delta(A)$ 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 optimality condition for correlated equilibrium (condition (1)) 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 3.1 and does not use common prior. □

The proof the only if of Theorem 2 then follows from the above lemma exactly as the proof of the only if in Theorem 1 follows from Lemma 3.1; it also does not use common prior.

For the if direction of Theorem 2, we also follow the strategy of proof for Theorem 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 CED 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 1, we have the following finite-level version of Theorem 2.

Theorem 2 (Finite-level version). *Fix a $l \geq 1$ and a CED $\mu \in \Delta(A)$. If for every player i,*

$$a_i, a_i' \in Y_i^l, a_i \neq a_i' \Longrightarrow \mu(\cdot|a_i) \neq \mu(\cdot|a_i'), \tag{8}$$

then there exists an a correlated equilibrium (λ, T, σ) that obtains μ in which players condition their actions on their l-th order beliefs (i.e., $\delta_i^l(t_i) = \delta_i^l(t_i') \Rightarrow \sigma_i(t_i) = \sigma_i(t_i')$).

Conversely, if players condition their actions on their l-th order beliefs in a correlated equilibrium that obtains μ, then (8) holds for every player i.

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

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

$$\sum_{a_{-i} \in A_i} u_i(a_i, a_{-i}) \mu(a_i, a_{-i}) > \sum_{a_{-i} \in A_i} u_i(a_i', a_{-i}) \mu(a_i, a_{-i}), \tag{9}$$

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

Myerson (1997) calls μ's incentives *elementary* if (9) is satisfied for every pair of distinct a_i and a_i' in A_i.

Proposition 4.2. *A CED with strict incentives on the support is intrinsic.*

The proof of the proposition is as follows: if incentives of a CED 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 CED:*

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 CED.*

More generally in this game, a CED 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 CED that is not intrinsic.*

Note that $p = q = r = 1/5$ *and* $s = 2/5$ *is an intrinsic CED with* $Y_1^1 = Y_2^1 = \{A\}$*; on the other hand,* $Y_i^2 = Y_i = \emptyset$ *for both* i*. Thus, it is an example where* Y_i^2 *makes a difference.*

Therefore, the set of intrinsic CED's in this game consists of all correlated equilibrium distribution 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 CED; the same reasoning from the previous example applies. Proposition 4.6 shows that this is a general phenomenon in two-person games: a non-degenerate mixed Nash equilibrium cannot be an intrinsic correlated equilibrium in generic two-person games.*

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

Here is a direct argument for the non-existence. For the sake of contradiction suppose that $(1/2A + 1/2B, 1/2A + 1/2B)$ *can be obtained from an intrinsic correlated equilibrium. Then, at every type in the equilibrium each player must believe that he gets the minmax* *value of the game, which is 0; this is a general property of correlated equilibrium in zero-sum games and is proved in Aumann (1974). Then, every type of player 1 must have first order belief* $1/2\ A + 1/2\ B$*, because this is the unique belief that gives an expected payoff 0 (given either A or B played by player 1). Likewise, every type of player 2 must have first order belief* $1/2\ A + 1/2\ B$*. Then, a simple induction argument shows that every type of player 1 (respectively, of player 2) have the same k-th order belief, for any* $k \geq 1$*. Therefore, there is a unique hierarchy of belief for player 1. By the "intrinsicness" requirement, player 1 can only play one action in the unique hierarchy of belief, but two actions are played in the equilibrium, thus a contradiction.*

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 CED:*

$Y_1^1 = \{A, B, C\}$ *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 CED 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* $(1/4A + 1/4B + 1/2C, 1/2A + 1/2B)$*:*

$T_1 = \{A(1), A(2), B(1), B(2), C\}, T_2 = \{A(1), A(2), B\}$, $\sigma_1(A(1)) = \sigma_1(A(2)) = \sigma_2(A(1)) = \sigma_2(A(2)) = A$, $\sigma_1(B(1)) = \sigma_1(B(2)) = \sigma_2(B) = B$, $\sigma_1(C) = C$, *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 CED 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) CED 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 CED is not intrinsic. One can check that it is an extreme point in the set of CED's (see Proposition 4.3).*

4.2 Geometric Properties

It turns out that intrinsic correlated distributions have nice geometric structure. The following proposition (Proposition 4.3) shows that intrinsic correlated equilibrium is related to a notion of irreducibility (analogous to that of Markov chains) and to extreme point in the set of CED's. As a by-product, it shows in a precise sense that "most" of the CED's are intrinsic. In addition, the geometry of intrinsic CED's enables us to prove that for a generic two-person game, any non-degenerate mixed-strategy Nash equilibrium is not an intrinsic CED (Proposition 4.6).

For a fixed CED $\mu \in \Delta(A)$, with $Q_i = \operatorname{supp}(\operatorname{marg}_{A_i} \mu)$ 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 CED 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 CED, and μ can be written uniquely as convex combination of μ^k's. We say that μ^k is an *irreducible sub-distribution* of μ.

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

As a concrete illustration, the CED below (where $\{A, B, C, D\}$ are actions for each of the two players) has three irreducible sub-distributions: 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 4.3. *Suppose that a correlated equilibrium distribution μ has irreducible sub-distributions 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 set of CED's.*

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, given the iterated construction of Y_i^l in Equation (7).

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

Suppose μ^1 and μ^2 are two CED's such that $\mu = \mu^1/2 + \mu^2/2$ and $\text{supp}\,\mu^1 = \text{supp}\,\mu^2 = \text{supp}\,\mu$. 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 $\mu^1(\cdot|a_i) = \mu^2(\cdot|a_i)$ 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 CED's. □

Corollary 4.4. *An irreducible and non-extreme CED is intrinsic.*

Any CED is a natural sum of irreducible CED's that are its sub-distributions, just like any graph is a natural sum of sub-graphs that are its connected components. Whenever analyzing an CED in the interim stage (i.e., when types are realized), it's without loss of generality to restrict to irreducible CED. Furthermore, by Point 3 of Proposition 4.3, to check if a CED is intrinsic, it is necessary and sufficient to check if each of its irreducible sub-distributions is intrinsic.

When restricting to irreducible CED, Corollary 4.4 tells us that "most" of the CED's are intrinsic, in the sense that "most" of the points in a convex set are not extreme points. For any finite game, the number of irreducible and non-intrinsic CED's is finite, because the set of CED's is a polytope, so it has a finite number of extreme points; and in general the number of irreducible CED's is infinite.

Moreover, if two distinct irreducible CED's μ^1 and μ^2 (not necessarily themselves intrinsic) place positive probability on a same action (i.e., 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 $\gamma \in (0, 1)$.

This suggests that intrinsic CED's have convexity property. This is indeed the case; the proof can be found in Appendix C.

Proposition 4.5. *The set of intrinsic CED's is convex.*

Although it is convex, the set of intrinsic CED's is not necessarily closed; see Example 4.1. This should not be surprising, since intrinsic CED is analogous to CED with strict incentive inequalities (cf. Proposition 4.2).

Finally, the geometry of intrinsic CED's enables us to show that for a generic two-person game, any non-degenerate mixed-strategy Nash equilibrium is not an intrinsic CED. Geometrically, this is because for generic two-person games, Nash equilibria are extreme points in the set of CED's. This, however, is not enough, because being an extreme point is only a necessarily condition for being non-intrinsic, when the CED is irreducible (clearly a Nash equilibrium is irreducible).

Conceptually, 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, the "intrinsicness" requires the presence of different hierarchies of beliefs to purify the mixed strategy — the source of mixing is the belief hierarchies. Thus, non-degenerate mixed Nash equilibrium cannot be intrinsic.

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 $\operatorname{BR}_j(x) = \{a_j \in A_j : u_j(a_j, x) \geq u_j(a'_j, x) \text{ for all } a'_j \in A_j\}$. This is a well-known genericity class of two-person games, and a good reference on it is von Stengel (2002).

Harsanyi (1973) proved that for these generic games, any mixed-strategy Nash equilibrium can be purified via vanishing payoff shocks that are independent across players. As mentioned in the introduction, we interpret this difference as showing that to purify mixed-strategy Nash equilibrium one must go "outside" of the game, either by introducing private randomization or via vanishing payoff shocks.

Proposition 4.6. *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 CED.*

Proof of Proposition 4.6. Since (x, y) is a Nash equilibrium, we have $\text{supp}(x) \subseteq BR_1(y)$ and $\text{supp}(y) \subseteq BR_2(x)$. Thus, $|\,\text{supp}(x)| + |\,\text{supp}(y)| \leq |\,\text{BR}_1(y)| + |\,\text{BR}_2(x)|$. By the genericity of the game, we have $|\,\text{supp}(x)| = |BR_2(x)|$ and $|\,\text{supp}(y)| = |BR_1(y)|$.

Theorem 2.10 of von Stengel (2002) (which again uses the genericity condition) implies that the convex set $C = \{z \in \Delta(A_1) : \text{supp}(z) = \text{supp}(x) \text{ and } \text{BR}_2(z) = \text{BR}_2(x)\}$ is of dimension 0, i.e. $C = \{x\}$ [3]. Fix any $a_2 \in \text{supp}(y)$, we claim that x is an extreme point of $\beta_2^A(a_2)$.

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 $\text{supp}(z_1) = \text{supp}(x) = \text{supp}(z_2)$. 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 $\text{BR}_2(x) \subseteq \text{BR}_2(z_1) \cap \text{BR}_2(z_2)$; this means that $BR_2(x) = \text{BR}_2(z_1) = \text{BR}_2(z_2)$, because $|\,\text{BR}_2(z_1)| \leq |\,\text{supp}(z_1)| = |\,\text{supp}(x)| = |BR_1(x)|$ 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 $a_1 \in \text{supp}(x)$. Our desired conclusion then follows from the characterization of intrinsic CED's in Theorem 2. □

5 Related Literature and Extension

5.1 Relation 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[4] $((\tilde{\lambda}_i)_{i \in N}, T)$, 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 (2). 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.

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

[4] 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 type $t_i \in T_i$ 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 formulation: $((\lambda_i)_{i\in N}, T, \sigma)$, 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 (2), and if condition $(*)$ 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 the type space $((\tilde{\lambda}_i)_{i\in N}, T)$.

Brandenburger and Friedenberg are interested in the set of actions that are consistent with epistemic conditions RCBR, CI and SUFF:

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

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 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 and independent paper by Peysakhovich (2009) provides another partial answer: if $\mu \in \Delta(A)$ is a correlated equilibrium distribution, 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 $i \in N$.

5.2 Private Randomization

We now relax our restriction to pure strategy.

As before that we have type space $((\lambda_i)_{i\in N}, T)$, where $\lambda_i : T_i \to \Delta(T_{-i})$ for each $i \in N$. Players are now allowed to use private randomization: $\sigma_i : T_i \to \Delta(A_i)$.

For common prior $\lambda \in \Delta(T)$, let $\lambda_i(t_i) = \lambda(\cdot|t_i)$ if $\lambda(t_i) > 0$.

The equilibrium condition (1) and Definition 2.1 of a posteriori and correlated equilibria still apply without change.

For every player i and type $t_i \in T_i$, the first order belief at t_i is given by $\delta_i^1(t_i) \in \Delta(A_{-i})$, where for every $a_{-i} \in A_{-i}$,

$$\delta_i^1(t_i)(a_{-i}) = \sum_{t_{-i} \in T_{-i}} \lambda_i(t_i)(t_{-i}) \prod_{j \neq i} \sigma_j(t_j)(a_j).$$

The tuple (λ, T, σ) *obtains* a distribution $\mu \in \Delta$, where for every $a \in A$,

$$\mu(a) = \sum_{t \in T} \lambda(t) \prod_{i \in N} \sigma_i(t_i)(a_i).$$

The following theorem is a reinterpretation of Peysakhovich (2009)'s main result; details of the proof can be found in his paper.

Theorem (Peysakhovich)**.** *For any CED $\mu \in \Delta(A)$, there exists a correlated equilibrium (λ, T, σ) that obtains μ such that players condition their actions only on their first order beliefs, i.e., $\delta_i^1(t_i) = \delta_i^1(t_i') \Rightarrow \sigma_i(t_i) = \sigma_i(t_i')$.*

Therefore, we have an interesting trade-off between mixed strategy and higher order beliefs. On the one hand, every CED can be obtained from a correlated equilibrium in which every player plays randomized actions contingent on his first-order belief. On the other hand, "most" CED's (cf. Proposition 4.3) can be obtained from correlated equilibria 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.

We leave the analogous result for a posteriori equilibrium to future works.

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 Proposition 3.5

We first formally specify the 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 (4)), and let $W_i^1(l-1)$ be the $W_i^1(R^{l-1})$, i.e., the W_i^1 obtained in Equation (5) when $Q = R^{l-1}$. And for each $i \in N$ and $\gamma \in \beta_i^{l-1}(W_i^1(l-1))$, fix an $a^{l-1}(\gamma) \in W_i^1(l-1)$ such that $\beta_i^{l-1}(a^{l-1}(\gamma)) = \gamma$; note that if β_i^{l-1} is injective on $W_i^1(l-1)$, there is a unique choice of $a^{l-1}(\gamma)$.

For each $i \in N$, let

$$\begin{aligned} R_i^{l,1} &= \left(R_i^{l-1} \setminus W_i^1(l-1)\right) \cup \{a^{l-1}(\gamma) : \gamma \in \beta_i^{l-1}(W_i^1(l-1))\}, \qquad (10)\\ R_i^{l,k} &= \{a_i \in R_i^{l,1} : \exists \mu \in \Delta(R_{-i}^{l,k-1}) \text{ s.t. } a_i \text{ is optimal under } \mu\}, k \geq 2,\\ R_i^l &= \bigcap_{k \geq 1} R_i^{l,k}. \end{aligned}$$

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 A.1. *$R = \prod_{i \in N} R_i$ is a non-empty, injective BRS. And by some choice of $a^{l-1}(\gamma)$ for each l and γ in (10), we can obtain any maximal (in the set-inclusion partial order) 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 $\gamma \in \beta_i^{l-1}(W_i^1(l-1))$.

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 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 $a_i' \neq a_i \in W_i^1(Q)$, we have $\beta_i^Q(a_i) \neq \beta_i^Q(a_i')$; and notice that $W_i^1(1) \cap Q_i \subseteq W_i^1(Q)$. Thus, so by some choices of $a^1(\gamma)$ in Equation (10), 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 (10), we have $Q_i \subseteq R_i$. But this means that $Q_i = R_i$ since Q is a maximal injective BRS. $\square$

B Proof of If in Theorem 2

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

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

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 $\text{marg}_{A_i} \mu$. 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 B.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 B.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 B.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 B.1); and $\lambda^{l-1}(\{c(a_i)(1), c(a_i)(2)\} \times \{a_i\} \times T_{-(i,j)}^{l-1}) = \mu(c(a_i), a_i) > 0$. 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 B.1 to λ^{l-1}, T^{l-1}, $(Y_i^{l-1}\setminus Y_i^l)_{i\in N}$ and $\{(\nu(a_i,1),\nu(a_i,2))\}_{a_i\in Y_i^{l-1}\setminus Y_i^l, i\in N}$ 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 $T = T^L$ ($T_i = T_i^L = \{a_i(k) : a_i \in Q_i\setminus Y_i, k \in \{1,2\}\}\cup Y_i$), $\lambda = \lambda^L$, 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 B.1. Without loss of generality suppose that $N = \{1, \ldots, n\}$.

Let $\mu^1 \in \Delta(\tilde{X}_1 \times \prod_{2\le i\le n} X_i)$ 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 $2 \le l \le n$, let $\mu^l \in \Delta(\prod_{1\le j\le l} \tilde{X}_j \times \prod_{l+1\le i\le n} X_i)$ be such that for every $x_l \in Z_l$, $(x_1, \ldots, x_{l-1}) \in \prod_{1\le i\le l-1} \tilde{X}_i$ and $(x_{l+1}, \ldots, x_n) \in \prod_{l+1\le i\le n} X_i$:

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

and

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

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

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

□

C Proofs for Section 4.2

Proof of Proposition 4.5. Suppose that $\mu^1,\mu^2\in\Delta(A)$ are two intrinsic CED's; 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 $Q_i^1=\operatorname{supp}(\operatorname{marg}_{A_i}\mu^1)$ and $Q_i^2=\operatorname{supp}(\operatorname{marg}_{A_i}\mu^2)$ 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 CED. $\square$

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] Cho, I. K. (1987) A Refinement of Sequential Equilibrium. Econometrica, Volume 55, No. 6, pp. 1367-1389.

[7] Harsanyi, J. C. (1973) Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. International Journal of Game Theory, Volume 2, No. 1, pp. 1-23.

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

[9] 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.

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

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

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

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

[14] 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.

[15] 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.