Weak Belief and Permissibility*

Emiliano Catonini[a] and Nicodemo De Vito[b]
[a]ICEF - National Research University Higher School of Economics, Russian Federation
[b]Department of Decision Sciences - Bocconi University

Abstract

We provide epistemic foundations for permissibility (Brandenburger, 1992), a strategic-form solution concept for finite games which coincides with the Dekel-Fudenberg procedure, i.e., the elimination of all weakly dominated strategies, followed by the iterated elimination of strictly dominated strategies. We show that permissibility characterizes the behavioral implications of "cautious rationality and common weak belief of cautious rationality" in the canonical, universal type structure for lexicographic beliefs. For arbitrary type structures, we show that the behavioral implications of these epistemic assumptions are characterized by the solution concept of full weak best response set, a weak dominance analogue of best response set (Pearce, 1984).

KEYWORDS: Pemissibility, Dekel-Fudenberg Procedure, Infinitely More Likely, Lexicographic Probability Systems, Rationality.

1 Introduction

Permissibility (Brandenburger, 1992) is a solution concept for finite games in strategic form, and it is based on an iterative procedure. The central feature of this procedure is that the beliefs of each player over his relevant space of uncertainty are represented by lexicographic probability systems (henceforth, LPS's); that is, each player has a finite sequence of probability measures on the set of co-players' strategies, and he uses them lexicographically to determine his preferences over his own strategies (see Blume et al., 1991). A strategy survives a step of the permissibility procedure if it is a lexicographic best reply to an LPS that satisfies two conditions: (i) each strategy of the co-players is assigned strictly positive probability by some component measure of the LPS; (ii) the first component measure of the LPS assigns probability one only to those strategies which have survived the previous steps of the procedure. Brandenburger (1992) proved that permissibility coincides with the Dekel-Fudenberg procedure (Dekel and Fudenberg,

*We thank three anonymous reviewers from the 13th Conference on Logic and the Foundations of Game and Decision Theory (LOFT 2018), Milan, 2018, for useful comments on a previous version of this paper. We also thank Pierpaolo Battigalli and the seminar participants at Bocconi University for their suggestions. Financial support from European Research Council (STRATEMOTIONS–GA 324219) is gratefully acknowledged. The study has been funded within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE) by the Russian Academic Excellence Project '5-100.

1990), i.e., the elimination of all weakly dominated strategies followed by the iterated elimination of strictly dominated strategies.

The aim of this paper is to provide epistemic foundations for permissibility. We base our analysis on two key notions: cautious rationality and weak belief. Cautious rationality is the combination of lexicographic expected utility maximization and a cautious attitude of the player towards the primitive, payoff-relevant uncertainty. Specifically, each payoff-relevant event is assigned strictly positive probability by some component measure of the player's LPS, i.e., it is considered possible under some theory of the world that the player entertains.

We say that a player weakly believes an event E if he deems E "infinitely more likely" than its complement. Loosely speaking, a player deems an event E infinitely more likely than F if he strictly prefers to bet on E rather than on F regardless of the size of the winning prizes for the two bets (given the same losing outcome). This preference-based notion of "infinitely more likely" is due to Lo (1999), and it is weaker than the one put forward by Blume et al. (1991). In particular, the first is monotone, whereas the second is not. Since the set of permissible strategies can be formally defined as the ouptut of an infinite iteration of a *monotone* operator between subsets of strategy profiles (see Definition 4 below), the "infinitely more likely" relation due to Lo (1999) is suitable for the epistemic analysis of permissibility. Indeed, as we will show, an event E is weakly believed if and only if it is assigned probability one by the first component measure of the LPS representing the preference relation.[1] By contrast, Brandenburger et al. (2008) adopt the (non-monotone) notion of "infinitely more likely than" of Blume et al. (1991) for their epistemic analysis of iterated admissibility, and they leave the epistemic foundations for permissibility as an open question (see Brandenburger et al., 2008, p. 333).

With this, we show that the permissibility set characterizes the behavioral implications of cautious rationality and common weak belief of cautious rationality in the canonical, universal type structure for LPS's. The canonical type structure represents all hierarchies of lexicographic beliefs on strategies, without imposing extraneous restrictions on players' beliefs. For arbitrary type structures, we show that the behavioral implications of cautious rationality and common weak belief of cautious rationality are characterized by the solution concept of *full weak best response set*, a weak dominance analogue of best response set—a concept, due to Pearce (1984), based on strict dominance.

The epistemic foundations of the Dekel-Fudenberg procedure have been studied also by Asheim and Dufwenberg (2003), Hu (2007) and Perea (2012). We elaborate on the comparison between our analysis and the aforementioned papers in Section 5. Here we just note that our notion of cautious rationality and a strengthening of weak belief allow a transparent comparison with a possible epistemic foundation for iterated admissibility—see Section 5 and Catonini and De Vito (2018).

The paper is structured as follows. Section 2 introduces LPS's and weak belief. Section 3 presents the solution concepts: permissibility and (full) weak best response sets. Section 4 introduces the formalism of type structures and carries on the epistemic analysis of the solution concepts. Section 5 discusses the relationship with other analyses of the Dekel-Fudenberg procedure in the literature, and with our epistemic analysis of iterated admissibility. Appendix A illustrates the preference-based foundation of weak belief.

[1] This LPS-based notion of weak belief is called "first-order knowledge" in Brandenburger (1992). It should be noted that Brandenburger did not provide a preference-based foundation of this concept.

Appendix B collects the proofs omitted from Section 3 and Appendix C those omitted from Section 4.

2 Lexicographic beliefs and weak belief

2.1 Lexicographic probability systems

Throughout, fix a Polish space X. The space X is endowed with the Borel σ-field, the elements of which are called *events*. We let $\mathcal{M}(X)$ denote the set of Borel probability measures on X, and we endow $\mathcal{M}(X)$ with the weak* topology, so that it becomes a Polish space. We endow the product of Polish spaces with the product topology and endow a subset of a Polish space with the subspace topology. We let $\mathcal{N}(X)$ (resp. $\mathcal{N}_n(X)$) denote the set of all finite (resp. length-n) sequences of Borel probability measures on X, that is,

$$\begin{aligned}\mathcal{N}(X) &= \cup_{n\in\mathbb{N}}\mathcal{N}_n(X)\\ &= \cup_{n\in\mathbb{N}}(\mathcal{M}(X))^n.\end{aligned}$$

Each $\overline{\mu}=(\mu^1,...,\mu^n)\in\mathcal{N}(X)$ is called **lexicographic probability system** (LPS). The set $\mathcal{N}(X)$ is endowed with the *direct sum topology*, so that $\mathcal{N}(X)$ is a Polish space.

For every Borel probability measure μ on a Polish space X, the support of μ, denoted by $\mathrm{Supp}\mu$, is the smallest closed set $C\subseteq X$ such that $\mu(C)=1$. The support of an LPS $\overline{\mu}=(\mu^1,...,\mu^n)\in\mathcal{N}(X)$ is thus defined as $\mathrm{Supp}\overline{\mu}=\cup_{l\leq n}\mathrm{Supp}\mu^l$. So, an LPS $\overline{\mu}=(\mu^1,...,\mu^n)\in\mathcal{N}(X)$ is of **full-support** if $\cup_{l\leq n}\mathrm{Supp}\mu^l=X$. We write $\mathcal{N}_n^+(X)$ for the set of all full-support, length-n LPS's and $\mathcal{N}^+(X)$ for the set of full-support LPS's.

Suppose we are given Polish spaces X and Y, and a Borel map $f:X\to Y$. The map $\widetilde{f}:\mathcal{M}(X)\to\mathcal{M}(Y)$, defined by

$$\widetilde{f}(\mu)(E)=\mu\left(f^{-1}(E)\right),$$

where $\mu\in\mathcal{M}(X)$ and $E\subseteq Y$ is an event, is called the image (or pushforward) measure map of f. For each $n\in\mathbb{N}$, the map $\widehat{f}_{(n)}:\mathcal{N}_n(X)\to\mathcal{N}_n(Y)$ is defined by

$$(\mu^1,...,\mu^n)\mapsto\widehat{f}_{(n)}\left((\mu^1,...,\mu^n)\right)=\left(\widetilde{f}\left(\mu^k\right)\right)_{k\leq n}.$$

Thus the map $\widehat{f}:\mathcal{N}(X)\to\mathcal{N}(Y)$ defined by

$$\widehat{f}(\overline{\mu})=\widehat{f}_{(n)}(\overline{\mu}),\ \overline{\mu}\in\mathcal{N}_n(X),$$

is called the **image LPS map of** f, and it is Borel measurable.[2]

Furthermore, given Polish spaces X and Y, we let Proj_X denote the canonical projection from $X\times Y$ onto X. Define the marginal measure of $\mu\in\mathcal{M}(X\times Y)$ on X as $\mathrm{marg}_X\mu=\widetilde{\mathrm{Proj}_X}(\mu)$. Consequently, the marginal of $\overline{\mu}\in\mathcal{N}(X\times Y)$ on X is defined by $\overline{\mathrm{marg}}_X\overline{\mu}=\widehat{\mathrm{Proj}_X}(\overline{\mu})$. Finally, we let Id_X denote the identity map on X, that is, $\mathrm{Id}_X(x)=x$ for all $x\in X$.

[2]For details and proofs related to Borel measurability and continuity of the involved maps, the reader can consult Catonini and De Vito (2016).

2.2 Weak belief

The notion of **weak belief** captures the idea that an event (i.e., a Borel set) $E \subseteq X$ is "infinitely more likely" than its complement.

Following Lo (1999), we say that a player deems event E infinitely more likely than event F if she prefers to bet on E rather than on F no matter the prizes for the two bets. We formalize this preference-based notion in Appendix A, where we introduce the appropriate language. Here, we provide the equivalent definition of "infinitely more likely" in terms of the LPS that represents the player's preferences. This equivalence is formally proved in Appendix A.

Given an LPS $\overline{\mu} = (\mu^1, ..., \mu^n) \in \mathcal{N}(X)$ and an event $E \subseteq X$, let

$$\mathcal{I}_{\overline{\mu}}(E) = \inf \left\{ l \in \{1, ..., n\} \,\middle|\, \mu^l(E) > 0 \right\}.$$

Definition 1 *Fix two disjoint events $E, F \subseteq X$ with $E \neq \emptyset$. Say that E is* ***infinitely more likely than*** *F under $\overline{\mu}$ if $\mathcal{I}_{\overline{\mu}}(E) < \mathcal{I}_{\overline{\mu}}(F)$.*

It is straightforward to see that "infinitely more likely" is monotone.

Remark 1 *If E is infinitely more likely than F under $\overline{\mu}$ and G is an event such that $E \subseteq G$, then G is infinitely more likely than F under $\overline{\mu}$.*

Now we provide the LPS-based notion of weak belief.

Definition 2 *Fix a Polish space X and a non-empty event $E \subseteq X$. We say that E is* ***weakly believed under*** *$\overline{\mu} = (\mu^1, ..., \mu^n) \in \mathcal{N}(X)$ if $\mu^1(E) = 1$.*

As anticipated, weak belief of E corresponds to the desired infinitely more likely relation between E and its complement.

Proposition 1 *A non-empty event $E \subseteq X$ is weakly believed under $\overline{\mu}$ if and only if E is infinitely more likely than $X \backslash E$ under $\overline{\mu}$.*

Proof: Suppose that $E \subseteq X$ is weakly believed under $\overline{\mu} = (\mu^1, ..., \mu^n)$. Thus $\mu^1(E) = 1$. This means that $\mathcal{I}_{\overline{\mu}}(E) = 1 < \mathcal{I}_{\overline{\mu}}(X \backslash E)$. Conversely, suppose that E is infinitely more likely than $X \backslash E$ under $\overline{\mu}$. Thus $\mathcal{I}_{\overline{\mu}}(E) < \mathcal{I}_{\overline{\mu}}(X \backslash E)$. Then $\mathcal{I}_{\overline{\mu}}(X \backslash E) > 1$, which implies $\mu^1(X \backslash E) = 0$, and so $\mu^1(E) = 1$. ■

Weak belief satisfies the following properties that will be useful in our analysis.

Property 1: (Marginalization) If E is a non-empty event in $X \times Y$ which is weakly believed under $\overline{\mu} \in \mathcal{N}(X \times Y)$ and $\text{Proj}_X(E)$ is Borel, then $\text{Proj}_X(E)$ is weakly believed under $\overline{\text{marg}}_X \overline{\mu}$.

Property 2: (Conjunction and Disjunction) Fix non-empty events $E_1, E_2, ...$ in X. Suppose that, for each k, E_k is weakly believed under $\overline{\mu} \in \mathcal{N}(X)$. Thus $\cap_k E_k$ and $\cup_k E_k$ are weakly believed under $\overline{\mu}$.

Property 3: (Monotonicity) Fix a non-empty event $E \subseteq X$ which is weakly believed under $\overline{\mu} \in \mathcal{N}(X)$. If event $F \subseteq X$ is such that $E \subseteq F$, then F is weakly believed under $\overline{\mu}$.

A stronger notion than weak belief is that of **certain belief**: A non-empty event $E \subseteq X$ is certainly believed under $\overline{\mu} = (\mu^1, ..., \mu^n) \in \mathcal{N}(X)$ if $\mu^l(E) = 1$ for each $l = 1, ..., n$. In other words, E is certainly believed under $\overline{\mu}$ if its complement, viz. $X \backslash E$, is deemed subjectively impossible by the decision maker.[3] By constrast, weak belief in E does not rule out the possibility that $X \backslash E$ occurs, i.e., $\mu^l(X \backslash E) > 0$ for some $l > 1$. Certain belief also satisfies Properties 1-3 above, and it concides with weak belief when preferences conform to subjective expected-utility theory.

3 Permissibility and Weak Best Response Sets

Throughout, we consider finite two-player games. A **finite two-player game** is a structure $G = \langle I, (S_i, \pi_i)_{i \in I} \rangle$, where I is a two-player set and, for every $i \in I$, S_i is the (finite) set of strategies with $|S_i| \geq 2$ and $\pi_i : S \to \mathbb{R}$ is the payoff function.[4] Each strategy set S_i is given the obvious topology, i.e., the discrete topology. For notational convenience, given a mixed strategy profile $\sigma \in \prod_{i \in I} \mathcal{M}(S_i)$, we will denote player i's expected utility simply by $\pi_i(\sigma_i, \sigma_{-i})$, i.e., $\pi_i(\sigma_i, \sigma_{-i}) = \sum_{(s_i, s_{-i}) \in S_i \times S_{-i}} \sigma_i(s_i) \sigma_{-i}(s_{-i}) \pi_i(s_i, s_{-i})$. Similarly, given a pure strategy $s_i \in S_i$ and a probability measure $\mu_i \in \mathcal{M}(S_i)$, we will denote player i's expected utility by $\pi_i(s_i, \mu_i) = \sum_{s_{-i} \in S_{-i}} \pi_i(s_i, s_{-i}) \mu_i(s_{-i})$.[5] With an abuse of notation, we will also identify the pure strategy $s_i \in S_i$ with the mixed strategy $\sigma_i \in \mathcal{M}(S_i)$ such that $\sigma_i(s_i) = 1$.

For any two vectors $x = (x_l)_{l=1}^n, y = (y_l)_{l=1}^n \in \mathbb{R}^n$, we write $x \geq_L y$ if either (1) $x_l = y_l$ for every $l \leq n$, or (2) there exists $m \leq n$ such that $x_m > y_m$ and $x_l = y_l$ for every $l < m$.

In the remainder of this section, we fix a finite two-player game $G = \langle I, (S_i, \pi_i)_{i \in I} \rangle$.

Definition 3 *A strategy $s_i \in S_i$ is **optimal under** $\overline{\mu}_i = (\mu_i^1, ..., \mu_i^n) \in \mathcal{N}(S_{-i})$ if*

$$\left(\pi_i(s_i, \mu_i^l)\right)_{l=1}^n \geq_L \left(\pi_i(s_i', \mu_i^l)\right)_{l=1}^n, \ \forall s_i' \in S_i.$$

*We say that s_i is a **lexicographic best reply to** $\overline{\mu}_i$ if it is optimal under $\overline{\mu}_i$.*

Fix a player $i \in I$, and a set $Q_{-i} \subseteq S_{-i}$. We let $r_i(\overline{\mu}_i)$ denote the set of player i's strategies which are optimal under $\overline{\mu}_i \in \mathcal{N}(S_{-i})$, while $\mathcal{W}^+(Q_{-i})$ denotes the set of all full-support LPS's $\overline{\mu}_i$ such that $\mu_i^1(Q_{-i}) = 1$. Note that $\mathcal{W}^+(S_{-i}) = \mathcal{N}^+(S_{-i})$.

[3] In the language of decision theory, $X \backslash E$ is a Savage-null event—see Appendix A.

[4] Our notation is standard. Given a list $X_1, ..., X_n$ of sets, we write $X = \prod_{l=1,...,n} X_l$ and $x = (x_1, ..., x_n) \in X$. Moreover, given a player i, we denote by $-i$ the co-players. Here, we restrict our set-up to the two-player case; the analysis can be trivially extended to more than two players.

[5] We abuse notation by writing $\sigma_i(s_i)$ (or $\mu_i(s_{-i})$) instead of $\sigma_i(\{s_i\})$ (or $\mu_i(\{s_{-i}\})$).

Let $\mathcal{Q}$ be the collection of all subsets of S with the cross-product form $Q = \prod_{i\in I} Q_i$, where $Q_i \subseteq S_i$ for every i. We now introduce an operator $\rho : \mathcal{Q} \to \mathcal{Q}$ as follows. For each $Q \in \mathcal{Q}$, define the following sets:

$$\rho_i(Q_{-i}) = \left\{ s_i \in S_i : \exists \overline{\mu}_i \in \mathcal{W}^+(Q_{-i}), s_i \in r_i(\overline{\mu}_i) \right\},$$

$$\rho(Q) = \prod_{i\in I} \rho_i(Q_{-i}).$$

In words, $\rho(Q)$ is the set of strategies that are lexicographic best replies to some full support LPS under which Q_{-i} is weakly believed. Note that $\rho(\emptyset) = \emptyset$.

Remark 2 *The operator ρ is* monotone*: For every pair of subsets $E, F \in \mathcal{Q}$, if $E \subseteq F$ then $\rho(E) \subseteq \rho(F)$. This follows from the monotonicity property of weak belief (Property 3).*

We define the k-th iteration of ρ (the k-fold composition of ρ with itself) recursively as follows. For each $Q \in \mathcal{Q}$, define $\rho^0(Q) = Q$ for convenience; then for each $k \geq 1$,

$$\rho^k(Q) = \rho(\rho^{k-1}(Q)).$$

Note that, by the monotonicity of ρ, the sequence of subsets $(\rho^k(S))_{k=1}^{\infty}$ is weakly decreasing, i.e., $\rho^{k+1}(S) \subseteq \rho^k(S)$ for each $k \geq 1$. Therefore define

$$\rho^{\infty}(S) = \bigcap_{k\geq 1} \rho^k(S).$$

Since each strategy set S_i is finite, there exists $M \in \mathbb{N}$ such that $\rho^{\infty}(S) = \rho^M(S) \neq \emptyset$.

Definition 4 (Brandenburger, 1992) *A strategy profile $s \in S$ is* ***permissible*** *if $s \in \rho^{\infty}(S)$.*

It is possible to provide a characterization of the set $\rho^{\infty}(S)$ in terms of dominated strategies. The following definitions are standard.

Definition 5 *Fix a set $Q \in \mathcal{Q}$. A strategy $s_i \in S_i$ is* ***weakly dominated with respect to*** *Q if there exists a mixed strategy $\sigma_i \in \mathcal{M}(S_i)$ with $\sigma_i(Q_i) = 1$ such that $\pi_i(\sigma_i, s_{-i}) \geq \pi_i(s_i, s_{-i})$ for every $s_{-i} \in Q_{-i}$ and $\pi_i(\sigma_i, s'_{-i}) > \pi_i(s_i, s'_{-i})$ for some $s'_{-i} \in Q_{-i}$. If strategy $s_i \in S_i$ is weakly dominated with respect to S, we simply say that s_i is* ***weakly dominated****.*

Definition 6 *Fix a set $Q \in \mathcal{Q}$. A strategy $s_i \in S_i$ is* ***strictly dominated with respect to*** *Q if there exists a mixed strategy $\sigma_i \in \mathcal{M}(S_i)$ such that $\pi_i(\sigma_i, s'_{-i}) > \pi_i(s_i, s'_{-i})$ for every $s'_{-i} \in Q_{-i}$. If strategy $s_i \in S_i$ is strictly dominated with respect to S, we simply say that s_i is* ***strictly dominated****.*

We write $\mathrm{NWD}_i(Q)$ (resp., $\mathrm{ND}_i(Q)$) for the set of player i's strategies which are not weakly (resp., strictly) dominated with respect to $Q \in \mathcal{Q}$. We also write $\mathrm{NWD}(Q) = \prod_{i \in I} \mathrm{NWD}_i(Q)$ and $\mathrm{ND}(Q) = \prod_{i \in I} \mathrm{ND}_i(Q)$.

As we did for the operator ρ, we now define the k-th iteration of the operator $\mathrm{ND} : \mathcal{Q} \to \mathcal{Q}$ in a recursive way. For each $Q \in \mathcal{Q}$, define $\mathrm{ND}^0(Q) = Q$; then, for each $k \geq 1$, let

$$\mathrm{ND}^k(Q) = \mathrm{ND}(\mathrm{ND}^{k-1}(Q)).$$

The **Dekel-Fudenberg procedure** (Dekel and Fudenberg, 1990) is an iterative procedure in which one round of elimination of all weakly dominated strategies is followed by iterated elimination of strictly dominated strategies. The set of strategies surviving the Dekel-Fudenberg procedure is defined as

$$\mathrm{ND}^\infty(\mathrm{NWD}(S)) = \bigcap_{k \geq 1} \mathrm{ND}^{k-1}(\mathrm{NWD}(S)).$$

By finiteness of each strategy set S_i, it follows that $\mathrm{ND}^\infty(\mathrm{NWD}(S)) \neq \emptyset$.

The following result is due to Brandenburger (1992).

Proposition 2 *For each $k \geq 1$, it holds that*

$$\rho^k(S) = \mathrm{ND}^{k-1}(\mathrm{NWD}(S)).$$

Therefore, a strategy profile is permissible if and only if it survives the Dekel-Fudenberg procedure.

It is also possible to provide an alternative characterization of permissible strategies in terms of an analogue of best response set (Pearce, 1984).

Definition 7 *Fix a set $Q \in \mathcal{Q}$.*

*(i) Q is a **weak best response set (WBRS)** if for each $i \in I$ and each $s_i \in Q_i$, there exists $\overline{\mu}_i \in \mathcal{W}^+(Q_{-i})$ such that $s_i \in r_i(\overline{\mu}_i)$.*

*(ii) Q is a **full WBRS** if for each $i \in I$ and each $s_i \in Q_i$, there exists $\overline{\mu}_i \in \mathcal{W}^+(Q_{-i})$ such that $s_i \in r_i(\overline{\mu}_i)$ and $r_i(\overline{\mu}_i) \subseteq Q_i$.*

In words, a WBRS is a collection of strategy profiles with the property that every strategy of every player is a lexicographic best reply to some full-support LPS under which the set of the opponent's strategies is weakly believed. A WBRS is full if, in addition, all lexicographic best replies to each such LPS also belong to the set.

Proposition 3 *The set $\rho^\infty(S)$ is the unique full WBRS such that $Q \subseteq \rho^\infty(S)$ for every WBRS $Q \in \mathcal{Q}$.*

Corollary 1 *A strategy profile $s \in S$ is permissible if and only if $s \in Q$ for some WBRS $Q \in \mathcal{Q}$.*

WBRS's can be characterized in terms of dominated strategies (cf. Brandenburger et al., 2008). To this end, we need an additional definition. Say that a strategy $s_i' \in S_i$ **supports** $s_i \in S_i$ if there exists a mixed strategy σ_i with $s_i' \in \text{Supp}\sigma_i$ and $\pi_i(\sigma_i, s_{-i}) = \pi_i(s_i, s_{-i})$ for all $s_{-i} \in S_{-i}$. We let $\text{su}(s_i)$ denote the set of all $s_i' \in S_i$ that support s_i. So, the set $\text{su}(s_i)$ consists of all strategies for player i that are part of some convex combination equivalent to s_i.

Proposition 4 *A set $Q \in \mathcal{Q}$ is a WBRS if and only if, for each $i \in I$,*
(i) each $s_i \in Q_i$ is not weakly dominated, and
(ii) each $s_i \in Q_i$ is not strictly dominated with respect to $S_i \times Q_{-i}$.
A set $Q \in \mathcal{Q}$ is a full WBRS if and only if it is a WBRS such that, for each $i \in I$,
(iii) $\text{su}(s_i) \subseteq Q_i$ for every $s_i \in Q_i$.

4 Epistemic analysis

This section provides an epistemic foundation of permissible strategies and full weak best response sets in finite games. We consider epistemic type structures where types map to LPS's. We formalize (lexicographic) rationality, cautious rationality, weak belief of cautious rationality, etc., as events in a type structure. We show that full WBRS's capture the behavioral implications of cautious rationality and common weak belief of cautious rationality across all type structures. Moreover, permissibility captures the behavioral implications of cautious rationality and common weak belief of cautious rationality in the so-called universal type structure.

4.1 Lexicographic type structures

Hierarchies of beliefs are an essential element of epistemic analysis, as they will be used to formally define the notion of cautious rationality and common weak belief in cautious rationality. A hierarchy of beliefs specifies a player's belief over the space of primitive uncertainty (e.g., the set S_{-i} of strategies of player i's opponent), his belief over the opponent's beliefs, and so on. We adopt the formalism of type structures to model belief hiearchies.

In the remainder of this subsection, fix a finite two-player game $G = \langle I, (S_i, \pi_i)_{i\in I}\rangle$.

Definition 8 *An $(S_i)_{i\in I}$-**based lexicographic type structure** is a structure $\mathcal{T} = \langle S_i, T_i, \beta_i\rangle_{i\in I}$, where*

1. *for each $i \in I$, T_i is a Polish space;*

2. for each $i \in I$, the function $\beta_i : T_i \to \mathcal{N}(S_{-i} \times T_{-i})$ is measurable.

*We call each space T_i **type space** and we call each β_i **belief map**. Members of type spaces, viz. $t_i \in T_i$, are called **types**. Each element $(s_i, t_i)_{i \in I} \in S \times T$ is called **state (of the world)**.*

Definition 8 is a natural generalization of the standard definition of type structures with beliefs represented by probability measures, i.e., length-1 LPS (cf. Heifetz and Samet, 1998). The formalism of lexicographic type structures was first introduced by Brandenburger et al. (2008),[6] and later analyzed in Catonini and De Vito (2016, 2018). In what follows, we will omit the qualifier "lexicographic," and simply speak of type structures.

Type structures generate a collection of hierarchies of beliefs for each player. For instance, type t_i's first-order belief is an LPS on S_{-i}, and is given by $\overline{\mathrm{marg}}_{S_{-i}} \beta_i(t_i)$. A standard inductive procedure (see Catonini and De Vito, 2016, for details) shows how it is possible to provide an explicit description of a hierarchy induced by a type. This raises the question as to whether there exists a type structure which generates all possible hierarchies of beliefs, so that every type structure can be mapped into it in a unique belief-preserving way. In Catonini and De Vito (2016), we show that such type structure can be constructed by taking the set of types to be the collection of all possible hierarchies of beliefs that satisfy a coherence condition; this is the so called *canonical type structure for LPS's*, and every type structure can be viewed as a "sub-structure" of it.[7] The details of the construction are not relevant for the statements and proofs of our results. Instead, we will make use of two properties of the canonical type structure, namely belief-completeness and universality, as we review next.

Definition 9 *Let $\mathcal{T} = \langle S_i, T_i, \beta_i \rangle_{i \in I}$ and $\mathcal{T}' = \langle S_i, T'_i, \beta'_i \rangle_{i \in I}$ be two $(S_i)_{i \in I}$-based type structures. For each $i \in I$, let $\varphi_i : T_i \to T'_i$ be a measurable map such that*

$$\beta'_i \circ \varphi_i = \left(\widehat{\mathrm{Id}_{S_{-i}}, \varphi_{-i}}\right) \circ \beta_i. \tag{4.1}$$

*Then the function $(\varphi_i)_{i \in I} : T \to T'$ is called **type morphism (from $\mathcal{T}$ to $\mathcal{T}'$)**.*

The notion of type morphism captures the idea that a type structure $\mathcal{T}$ is "contained in" another type structure $\mathcal{T}'$ if $\mathcal{T}$ can be mapped into $\mathcal{T}'$ in a way that preserves the beliefs associated with types. Condition (4.1) in the definition of type morphism expresses consistency between the function $\varphi_i : T_i \to T'_i$ and the induced function $\left(\widehat{\mathrm{Id}_{S_{-i}}, \varphi_{-i}}\right) : \mathcal{N}(S_{-i} \times T_{-i}) \to \mathcal{N}\left(S_{-i} \times T'_{-i}\right)$. That is, the following diagram commutes:

$$\begin{array}{ccc} T_i & \xrightarrow{\beta_i} & \mathcal{N}(S_{-i} \times T_{-i}) \\ \downarrow \varphi_i & & \downarrow (\widehat{Id_{S_{-i}}, \varphi_{-i}}) \\ T'_i & \xrightarrow{\beta'_i} & \mathcal{N}(S_{-i} \times T'_{-i}) \end{array}$$

[6]The definition of lexicographic type structure in Brandenburger et al. (2008) requires that each belief be represented by a *mutually singular* LPS, i.e., an LPS in which, roughly speaking, the supports of the component measures are disjoint.

[7]This is in line with analogous results on hierarchies of both ordinary probabilities and conditional probability systems (cf. Mertens and Zamir, 1985, and Battigalli and Siniscalchi, 1999).

The notion of type morphism does not make any reference to hierarchies of LPS's. But, as one should expect, the important property of type morphisms is that they preserve the explicit description of belief hierarchies (for details, see Catonini and De Vito, 2016).

Definition 10 *An $(S_i)_{i \in I}$-based type structure $\mathcal{T} = \langle S_i, T_i, \beta_i \rangle_{i \in I}$ is*

- ***belief-complete*** *if each belief map β_i is onto;*
- ***universal*** *if for every other $(S_i)_{i \in I}$-based type structure $\mathcal{T}' = \langle S_i, T_i', \beta_i' \rangle_{i \in I}$ there is a unique type morphism from $\mathcal{T}'$ to $\mathcal{T}$*

The canonical type structure for LPS's is an instance of a belief-complete and universal type structure. In particular, every isomorphic copy of the canonical type structure is universal and belief-complete.[8] So, universality implies belief-completeness. The converse is not true: There exist belief-complete type structures which are not necessarily universal—see Friedenberg and Keisler (2011) and Catonini and De Vito (2018) for examples.

4.2 Cautious rationality and common weak belief of cautious rationality

We now formalize the epistemic conditions of interest as restrictions on strategy-type pairs in a type structure associated with the game. Fix a finite two-player game $G = \langle I, (S_i, \pi_i)_{i \in I} \rangle$ and a type structure $\mathcal{T} = \langle S_i, T_i, \beta_i \rangle_{i \in I}$.

Definition 11 *Fix a player $i \in I$. A type $t_i \in T_i$ is* ***cautious (in $\mathcal{T}$)*** *if $\overline{\mathrm{marg}}_{S_{-i}} \beta_i(t_i) \in \mathcal{N}^+(S_{-i})$.*

In words, this notion of "cautiousness" requires that the first-order belief of a type be a full-support LPS. It is therefore a condition that can be expressed in terms of primitives of the model (i.e., hierarchies of beliefs), and it requires that every payoff-relevant event, viz. $\{s_{-i}\} \times T_{-i}$, be assigned strictly positive probability by at least one of the component measures of the LPS $\beta_i(t_i)$.[9] We let C_i denote the set of all cautious types for each player $i \in I$. We show in Appendix C that this and all the other sets we will refer to are Borel.

For strategy-type pairs we define the following notions.

[8] Alternatively put, the canonical type structure for LPS's is the unique universal type structure up to type isomorphism (i.e., a bijective type morphism whose inverse is measurable).

[9] Cautiousness is satisfied by full-support types, i.e., types $t_i \in T_i$ for which $\beta_i(t_i) \in \mathcal{N}^+(S_{-i} \times T_{-i})$. However, as illustrated in Catonini and De Vito (2018), the notion of full-support for types crucially depends on topological features of type structures which are unrelated to belief hierarchies. By contrast, the weaker notion of cautiousness does not hinge on the topology on types, and it is sufficient to capture cautious behavior.

Definition 12 *Fix a strategy-type pair $(s_i, t_i) \in S_i \times T_i$.*

1. *Say (s_i, t_i) is **rational (in $\mathcal{T}$)** if s_i is optimal under $\overline{\text{marg}}_{S_{-i}}\beta_i(t_i)$.*
2. *Say (s_i, t_i) is **cautiously rational (in $\mathcal{T}$)** if it is rational and t_i is cautious.*

We let R_i be the set of all rational strategy-type pairs. Then, we let R_i^1 denote the set all **cautiously rational** strategy-type pairs $(s_i, t_i) \in S_i \times T_i$, that is, $R_i^1 = R_i \cap C_i$.

Say that type t_i weakly believes a non-empty event $E_{-i} \subseteq S_{-i} \times T_{-i}$ if E_{-i} is weakly believed under $\beta_i(t_i)$. For each player $i \in I$, let $\mathbf{WB}_i : 2^{S_{-i} \times T_{-i}} \to 2^{S_i \times T_i}$ be the operator defined by

$$\mathbf{WB}_i\left(E_{-i}\right) = \left\{(s_i, t_i) \in S_i \times T_i \,| t_i \text{ weakly believes } E_{-i}\right\},$$

for every event $E_{-i} \subseteq S_{-i} \times T_{-i}$. (If E_{-i} is not Borel, let $\mathbf{WB}_i\left(E_{-i}\right) = \emptyset$.).

For each $m > 1$, define R_i^m recursively by

$$R_i^{m+1} = R_i^m \cap \mathbf{WB}_i\left(R_{-i}^m\right).$$

We write $R_i^0 = S_i \times T_i$ and $R_i^\infty = \cap_{m \in \mathbb{N}} R_i^m$ for each $i \in I$. If $(s_i, t_i)_{i \in I} \in \prod_{i \in I} R_i^{m+1}$, we say that there is **cautious rationality and mth-order weak belief of cautious rationality ($\mathbf{R}^{\mathrm{c}}m\mathbf{WBR}^{\mathrm{c}}$)** at this state. If $(s_i, t_i)_{i \in I} \in \prod_{i \in I} R_i^\infty$, we say that there is **cautious rationality and common weak belief of cautious rationality ($\mathbf{R}^{\mathrm{c}}\mathbf{CWBR}^{\mathrm{c}}$)** at this state. Note that, for each $m > 1$,

$$R_i^{m+1} = R_i^1 \cap \left(\cap_{l \leq m} \mathbf{WB}_i\left(R_{-i}^l\right)\right),$$

and each R_i^m is Borel in $S_i \times T_i$ (see Appendix C).

We point out that $\mathrm{R^cCWBR^c}$ is preserved under type morphism between type structures. Specifically, suppose there is a type morphism $(\varphi_i)_{i \in I}$ from $\mathcal{T}$ to $\mathcal{T}^*$. So, if there is $\mathrm{R^cCWBR^c}$ at $(s_i, t_i)_{i \in I}$ in $\mathcal{T}$, then there is also $\mathrm{R^cCWBR^c}$ at $(s_i, \varphi_i(t_i))_{i \in I}$ in $\mathcal{T}^*$, provided that a measurability condition on $(\varphi_i)_{i \in I}$ is satisfied.

To state this formally, recall that a Borel map $f : X \to Y$ between separable metrizable spaces is **bimeasurable** if $f(E)$ is Borel in Y provided E is Borel in X. In particular, if X is a finite, discrete topological space, then f is bimeasurable, since its range is a finite (so Borel) subset of Y.

Lemma 1 *Let $\mathcal{T} = \langle S_i, T_i, \beta_i \rangle_{i \in I}$ and $\mathcal{T}^* = \langle S_i, T_i^*, \beta_i^* \rangle_{i \in I}$ be type structures such that there exists a bimeasurable type morphism $(\varphi_i)_{i \in I} : T \to T^*$ from $\mathcal{T}$ to $\mathcal{T}^*$. The following statements hold:*

(i) for each $m \geq 1$, if $(s_i, t_i)_{i \in I} \in \prod_{i \in I} R_i^m$, then $(s_i, \varphi_i(t_i))_{i \in I} \in \prod_{i \in I} R_i^{*,m}$;
(ii) if $(s_i, t_i)_{i \in I} \in \prod_{i \in I} R_i^\infty$, then $(s_i, \varphi_i(t_i))_{i \in I} \in \prod_{i \in I} R_i^{*,\infty}$.

We will make use of Lemma 1 in the proof of one of the main results, which are presented in the next section.

4.3 Main results

We now state and prove the main results of this paper. For notational convenience, for each $m = 0, 1, ...$, let $S^m = \rho^m(S)$ and, for each $i \in I$, $S_i^m = \text{Proj}_{S_i}\rho^m(S)$. Moreover, we let $S^\infty = \rho^\infty(S)$ and, for each $i \in I$, $S_i^\infty = \text{Proj}_{S_i}\rho^\infty(S)$.

Theorem 1 *Fix a finite two-player game $G = \langle I, (S_i, \pi_i)_{i \in I} \rangle$.*
(i) Let $\mathcal{T} = \langle S_i, T_i, \beta_i \rangle_{i \in I}$ be a type structure associated with G. Then $\prod_{i \in I} \text{Proj}_{S_i}(R_i^\infty)$ is a full WBRS.
(ii) If $Q \in \mathcal{Q}$ is a full WBRS, then there exists a type structure $\mathcal{T}^* = \langle S_i, T_i^*, \beta_i^* \rangle_{i \in I}$ such that $Q = \prod_{i \in I} \text{Proj}_{S_i}(R_i^{*,\infty})$.

Theorem 2 *Fix a finite two-player game $G = \langle I, (S_i, \pi_i)_{i \in I} \rangle$ and an associated belief-complete type structure $\mathcal{T} = \langle S_i, T_i, \beta_i \rangle_{i \in I}$. The following statements hold:*
(i) for each $m \geq 0$, $\prod_{i \in I} \text{Proj}_{S_i}(R_i^m) = \prod_{i \in I} S_i^m$;
(ii) if $\mathcal{T}$ is universal, then $\prod_{i \in I} R_i^\infty \neq \emptyset$ and $\prod_{i \in I} \text{Proj}_{S_i}(R_i^\infty) = \prod_{i \in I} S_i^\infty$.

The proofs of Theorem 1 and Theorem 2 will make use of the following results.

Lemma 2 *Fix a finite two-player game $G = \langle I, (S_i, \pi_i)_{i \in I} \rangle$ and an associated type structure $\mathcal{T} = \langle S_i, T_i, \beta_i \rangle_{i \in I}$. The following statements hold:*
(i) for each $m \geq 1$, $\prod_{i \in I} \text{Proj}_{S_i}(R_i^m) \subseteq \prod_{i \in I} S_i^m$;
(ii) $\prod_{i \in I} \text{Proj}_{S_i}(R_i^\infty) \subseteq \prod_{i \in I} S_i^\infty$.

Say that a type structure $\mathcal{T} = \langle S_i, T_i, \beta_i \rangle_{i \in I}$ is **finite** if each type space T_i is a finite set, and it is endowed with the discrete topology.

Lemma 3 *Fix a finite two-player game $G = \langle I, (S_i, \pi_i)_{i \in I} \rangle$ and a full WBRS $Q \in \mathcal{Q}$. There exists a finite type structure $\mathcal{T}^* = \langle S_i, T_i^*, \beta_i^* \rangle_{i \in I}$ such that $\prod_{i \in I} \text{Proj}_{S_i}(R_i^{*,\infty}) = \prod_{i \in I} Q_i$.*

Proof of Theorem 1: **Part (i)**: The conclusion is immediate if $\prod_{i \in I} \text{Proj}_{S_i}(R_i^\infty) = \emptyset$. So we will assume that this set is non-empty. Let $s_i \in \text{Proj}_{S_i}(R_i^\infty)$, so that $(s_i, t_i) \in R_i^\infty$ for some $t_i \in T_i$. Thus $(s_i, t_i) \in R_i^1$, so we have to show that $\text{Proj}_{S_{-i}}\left(R_{-i}^\infty\right)$ is weakly believed under $\overline{\text{marg}}_{S_{-i}}\beta_i(t_i)$. Since $R_{-i}^1, R_{-i}^2, ...$ are weakly believed under $\beta_i(t_i)$, it follows from the conjuction property of weak belief (Property 2) that R_{-i}^∞ is weakly believed under $\beta_i(t_i)$. Using the marginalization property (Property 1), $\text{Proj}_{S_{-i}}\left(R_{-i}^\infty\right)$ is weakly believed under $\overline{\text{marg}}_{S_{-i}}\beta_i(t_i)$, as required.

Suppose now that s_i' is optimal under $\overline{\text{marg}}_{S_{-i}}\beta_i(t_i)$. Then $(s_i', t_i) \in R_i^1$, and since $\text{Proj}_{S_{-i}}\left(R_{-i}^\infty\right)$ is weakly believed under $\overline{\text{marg}}_{S_{-i}}\beta_i(t_i)$ (as shown above), we have $(s_i', t_i) \in R_i^\infty$. This shows that $s_i' \in \text{Proj}_{S_i}(R_i^\infty)$, and concludes the proof of part (i).

Part (ii): Follows from Lemma 3. ■

Proof of Theorem 2: **Part (i)**: The statement is trivially true for $m = 0$.

Fix $m \geq 1$, and suppose that the statement has been shown to hold for $m-1$. We show that the statement is true for m.

Fix a player $i \in I$. Lemma 2.(i) gives that $\mathrm{Proj}_{S_i}(R_i^m) \subseteq S_i^m$. Conversely, let $s_i \in S_i^m$. So there is $\overline{\nu}_i = (\nu_i^1, ..., \nu_i^n) \in \mathcal{N}^+(S_{-i})$ such that $\nu_i^1 \left(S_{-i}^{m-1}\right) = 1$, and s_i is a lexicographic best reply to $\overline{\nu}_i$. We now show the existence of an LPS $\overline{\mu}_i = (\mu_i^1, ..., \mu_i^n) \in \mathcal{N}(S_{-i} \times T_{-i})$ such that

(a) $\overline{\mathrm{marg}}_{S_{-i}} \overline{\mu}_i = \overline{\nu}_i$; and

(b) $(R_{-i}^k)_{k=0}^{m-1}$ are weakly believed under $\overline{\mu}_i$.

To this end, note that, by the induction hypothesis, for each $s_{-i} \in S_{-i}^{m-1}$ there exists $t_{s_{-i}} \in T_{-i}$ such that $\left(s_{-i}, t_{s_{-i}}\right) \in R_{-i}^{m-1}$. So, for each $s_{-i} \in S_{-i}^{m-1}$ we fix some $t_{s_{-i}} \in T_{-i}$ for which $\left(s_{-i}, t_{s_{-i}}\right) \in R_{-i}^{m-1}$. We also fix an arbitrary $t_{-i}^0 \in T_{-i}$, and we define the map $\psi_{-i}^{m-1} : S_{-i} \to S_{-i} \times T_{-i}$ as

$$\psi_{-i}^{m-1}(s_{-i}) = \begin{cases} (s_{-i}, t_{s_{-i}}), & \text{if } s_{-i} \in S_{-i}^{m-1}, \\ (s_{-i}, t_{-i}^0), & \text{if } s_i \in S_{-i} \backslash S_{-i}^{m-1}. \end{cases}$$

(Of course, the map ψ_{-i}^{m-1} is continuous, since strategy sets are endowed with the discrete topology.) Define $\overline{\mu}_i \in \mathcal{N}(S_{-i} \times T_{-i})$ by $\overline{\mu}_i = \widehat{\psi}_{-i}^{m-1}(\overline{\nu}_i)$. It readily follows that $\overline{\mu}_i$ satisfies property (a), since $\mathrm{Proj}_{S_{-i}} \circ \psi_{-i}^{m-1} = \mathrm{Id}_{S_{-i}}$. Property (b) also holds, in that

$$\mu_i^1 \left(R_{-i}^{m-1}\right) = \nu_i^1 \left((\psi_{-i}^{m-1})^{-1} \left(R_{-i}^{m-1}\right)\right) \geq \nu_i^1 \left(S_{-i}^{m-1}\right) = 1, \tag{4.2}$$

where the inequality follows by construction of ψ_{-i}^{m-1} (i.e., $\psi_{-i}^{m-1}\left(S_{-i}^{m-1}\right) \subseteq R_{-i}^{m-1}$ and so $S_{-i}^{m-1} \subseteq (\psi_{-i}^{m-1})^{-1}\left(R_{-i}^{m-1}\right)$). Therefore, R_{-i}^{m-1} is weakly believed under $\overline{\mu}_i$. By the monotonicity property of weak belief (Property 3), it follows that, for each $k < m-1$, R_{-i}^k is weakly believed under $\overline{\mu}_i$. By belief-completeness, there exists $t_i \in T_i$ such that $\beta_i(t_i) = \overline{\mu}_i$. This implies $(s_i, t_i) \in R_i^m$, hence $s_i \in \mathrm{Proj}_{S_i}(R_i^m)$.

Part (ii): Fix a player $i \in I$. Lemma 2.(ii) gives that $\mathrm{Proj}_{S_i}(R_i^\infty) \subseteq S_i^\infty$. Conversely, first note that, since S^∞ is a full WBRS (Proposition 3), Lemma 3 entails the existence of a finite type structure $\mathcal{T}^* = \langle S_i, T_i^*, \beta_i^* \rangle_{i \in I}$ such that $\mathrm{Proj}_{S_i}(R_i^{*,\infty}) = S_i^\infty$ for each $i \in I$. Hence, for every $s_i \in S_i^\infty$, there exists $t_i^* \in T_i^*$ such that $(s_i, t_i^*) \in R_i^{*,\infty}$. Since $\mathcal{T}$ is universal, there exists a type morphism, viz. $(\varphi_i^*)_{i \in I}$, from $\mathcal{T}^*$ to $\mathcal{T}$. Moreover, since $\mathcal{T}^*$ is finite, the map $(\varphi_i^*)_{i \in I}$ is bimeasurable. It thus follows from Lemma 1 that $(s_i, \varphi_i^*(t_i)) \in R_i^\infty$. This shows that $S_i^\infty \subseteq \mathrm{Proj}_{S_i}(R_i^\infty) \neq \emptyset$, as required. ■

5 Discussion

5.1 Transparency of cautiousness

Fix a type structure $\mathcal{T} = \langle S_i, T_i, \beta_i \rangle_{i \in I}$. Say that $\mathcal{T}$ is a **cautious type structure** if all the types of all players are cautious. In a cautious type structure, not only all the types are cautious, but there is common certain belief of this. In other words, there is **transparency** of cautiousness. We let $C^\infty \subseteq \prod_{i \in I} S_i \times T_i$ denote the event corresponding to transparency of cautiousness in $\mathcal{T}$, and we let C_i^∞ denote its projection on $S_i \times T_i$.

The finite type structure $\mathcal{T}^* = \langle S_i, T_i^*, \beta_i^* \rangle_{i \in I}$ we construct for Lemma 3 is cautious. Since cautiousness is preserved by type morphisms (cf. Lemma 1) and the image of $\mathcal{T}^*$ in the universal type structure $\mathcal{T}$ via bimeasurable type morphism $(\varphi_i)_{i \in I}$ is a self-evident event (cf. Battigalli and Friedenberg, 2012, Appendix A), we have $\prod_{i \in I} \varphi_i(T_i^*) \subseteq C^\infty$. This implies that the proof of Theorem 2 can be easily adapted to provide an alternative justification of permissibility that makes the transparency of cautiousness fully explicit. We can substitute the event "cautious rationality" with "rationality and transparency of cautiousness" in $\mathcal{T}$, that is, we can define R_i^1 as the set $R_i^c \cap C_i^\infty$ instead of just R_i^c. Then, Theorem 2 holds *verbatim* with the new definition of R^1. In words, permissibility characterizes the behavioral implications of rationality, transparency of cautiousness, and common weak belief thereof.

5.2 Comparison with Börgers (1994) and Hu (2007)

In the standard subjective expected utility framework, Börgers (1994) introduced the solution concept of perfect p-rationalizability, and showed its equivalence with the Dekel-Fudenberg procedure when $p \to 1$. Say that a player has p-belief of event E if he assigns probability at least p to E. Perfect p-rationalizability is implied by the following assumptions:

(1) players' beliefs have full support on each opponents' strategy sets, and they maximize their expected payoffs with respect to their beliefs; and

(2) players have common p-belief of (1).

Börgers (1994) did not provide an explicit formulation of such assumptions in an epistemic framework. A formal characterization is given in Hu (2007, Section 5) and reviewed by Dekel and Siniscalchi (2015, Section 12.5). Specifically, Theorem 12.11 in Dekel and Siniscalchi (2015) states that, for a given finite game $G = \langle I, (S_i, \pi_i)_{i \in I} \rangle$, there is a $\delta \in (0, 1)$ such that, for $p \geq \delta$, the Dekel-Fudenberg procedure characterizes the behavioral implications of the above epistemic assumptions in the canonical type structure with beliefs represented by probability measures.[10] As stressed by Dekel and Siniscalchi (2015), this epistemic foundation of the Dekel-Fudenberg procedure depends crucially on δ, which in turn depends on the game G. By contrast, Theorem 2 provides one common epistemic condition—across all games—that yield the Dekel-Fudenberg procedure. We view this as a desirable feature, because the epistemic assumption of $\mathrm{R^cCWBR^c}$ is not tailored to a specific game G.

The comparison between permissibility and perfect p-rationalizability (and their epistemic foundations) can be also understood by providing an analogue of weak belief in terms of *nonstandard* probability measures. As shown in Blume et al. (1991a, Section 6; see also Halpern, 2010), a preference relation admitting an LPS representation can be also described by an $\mathbb{F}$-valued probability measure, where $\mathbb{F}$ is a non-Archimedean ordered field that is a strict extension of the set of real numbers $\mathbb{R}$. Furthermore, as claimed in Halpern (2010, Section 7) and formally shown in Appendix A, weak belief of event $E \subseteq X$

[10] In fact, Dekel and Siniscalchi (2015) state the result for belief-complete type structures, with the additional technical assumption of compact type spaces and continuous belief maps. A result due to Friedenberg (2010) shows that, in such a case, every belief-complete type structure is, in a very specific sense, "equivalent" to the canonical type structure—see Theorem 3.1 and Proposition 4.1 in Friedenberg (2010) for a precise statement.

under LPS $\overline{\mu}$ admits the following analogue in terms of nonstandard probabilities: Let ν be the nonstandard probability representing the same preference relation as $\overline{\mu}$; then E is weakly believed under ν if $st(\nu(E)) = 1$. (Here, $st(x)$ denotes the closest standard real to $x \in \mathbb{F}$, called "the standard part of x.")

With this, we can now see how p-belief of E can be thought of as a "approximation" of weak belief of E in terms of standard probabilities, as long as p is sufficiently close to 1. Full-support probability measures are sufficient to describe cautious behavior, and so the avoidance of weakly dominated strategies (Pearce, 1984, Lemma 4). But weak belief coincides with certain belief when preferences conform to subjective expected-utility theory. So, for $p \to 1$, the notion of p-belief of E reflects the same idea as that of weak belief, that is, not-E can be assigned strictly positive but infinitesimal probability.

5.3 Comparison with Asheim and Dufwenberg (2003)

Asheim and Dufwenberg (2003) use an epistemic formalism related to, but different from, the one we have adopted in this paper. In their framework, each type $t_i \in T_i$ is associated with a preference relation on the set of acts on $S_{-i} \times T_{-i}$, and such preferences need not satisfy completeness or Archimedean continuity. Asheim and Dufwenberg restrict attention to finite type structures, and they provide the following epistemic foundation for the permissibility set (later analyzed also by Perea, 2012): There exists a type structure where permissibility characterizes the behavioral implications of "caution, rationality, and common certain belief of caution and primary belief in rationality." Caution coincides with full-support, and primary belief,[11] when preferences are complete and thus can be represented by an LPS, corresponds to weak belief. We now recast their analysis in the current framework of lexicographic type structures.

Fix a type structure $\mathcal{T} = \langle S_i, T_i, \beta_i \rangle_{i \in I}$. Say that type t_i certainly believes a non-empty event $E_{-i} \subseteq S_{-i} \times T_{-i}$ if E_{-i} is certainly believed under $\beta_i(t_i)$. For each player $i \in I$, let $\mathbf{B}_i : 2^{S_{-i} \times T_{-i}} \to 2^{S_i \times T_i}$ be the operator defined by

$$\mathbf{B}_i(E_{-i}) = \{(s_i, t_i) \in S_i \times T_i \,| t_i \text{ certainly believes } E_{-i}\},$$

for every event $E_{-i} \subseteq S_{-i} \times T_{-i}$; if E_{-i} is not Borel, then $\mathbf{B}_i(E_{-i}) = \emptyset$.[12]

We define the event "caution" as follows: for each $i \in I$,

$$\bar{C}_i = \left\{(s_i, t_i) \in S_i \times T_i | \beta_i(t_i) \in \mathcal{N}^+ (S_{-i} \times T_{-i})\right\}.$$

For each player $i \in I$, let $D_i^0 = \bar{C}_i \cap \mathbf{WB}_i(R_{-i})$, and, recursively, let $D_i^m = \mathbf{B}_i(D_{-i}^{m-1})$. We let $\bar{R}_i^1 = R_i \cap \bar{C}_i$, and, for each $m > 1$, define $\bar{R}_i^m$ recursively by

$$\bar{R}_i^m = \bar{R}_i^{m-1} \cap D_i^{m-1}.$$

We write $\bar{R}_i^\infty = \cap_{m \in \mathbb{N}} \bar{R}_i^m$ for each $i \in I$. Note that, for each $m > 1$,

$$\bar{R}_i^m = \bar{R}_i^1 \cap \left(\cap_{l=1}^{m-1} D_i^l\right).$$

[11] This notion is simply called "belief" in Asheim and Dufwenberg (2003); the terminology "primary belief" is used by Perea (2012).

[12] A straightforward modification of the proofs in Appendix C shows that the set $\mathbf{B}_i(E_{-i})$ is Borel in $S_i \times T_i$ for every event $E_{-i} \subseteq S_{-i} \times T_{-i}$.

The set $\prod_{i\in I}\bar{R}_i^\infty$ corresponds to the set of all states in which there is "caution, rationality, and common certain belief of caution and primary belief in rationality."

In the canonical type structure, this event does not yield the permissibility set. This is because $\bar{R}_i^\infty$ does not impose weak belief of rationality of the opponent. Hence, it is possible to show that $\mathrm{Proj}_{S_i}\bar{R}_i^\infty = \mathrm{Proj}_{S_i}\rho^1(S)$.

Instead, the event $\prod_{i\in I}\hat{R}_i^\infty$ where

$$\hat{R}_i^\infty = \bar{R}_i^1 \cap \left(\cap_{l=0}^\infty D_i^l\right) = \bar{R}_i^\infty \cap \mathbf{WB}_i\left(R_{-i}\right),\ i \in I,$$

captures the permissibility set in the canonical type structure. However, it is not true that the event

$$\hat{R}_i^n = \bar{R}_i^1 \cap \left(\cap_{l=0}^{n-1} D_i^l\right)$$

captures the strategies of player i that survives n steps of permissibility. This is because D_i^1 already imposes certain belief of the event that the opponent weakly believes in the rationality of player i. We now exhibit an example which shows the existence of a strategy $s_i \in S_i$ which survives the first two steps of permissibility, but $s_i \notin \mathrm{Proj}_{S_i}\hat{R}_i^2$ in every type structure, including the canonical one.

$1\backslash 2$	ℓ	c	r
u	$1,1$	$0,1$	$1,0$
m	$0,0$	$1,1$	$1,0$
d	$0,0$	$0,0$	$0,1$
b	$2,0$	$0,1$	$0,0$

In this two-player game, there is a unique permissible strategy profile, which results from the first three steps of the procedure:

$$\begin{aligned}\rho^1(S) &= \{u,m,b\} \times \{c,r\}\,,\\ \rho^2(S) &= \{u,m\} \times \{c\}\,,\\ \rho^3(S) &= \{m\} \times \{c\}\,.\end{aligned}$$

We now show that, although u survives the first two steps of the procedure, we cannot have $u \in \mathrm{Proj}_{S_1}\hat{R}_1^2$, independently of the type structure. To this end, fix any $(s_1,t_1) \in \hat{R}_1^2$ and write $\beta_1(t_1) = (\mu_1^1,...,\mu_1^n)$. Note that, by definition,

$$\hat{R}_1^2 = R_1 \cap \bar{C}_1 \cap \mathbf{WB}_1\left(R_2\right) \cap \mathbf{B}_1(\bar{C}_2 \cap \mathbf{WB}_2\left(R_1\right));$$

hence

$$\mu_1^1(R_2 \cap \bar{C}_2 \cap \mathbf{WB}_2\left(R_1\right)) = 1.$$

Next note that:

(a) every strategy-type pair $(s_2,t_2) \in R_2 \cap \bar{C}_2$ must satisfy $s_2 \neq \ell$, and

(b) $s_2 \neq r$ for every $(s_2,t_2) \in R_2 \cap \mathbf{WB}_2\left(R_1\right)$.

Indeed, ℓ is weakly dominated, so it cannot optimal under a full-support first-order belief of a type t_2. Property (b) holds because r is not a lexicographic best reply to any first-order belief under which $\{u,m,b\}$ is weakly believed. Since $(s_1,t_1) \in R_1$, this implies that $s_1 = m$ and so $u \notin \mathrm{Proj}_{S_1}\hat{R}_1^2$.

Since $\mathrm{Proj}_{S_1}\hat{R}_1^2$ is strictly included in $\mathrm{Proj}_{S_1}\rho^2(S) = \{u, m\}$, one may hope that $\mathrm{Proj}_{S_1}\hat{R}_1^1 = \mathrm{Proj}_{S_1}\rho^2(S)$. But this is not generally true either. To see this, fix a belief-complete type structure for the game above. Then $\mathrm{Proj}_{S_2}R_2 = \{\ell, c, r\}$ because R_2 does not require full-support. So, we can construct a full-support LPS $\overline{\mu}_1 = (\mu_1^1, \mu_1^2) \in \mathcal{N}^+(S_1 \times T_1)$ such that $\mu_1^1(\{\ell\} \times T_2) \geq 1/2$ and $\mu_1^1(R_2) = 1$. Then, by belief completeness, there exists $t_1 \in T_1$ such that $\beta_1(t_1) = \overline{\mu}_1$, and it can be easily seen that

$$(b, t_1) \in \hat{R}_1^1 = R_1 \cap \bar{C}_1 \cap \mathbf{WB}_1(R_2),$$

although $b \notin \mathrm{Proj}_{S_1}\rho^2(S)$. So $\mathrm{Proj}_{S_1}\hat{R}_1^1$ strictly includes $\mathrm{Proj}_{S_1}\rho^2(S)$. This is because $\hat{R}_i^1$ does not impose weak or certain belief in caution.

To summarize, $\prod_{i\in I}\bar{R}_i^\infty$ captures the permissibility set in an ad-hoc type structure, but not in the canonical one. This can be corrected with small amendment (event $\prod_{i\in I}\hat{R}_i^\infty$), but the iterative construction of the event still does not correspond to the steps of permissibility. The reference point of Asheim and Dufwenberg is actually not permissibility but the Dekel-Fudenberg procedure.[13] This explains the use of two different notions of belief, which mirrors the conceptual difference between the first step of the procedure (based on weak dominance) and the further steps (based on strict dominance).

Differently from Asheim and Dufwenberg (2003), we provide epistemic conditions in the canonical type structure for each step of permissibility by using only the notion of weak belief. Moreover, we answer the following "inverse question": Given a type structure, what are the behavioral implications of R$^{\mathrm{c}}$CWBR$^{\mathrm{c}}$? We conjecture that full weak best response sets also characterize the behavioral implications of $\prod_{i\in I}\hat{R}_i^\infty$ across all type structures. This analogue of Theorem 1 shall not mislead: Event R$^{\mathrm{c}}$CWBR$^{\mathrm{c}}$ is larger than $\prod_{i\in I}\hat{R}_i^\infty$, possibly also in terms of projections on strategy sets.

5.4 Cautious belief and iterated admissibility

The use of weak belief in place of certain belief also sheds light on the conceptual difference between permissibility and iterated admissibility (i.e., maximal iterated elimination of weakly dominated strategies). In Catonini and De Vito (2018), we show that iterated admissibility characterizes the behavioral implications of cautious rationality and common *cautious* belief of cautious rationality. Cautious belief strengthens weak belief by introducing a cautious attitude towards the believed event. An event $E \subseteq S_{-i} \times T_{-i}$ is cautiously believed under the LPS $\overline{\mu}_i = (\mu^1, ..., \mu^n)$ if (i) there is $m \leq n$ such that $\mu^l(E) = 1$ for each $l = 1, ..., m$, and (ii) for every $s_{-i} \in \mathrm{Proj}_{S_{-i}}E$, $\mu^j(\{s_{-i}\} \times T_{-i}) > 0$ for some $j \leq m$. Requirement (i) coincides with weak belief. Requirement (ii) means that player i considers every payoff-relevant subset of E infinitely more likely than not-E. Player i is then cautious towards the believed event in the same way player i is cautious in general: before entertaining the possibility that E does not occur, player i cautiously takes into consideration all the possible payoff-relevant implications of E. So, the conceptual difference between permissibility and iterated admissibility lies in the presence or not of this cautious attitude towards opponents' rationality and beliefs in rationality of all orders. In permissibility, when players weakly believe in the opponents' rationality, they

[13]It should be noted that Asheim and Dufwenberg use the Dekel-Fudenberg procedure as a primitive definition for the concept of permissible strategies—see Asheim and Dufwenberg (2003, Definition C.1).

can ignore some payoff-relevant implications of it, and only consider them together with other payoff-relevant events that are not necessarily compatible with the opponents' rationality. In iterated admissibility, players cautiously believe in the opponents' rationality by considering all its payoff-relevant implications before considering any other payoff-relevant event.

Appendix A: The Decision-theoretic Framework

Fix a lexicographic type structure $\mathcal{T} = \langle S_i, T_i, \beta_i \rangle_{i \in I}$, where each strategy set S_i is finite. To ease notation, it will be convenient to set $\Omega = S_{-i} \times T_{-i}$ and to drop i's subscript from LPS's $\overline{\mu}_i$ on Ω.

An act on Ω is a Borel measurable function $f : \Omega \to [0,1]$. We let ACT(Ω) denote the set of all acts on Ω. A Decision Maker has preferences over elements of ACT(Ω). For $x \in [0,1]$, write $\overrightarrow{x}$ for the constant act associated with x, i.e., $\overrightarrow{x}(\omega) = x$ for all $\omega \in \Omega$. Each constant act is identified with the associated outcome in a natural way. In what follows, we assume that the outcome space $[0,1]$ is in utils, i.e., material consequences are replaced by their von Neumann-Morgenstern utility. Given a Borel set $E \subseteq \Omega$ and acts $f, g \in$ACT(Ω), define $(f_E, g_{\Omega \backslash E}) \in$ACT($\Omega$) as follows:

$$(f_E, g_{\Omega\backslash E})(\omega) = \begin{cases} f(\omega), & \text{if } \omega \in E, \\ g(\omega), & \text{if } \omega \in \Omega\backslash E. \end{cases}$$

Let $\succsim$ be a preference relation on ACT(Ω) and write $\succ$ (resp. $\sim$) for strict preference (resp. indifference). The preference relation $\succsim$ satisfies the following axioms:

Axiom 1 Order: $\succsim$ is a complete, transitive, reflexive binary relation on ACT(Ω).

Axiom 2 Independence: For all $f, g, h \in$ACT(Ω) and $\alpha \in (0,1]$,

$$\begin{aligned} f &\succ g \text{ implies } \alpha f + (1-\alpha) h \succ \alpha g + (1-\alpha) h, \text{ and} \\ f &\sim g \text{ implies } \alpha f + (1-\alpha) h \sim \alpha g + (1-\alpha) h. \end{aligned}$$

Moreover, let $\succsim_E$ denote the *conditional preference* given E, that is, $f \succsim_E g$ if and only if $(f_E, h_{\Omega\backslash E}) \succsim (g_E, h_{\Omega\backslash E})$ for some $h \in$ACT(Ω). Standard results (see Blume et al., 1991a, for a proof) show that, under Axioms 1 and 2, $(f_E, h_{\Omega\backslash E}) \succsim (g_E, h_{\Omega\backslash E})$ holds for all $h \in$ACT(Ω) if it holds for some h.

Throughout, we maintain the assumption that $\overline{\mu}$ is a Lexicographic Expected Utility representation of $\succsim$, i.e., $\succsim = \succsim^{\overline{\mu}}$. (This makes sense, since each Lexicographic Expected Utility representation satisfies Axioms 1 and 2.) Now we can introduce the notions of more likely and infinitely more likely.

Definition A.1 *Fix $\overline{\mu} = (\mu^1, ..., \mu^n) \in \mathcal{N}(\Omega)$ and events $E, F \subseteq \Omega$ with $E \neq \emptyset$. Say that E is* ***more likely than*** *F under $\succsim^{\overline{\mu}}$ if for all $x, y \in [0,1]$ with $x > y$,*

$$(\overrightarrow{x}_E, \overrightarrow{y}_{\Omega\backslash E}) \succsim^{\overline{\mu}} (\overrightarrow{x}_F, \overrightarrow{y}_{\Omega\backslash F}).$$

Say that E is deemed ***infinitely more likely than*** *F under $\succsim^{\overline{\mu}}$ (Lo, 1999), and write $E \gg^{\bar{\mu}} F$, if for all $x, y, z \in [0,1]$ with $x > y$,*

$$(\overrightarrow{x}_E, \overrightarrow{y}_{\Omega\backslash E}) \succ^{\overline{\mu}} (\overrightarrow{z}_F, \overrightarrow{y}_{\Omega\backslash F}).$$

In words, E is more likely than F if the Decision Maker prefers to bet on E rather than on F given the same prizes for the two bets; this choice theoretic notion is due to Savage (1972). On the other hand, E is infinitely more likely than F if the Decision Maker *strictly* prefers to bet on E rather than on F, and increasing the prize by any extent for the second bet does not induce his strict preference to change.

Recall that an event $E \subseteq \Omega$ is **Savage-null** under $\succsim$ if $f \sim_E g$ for all $f, g \in$ACT(Ω). Say that E is **non-null** under $\succsim$ if it is not Savage-null under $\succsim$. Thus, note that if $E \gg^{\bar{\mu}} F$, then E is non-null under $\succsim^{\overline{\mu}}$, while F may, but *need not*, be Savage-null under $\succsim^{\overline{\mu}}$. When $\succsim^{\overline{\mu}}$ has a subjective expected utility representation, $E \gg^{\bar{\mu}} F$ implies that F is Savage-null.

The likelihood relation $\gg^{\bar{\mu}}$ possesses some natural properties. First, it is irreflexive, asymmetric and transitive. Moreover, it is monotone; that is, if $E \gg^{\bar{\mu}} F$, then

(P1) E is infinitely more likely than any Borel subset of F; and

(P2) any Borel superset of E is infinitely more likely than F.

The final step to motivate the analysis of Section 2.2 is to characterize the likelihood order $\gg^{\bar{\mu}}$ between pairwise *disjoint* events in terms of LPS's representing $\succsim^{\overline{\mu}}$ as in Definition 1. Recall from Section 2.2 that, given an LPS $\overline{\mu} = (\mu^1, ..., \mu^n) \in \mathcal{N}(\Omega)$ and non-empty event $E \subseteq \Omega$, we let

$$\mathcal{I}_{\overline{\mu}}(E) = \inf \left\{ l \in \{1, ..., n\} \left| \mu^l(E) > 0 \right. \right\}.$$

Proposition A.1 *Fix $\overline{\mu} = (\mu^1, ..., \mu^n) \in \mathcal{N}(\Omega)$ and events $E, F \subseteq \Omega$ with $E \neq \emptyset$.*

1. *E is more likely than F under $\succsim^{\overline{\mu}}$ if and only if*

$$\left(\mu^l(E)\right)_{l=1}^n \geq_L \left(\mu^l(F)\right)_{l=1}^n.$$

2. *$E \gg^{\bar{\mu}} F$ if and only if $\mathcal{I}_{\overline{\mu}}(E) < \mathcal{I}_{\overline{\mu}}(F)$.*

Proof: The proof of part 1 is left to the reader. We prove part 2. To this end, we first record the following fact: Fix $x, y, z \in [0,1]$ with $x > y$. We have that

$$(\overrightarrow{x}_E, \overrightarrow{y}_{\Omega\backslash E}) \succ^{\overline{\mu}} (\overrightarrow{z}_F, \overrightarrow{y}_{\Omega\backslash F}) \Longleftrightarrow \left((x-y)\,\mu^l(E)\right)_{l=1}^n >_L \left((z-y)\,\mu^l(F)\right)_{l=1}^n. \tag{A.1}$$

To see this, note that $(\overrightarrow{x}_E, \overrightarrow{y}_{\Omega\backslash E}) \succ^{\overline{\mu}} (\overrightarrow{z}_F, \overrightarrow{y}_{\Omega\backslash F})$ holds if and only if

$$\left(\int_E x d\mu^l + \int_{\Omega\backslash E} y d\mu^l\right)_{l=1}^n >_L \left(\int_F z d\mu^l + \int_{\Omega\backslash F} y d\mu^l\right)_{l=1}^n,$$

that is, if and only if

$$\left(x\mu^{l}\left(E\right)+y\mu^{l}\left(\Omega\backslash E\right)\right)_{l=1}^{n}>_{L}\left(z\mu^{l}\left(F\right)+y\mu^{l}\left(\Omega\backslash F\right)\right)_{l=1}^{n},$$

and so if and only if $\left(\left(x-y\right)\mu^{l}\left(E\right)\right)_{l=1}^{n}>_{L}\left(\left(z-y\right)\mu^{l}\left(F\right)\right)_{l=1}^{n}$.

Suppose that $\mathcal{I}_{\overline{\mu}}\left(E\right)<\mathcal{I}_{\overline{\mu}}\left(F\right)$. Then, for every $x,y,z\in[0,1]$ with $x>y$,

$$\left(\left(x-y\right)\mu^{l}\left(E\right)\right)_{l=1}^{n}>_{L}\left(\left(z-y\right)\mu^{l}\left(F\right)\right)_{l=1}^{n}.$$

By (A.1), it follows that $(\overrightarrow{x}_{E},\overrightarrow{y}_{\Omega\backslash E})\succ^{\overline{\mu}}(\overrightarrow{z}_{F},\overrightarrow{y}_{\Omega\backslash F})$. Hence $E\gg^{\bar{\mu}}F$.

Suppose that $\mathcal{I}_{\overline{\mu}}\left(E\right)\geq\mathcal{I}_{\overline{\mu}}\left(F\right)$. Then, for every $x,y,z\in[0,1]$ with $x>y$ and $x<z\cdot\mu^{\mathcal{I}_{\overline{\mu}}(F)}\left(F\right)$,

$$\left(\left(z-y\right)\mu^{l}\left(F\right)\right)_{l=1}^{n}>_{L}\left(\left(x-y\right)\mu^{l}\left(E\right)\right)_{l=1}^{n}.$$

By (A.1), it follows that $(\overrightarrow{z}_{F},\overrightarrow{y}_{\Omega\backslash F})\succ^{\overline{\mu}}(\overrightarrow{x}_{E},\overrightarrow{y}_{\Omega\backslash E})$. So, it is not true that $E\gg^{\bar{\mu}}F$.■

We conclude this section by providing an additional perspective on weak belief in terms of infinitesimal nonstandard real numbers. As is well known (see Blume et al. 1991a, Section 6), a preference relation admitting a Lexicographic Expected Utility representation can be equivalently described by an $\mathbb{F}$-valued probability measure on Ω, where $\mathbb{F}$ is a non-Archimedean ordered field which is a strict extension of the set of real numbers $\mathbb{R}$. For instance, the LPS $\overline{\mu}=(\mu^{1},\mu^{2})$ can be represented by a nonstandard real valued probability measure $\nu=(1-\varepsilon)\mu^{1}+\varepsilon\mu^{2}$, where $\varepsilon>0$ is an infinitesimal nonstandard real such that for each real number $x>0$ and each $n\in\mathbb{N}$, it is the case that $x>n\varepsilon$.

Given nonstandard reals x and y, say that x is **infinitely greater** than y if $x>ny$ for each $n\in\mathbb{N}$. The notion of infinitely more likely in Definition 1 corresponds exactly to the "infinitely greater" relation between the nonstandard probability values that provide an equivalent representation of preferences. With this, we now show that weak belief can be given a nonstandard characterization as follows. Fix a non-empty event $E\subseteq\Omega$. Then E is weakly believed under $\mathbb{F}$-valued probability measure ν if and only if

$$\nu\left(E\right)>n\nu\left(\Omega\backslash E\right)$$

for every $n\in\mathbb{N}$, which is equivalent to say that

$$\nu\left(\Omega\backslash E\right)<\frac{1}{1+n}$$

for every $n\in\mathbb{N}$. For each non-standard real x, let $st\left(x\right)$ denote its standard part. So, in light of the above, event E is weakly believed under ν if and only if $st\left(\nu\left(\Omega\backslash E\right)\right)=0$, or, equivalently, $st\left(\nu\left(E\right)\right)=1$ (cf. Halpern, 2010).

Appendix B: Proofs for Section 3

Proof of Proposition 3: Note that a set Q is a WBRS if $Q\subseteq\rho(Q)$. The set of permissible strategies $\rho^{\infty}(S)$ is a WBRS, since $\rho^{\infty}(S)=\rho(\rho^{\infty}(S))$. It is immediate to see that $\rho^{\infty}(S)$ is a full WBRS. Fix some WBRS $Q\in\mathcal{Q}$. Using the monotonicity property of ρ (Remark 2) and the fact that $Q \subseteq \rho(Q)$, an easy induction argument shows that $Q \subseteq \rho^k(Q) \subseteq \rho^k(S)$ for all $k \geq 1$. Therefore, $Q \subseteq \rho^\infty(S)$. ■

For the proof of Proposition 4, we first note the following facts pertaining the set $\mathrm{su}(s_i)$ for some $s_i \in S_i$.

Remark B.1 *Fix $s_i \in S_i$ and $\nu_i \in \mathcal{M}(S_{-i})$. If $s_i \in r_i(\nu_i)$, then $\mathrm{su}(s_i) \subseteq r_i(\nu_i)$ (cf. Brandenburger et al., 2012, Lemma 5.2). Specifically, if s_i is the unique best reply to ν_i, then $r_i(\nu_i) = \mathrm{su}(s_i) = \{s_i\}$.*

Remark B.2 *Fix $s_i \in S_i$. If s_i is not weakly dominated, then there exists a full-support $\nu_i \in \mathcal{M}(S_{-i})$ such that $s_i \in r_i(\nu_i)$ and $r_i(\nu_i) = \mathrm{su}(s_i)$. This result is Lemma D.4 in Brandenburger et al. (2008), which is restated here in terms of the intended interpretation of the geometric notions in game theory (see Brandenburger et al., 2008, p.343).*

We break the proof of Proposition 4 in two lemmas.

Lemma B.1 *A set $Q \in \mathcal{Q}$ is a WBRS if and only if, for each $i \in I$,*
(i) each $s_i \in Q_i$ is not weakly dominated, and
(ii) each $s_i \in Q_i$ is not strictly dominated with respect to $S_i \times Q_{-i}$.

Proof: Fix a player $i \in I$. If $s_i \in Q_i$ is not weakly dominated, then by Lemma 4 in Pearce (1984) there exists $\mu_i^2 \in \mathcal{M}(S_{-i})$ such that $\mathrm{Supp}\mu_i^2 = S_{-i}$ and $\pi_i(s_i, \mu_i^2) \geq \pi_i(s_i', \mu_i^2)$ for every $s_i' \in S_i$. Moreover, if $s_i \in Q_i$ is not strictly dominated with respect to $S_i \times Q_{-i}$, then Lemma 3 in Pearce (1984) yields the existence of $\mu_i^1 \in \mathcal{M}(S_{-i})$ such that $\mu_i^1(Q_{-i}) = 1$ and $\pi_i(s_i, \mu_i^1) \geq \pi_i(s_i', \mu_i^1)$ for every $s_i' \in S_i$. Let $\overline{\mu}_i = (\mu_i^1, \mu_i^2)$. We clearly have $\overline{\mu}_i \in \mathcal{W}^+(Q_{-i})$ and $s_i \in r_i(\overline{\mu}_i)$, hence $s_i \in \rho_i(Q_{-i})$. Since $i \in I$ and s_i are arbitrary, this shows that $Q \subseteq \rho(Q)$, i.e., Q is a WBRS.

Conversely, fix a WBRS $Q \in \mathcal{Q}$. Then, for each $s_i \in Q_i$, we can take the first component probability of the LPS $\overline{\mu}_i \in \mathcal{W}^+(Q_{-i})$ for which $s_i \in r_i(\overline{\mu}_i)$, so to claim the existence of $\mu_i^1 \in \mathcal{M}(S_{-i})$ such that $\mu_i^1(Q_{-i}) = 1$ and $\pi_i(s_i, \mu_i^1) \geq \pi_i(s_i', \mu_i^1)$ for every $s_i' \in S_i$. By Lemma 3 in Pearce (1984), it follows that $s_i \in Q_i$ is not strictly dominated with respect to $S_i \times Q_{-i}$. Moreover, by Proposition 1 in Blume et al. (1991b), there exists $\nu_i \in \mathcal{M}(S_{-i})$, with $\mathrm{Supp}\nu_i = S_{-i}$, which is preference equivalent to $\overline{\mu}_i$. Thus, $\pi_i(s_i, \nu_i) \geq \pi_i(s_i', \nu_i)$ for every $s_i' \in S_i$. Then, using again Lemma 4 in Pearce (1984), we can conclude that each $s_i \in Q_i$ is not weakly dominated. ■

Lemma B.2 *Fix a WBRS Q and $s_i \in Q_i$. Then $\mathrm{su}(s_i) \subseteq Q_i$ if and only if there exists $\overline{\mu}_i \in \mathcal{W}^+(Q_{-i})$ such that $s_i \in r_i(\overline{\mu}_i)$ and $r_i(\overline{\mu}_i) \subseteq Q_i$.*

Proof: Suppose $\mathrm{su}(s_i) \subseteq Q_i$. Lemma B.1 entails that s_i is not weakly dominated. So, by Remark B.2, there exists a full-support $\mu_i^2 \in \mathcal{M}(S_{-i})$ such that $s_i \in r_i(\mu_i^2)$ and $r_i(\mu_i^2) = \mathrm{su}(s_i)$. Moreover, by Lemma B.1, s_i is not strictly dominated with respect to $S_i \times Q_{-i}$. Hence, there exists $\mu_i^1 \in \mathcal{M}(S_{-i})$ such that $s_i \in r_i(\mu_i^1)$ and $\mu_i^1(Q_{-i}) = 1$. Let $\overline{\mu}_i = (\mu_i^1, \mu_i^2)$. We clearly have $s_i \in r_i(\overline{\mu}_i)$ and $\overline{\mu}_i \in \mathcal{W}^+(Q_{-i})$. We now claim that, if $s_i' \in r_i(\overline{\mu}_i)$ then $s_i' \in r_i(\mu_i^2)$; this will imply $r_i(\overline{\mu}_i) \subseteq Q_i$, since $r_i(\mu_i^2) = \mathrm{su}(s_i) \subseteq Q_i$.

So, let $s_i' \in r_i(\overline{\mu}_i)$; it is enough to consider two cases:

(a) $\pi_i(s_i', \mu_i^1) = \pi_i(s_i'', \mu_i^1)$ for all $s_i'' \in S_i$, and $\pi_i(s_i', \mu_i^2) \geq \pi_i(s_i'', \mu_i^2)$ for all $s_i'' \in S_i$. Of course, this implies $s_i' \in r_i(\mu_i^2)$.

(b) $\pi_i(s_i', \mu_i^1) > \pi_i(s_i'', \mu_i^1)$ for all $s_i'' \in S_i$. In this case, s_i' is the unique best reply to μ_i^1, and so it must be $s_i' = s_i$. Since $s_i \in r_i(\mu_i^2)$, this proves the claim.

Conversely, if $s_i \in r_i(\overline{\mu}_i)$ for some $\overline{\mu}_i \in \mathcal{W}^+(Q_{-i})$, then, by Proposition 1 in Blume et al. (1991b), there exists $\nu_i \in \mathcal{M}(S_{-i})$, with $\text{Supp}\nu_i = S_{-i}$, which is preference equivalent to $\overline{\mu}_i$, and such that $s_i \in r_i(\nu_i)$. By Remark B.1, it follows that $\text{su}(s_i) \subseteq r_i(\nu_i)$. Clearly, if $s_i' \in r_i(\nu_i)$, then $s_i' \in r_i(\overline{\mu}_i)$, because ν_i and $\overline{\mu}_i$ are preference equivalent. Since $r_i(\overline{\mu}_i) \subseteq Q_i$, this implies $\text{su}(s_i) \subseteq Q_i$, as required.

Proof of Proposition 4: Immediate from Lemma B.1 and Lemma B.2. ■

Appendix C: Proofs for Section 4

We first show that, for a given type structure $\mathcal{T} = \langle S_i, T_i, \beta_i \rangle_{i \in I}$, the sets R_i^m, $m > 1$, as defined in the main text, are Borel subsets of $S_i \times T_i$.

Lemma C.1 *Fix a non-empty event $E \subseteq S_{-i} \times T_{-i}$. Then the set of all $\overline{\mu} \in \mathcal{N}(S_{-i} \times T_{-i})$ under which E is weakly believed is Borel in $\mathcal{N}(S_{-i} \times T_{-i})$.*

Proof: By Theorem 17.24 in Kechris (1995), it follows that, for a given event $E \subseteq S_{-i} \times T_{-i}$, the set of probability measures μ satisfying $\mu(E) = p$ for $p \in \mathbb{Q} \cap [0,1]$ is measurable in $\mathcal{M}(S_{-i} \times T_{-i})$. So the set of all $\mu \in \mathcal{M}(S_{-i} \times T_{-i})$ satisfying $\mu(E) = 1$ is Borel in $\mathcal{M}(S_{-i} \times T_{-i})$. Now, fix $n \in \mathbb{N}$. By the above argument and by definition of $\mathcal{N}_n(S_{-i} \times T_{-i})$, it turns out that the set

$$\begin{aligned}\mathcal{U}_n^1 &= \left\{\overline{\mu} \in \mathcal{N}_n(S_{-i} \times T_{-i}) \,\middle|\, \mu^1(E) = 1\right\} \\ &= \left\{\mu \in \mathcal{M}(S_{-i} \times T_{-i}) \,\middle|\, \mu^1(E) = 1\right\} \times (\mathcal{M}(S_{-i} \times T_{-i}))^{n-1}\end{aligned}$$

is Borel in $\mathcal{N}_n(S_{-i} \times T_{-i})$. The set of all $\overline{\mu} \in \mathcal{N}(S_{-i} \times T_{-i})$ under which E is weakly believed is given by $\cup_{n \in \mathbb{N}} \mathcal{U}_n^1$, so Borel in $\mathcal{N}(S_{-i} \times T_{-i})$. ■

By the measurability of each belief map in a type structures, we have the following.

Corollary C.1 *For every $i \in I$, if $E \subseteq S_{-i} \times T_{-i}$ is a non-empty event, then $\mathbf{WB}_i(E)$ is a Borel subset of $S_i \times T_i$.*

We let $\mathcal{C}_i$ denote the set of all $\overline{\mu} \in \mathcal{N}(S_{-i} \times T_{-i})$ such that $\overline{\text{marg}}_{S_{-i}}\overline{\mu} \in \mathcal{N}^+(S_{-i})$.

Lemma C.2 *The set $\mathcal{C}_i$ is Borel in $\mathcal{N}(S_{-i} \times T_{-i})$.*

Proof: Note that $\overline{\mu} = (\mu^1, ..., \mu^n) \in \mathcal{C}_i$ if and only if, for each $s_{-i} \in S_{-i}$, there is $l \leq n$ such that $\mu^l(\{s_{-i}\} \times T_{-i}) > 0$. Let

$$\mathcal{C}_i^n = \cap_{s_{-i} \in S_{-i}} \left(\mathcal{N}_n(S_{-i} \times T_{-i}) \backslash \left(\cap_{l \leq n} \left\{\overline{\mu} \in \mathcal{N}_n(S_{-i} \times T_{-i}) \,\middle|\, \mu^l(\{s_{-i}\} \times T_{-i}) = 0\right\}\right)\right).$$

Each set $\left\{\overline{\mu} \in \mathcal{N}_n(S_{-i} \times T_{-i}) \,\middle|\, \mu^l(\{s_{-i}\} \times T_{-i}) = 0\right\}$ is Borel by Theorem 17.24 in Kechris (1995). Therefore $\mathcal{C}_i = \cap_{n \in \mathbb{N}} \mathcal{C}_i^n$ is Borel. ■

Lemma C.3 *Fix $s_i \in S_i$. The set of all $\overline{\mu} \in \mathcal{N}(S_{-i} \times T_{-i})$ such that s_i is a lexicographic best reply to $\overline{\mathrm{marg}}_{S_{-i}}\overline{\mu}$ is Borel in $\mathcal{N}(S_{-i} \times T_{-i})$.*

To prove Lemma C.3, we need the following auxiliary result:

Lemma C.4 *Fix a type structure $\mathcal{T} = \langle S_i, T_i, \beta_i \rangle_{i \in I}$ and $s_i, s_i' \in S_i$. Let $O^{\mathcal{W}}_{s_i,s_i'}$ and $O^{\mathcal{S}}_{s_i,s_i'}$ be subsets of $\mathcal{M}(S_{-i} \times T_{-i})$ defined as follows:*

$$\begin{aligned} O^{\mathcal{W}}_{s_i,s_i'} &= \left\{\mu \in \mathcal{M}(S_{-i} \times T_{-i}) \,\middle|\, \pi_i(s_i, \mathrm{marg}_{S_{-i}}\mu) \geq \pi_i(s_i', \mathrm{marg}_{S_{-i}}\mu) \right\}, \\ O^{\mathcal{S}}_{s_i,s_i'} &= \left\{\mu \in \mathcal{M}(S_{-i} \times T_{-i}) \,\middle|\, \pi_i(s_i, \mathrm{marg}_{S_{-i}}\mu) > \pi_i(s_i', \mathrm{marg}_{S_{-i}}\mu) \right\}. \end{aligned}$$

Then $O^{\mathcal{W}}_{s_i,s_i'}$ and $O^{\mathcal{S}}_{s_i,s_i'}$ are closed and open in $\mathcal{M}(S_{-i} \times T_{-i})$, respectively.

Proof: First recall that the map $\widetilde{\mathrm{Proj}}_{S_{-i}} : \mathcal{M}(S_{-i} \times T_{-i}) \to \mathcal{M}(S_{-i})$ defined by

$$\mu \mapsto \mathrm{marg}_{S_{-i}}\mu,\ \mu \in \mathcal{M}(S_{-i} \times T_{-i}),$$

is continuous. Moreover, for each $\widetilde{s}_i \in S_i$, the function $\pi_i(\widetilde{s}_i, \cdot) : \mathcal{M}(S_{-i}) \to \mathbb{R}$ is also continuous. Define the real valued map $f_{s_i,s_i'} : \mathcal{M}(S_{-i}) \to \mathbb{R}$ as

$$f_{s_i,s_i'}\left(\mathrm{marg}_{S_{-i}}\mu\right) = \pi_i(s_i, \mathrm{marg}_{S_{-i}}\mu) - \pi_i(s_i', \mathrm{marg}_{S_{-i}}\mu),\ \mu \in \mathcal{M}(\Omega).$$

The map $f_{s_i,s_i'}$ is clearly continuous, and the set $O^{\mathcal{W}}_{s_i,s_i'}$ can be written as

$$\begin{aligned} O^{\mathcal{W}}_{s_i,s_i'} &= \left(\widetilde{\mathrm{Proj}}_{S_{-i}}\right)^{-1} \left\{\mathrm{marg}_{S_{-i}}\mu \in \mathcal{M}(S_{-i}) \,\middle|\, f_{s_i,s_i'}\left(\mathrm{marg}_{S_{-i}}\mu\right) \geq 0 \right\} \\ &= \left(\widetilde{\mathrm{Proj}}_{S_{-i}}\right)^{-1} \left(f^{-1}_{s_i,s_i'}([0,+\infty))\right) \\ &= \left(f_{s_i,s_i'} \circ \widetilde{\mathrm{Proj}}_{S_{-i}}\right)^{-1} ([0,+\infty)), \end{aligned}$$

i.e., $O^{\mathcal{W}}_{s_i,s_i'}$ is the inverse image of the set $[0,+\infty)$, closed in $\mathbb{R}$, under the continuous map $f_{s_i,s_i'} \circ \widetilde{\mathrm{Proj}}_{S_{-i}}$, hence $O^{\mathcal{W}}_{s_i,s_i'}$ is closed in $\mathcal{M}(\Omega)$. An analogous argument shows that set $O^{\mathcal{S}}_{s_i,s_i'}$ can be written as

$$O^{\mathcal{S}}_{s_i,s_i'} = \left(f_{s_i,s_i'} \circ \widetilde{\mathrm{Proj}}_{S_{-i}}\right)^{-1} ((0,+\infty)),$$

hence $O^{\mathcal{S}}_{s_i,s_i'}$ is open in $\mathcal{M}(S_{-i} \times T_{-i})$. ■

Proof of Lemma C.3.[14] Let $U^{s_i}_n$ be the set of all $\overline{\mu} \in \mathcal{N}_n(S_{-i} \times T_{-i})$ for which s_i is a lexicographic best reply to $\overline{\mathrm{marg}}_{S_{-i}}\overline{\mu}$. By Lemma C.3, the sets $O^{\mathcal{W}}_{s_i,s_i'}$ and $O^{\mathcal{S}}_{s_i,s_i'}$ are, respectively, closed and open in $\mathcal{M}(\Omega)$, hence the set $O^{\mathcal{E}}_{s_i,s_i'} = O^{\mathcal{W}}_{s_i,s_i'} \backslash O^{\mathcal{S}}_{s_i,s_i'}$ is closed in $\mathcal{M}(S_{-i} \times T_{-i})$. The set $U^{s_i}_n$ can be expressed as

$$U^{s_i}_n = \bigcap_{s_i' \neq s_i} \left(\begin{array}{c} \left(O^{\mathcal{S}}_{s_i,s_i'} \times \mathcal{N}_{n-1}(\Omega)\right) \cup \left(O^{\mathcal{E}}_{s_i,s_i'} \times O^{\mathcal{S}}_{s_i,s_i'} \times \mathcal{N}_{n-2}(\Omega)\right) \cup \\ \ldots \cup \left(O^{\mathcal{E}}_{s_i,s_i'} \times O^{\mathcal{E}}_{s_i,s_i'} \times \ldots \times O^{\mathcal{W}}_{s_i,s_i'}\right) \end{array} \right),$$

[14]The proof closely follows the lines of the proof of Lemma A.6 in Dekel et al. (2016).

and this shows that $U_n^{s_i}$ is Borel in $\mathcal{N}(S_{-i} \times T_{-i})$. The set of all $\overline{\mu} \in \mathcal{N}(S_{-i} \times T_{-i})$ for which s_i is a lexicographic best reply to $\overline{\operatorname{marg}}_{S_{-i}}\overline{\mu}$ can be written as $\cup_{n\in\mathbb{N}} U_n^{s_i}$, hence it is Borel. ■

Corollary C.2 *Fix a type structure $\mathcal{T} = \langle S_i, T_i, \beta_i \rangle_{i\in I}$. For every $i \in I$, the set R_i^1 is Borel in $S_i \times T_i$.*

Proof: By Lemma C.2 and measurability of β_i, it follows that $C_i = S_i \times \beta_i^{-1}(\mathcal{C}_i)$ is Borel in $S_i \times T_i$. Write R_i as

$$R_i = \cup_{s_i \in S_i} \left(\{s_i\} \times \beta_i^{-1}(L^{s_i})\right),$$

where L^{s_i} stands for the set of all $\overline{\mu}_i \in \mathcal{N}(S_{-i} \times T_{-i})$ such that s_i is a lexicographic best reply to $\overline{\operatorname{marg}}_{S_{-i}}\overline{\mu}_i$. By Lemma C.3 and measurability of β_i, it follows that R_i is Borel in $S_i \times T_i$. Since $R_i^1 = R_i \cap C_i$ the conclusion follows. ■

We can now state and prove the desired result.

Lemma C.5 *For each $i \in I$ and $m \geq 1$, R_i^m is Borel in $S_i \times T_i$.*

Proof: For each $i \in I$, by Corollary C.2, the set R_i^1 is Borel in $S_i \times T_i$. By Corollary C.1, the set $\mathbf{WB}_i\left(R_{-i}^m\right)$ is Borel in $S_i \times T_i$ provided that R_{-i}^m is Borel. Since $R_i^{m+1} = R_i^m \cap \mathbf{WB}_i\left(R_{-i}^m\right)$, the conclusion follows from an easy induction on m. ■

For the proof of Lemma 1 we need two auxiliary results. The first (Lemma C.6) states that cautious rationality is preserved under type morphisms between type structures. The second (Lemma C.7) states an analogous result for the operator $\mathbf{WB}_i$, provided that the type morphisms are bimeasurable.

Lemma C.6 *Let $\mathcal{T} = \langle S_i, T_i, \beta_i \rangle_{i\in I}$ and $\mathcal{T}^* = \langle S_i, T_i^*, \beta_i^* \rangle_{i\in I}$ be type structures such that there exists a type morphism $(\varphi_i)_{i\in I} : T \to T^*$ from $\mathcal{T}$ to $\mathcal{T}^*$. Fix a type $t_i \in T_i$. Then*

(i) t_i is cautious if and only if $\varphi_i(t_i)$ is cautious.

(ii) A strategy-type pair (s_i, t_i) is rational in $\mathcal{T}$ if and only $(s_i, \varphi_i(t_i))$ is rational in $\mathcal{T}^*$.

Proof: Part (i): Let $\beta_i(t_i) = \left(\beta_i^1(t_i), ..., \beta_i^n(t_i)\right)$ be LPS associated with t_i. Fix $s_{-i} \in S_{-i}$. For each $l \leq n$, by definition of type morphism, we have

$$\beta_i^{*,l}(\varphi_i(t_i))\left(\{s_{-i}\} \times T_{-i}^*\right) = \beta_i^l(t_i)\left(\left(\mathrm{Id}_{S_{-i}}, \varphi_{-i}\right)^{-1}\left(\{s_{-i}\} \times T_{-i}^*\right)\right) = \beta_i^l(t_i)\left(\{s_{-i}\} \times T_{-i}\right).$$

Therefore there exists $l \leq n$ such that $\beta_i^l(t_i)(\{s_{-i}\} \times T_{-i}) > 0$ if and only if $\beta_i^{*,l}(\varphi_i(t_i))\left(\{s_{-i}\} \times T_{-i}^*\right) > 0$. It follows that $\overline{\operatorname{marg}}_{S_{-i}}\beta_i(t_i) \in \mathcal{N}^+(S_{-i})$ if and only if $\overline{\operatorname{marg}}_{S_{-i}}\beta_i^*(\varphi_i(t_i)) \in \mathcal{N}^+(S_{-i})$.

Part (ii) follows from the observation that type morphisms preserve the marginal LPS on strategies (i.e., first-order beliefs). ■

Lemma C.7 *Let $\mathcal{T} = \langle S_i, T_i, \beta_i \rangle_{i\in I}$ and $\mathcal{T}^* = \langle S_i, T_i^*, \beta_i^* \rangle_{i\in I}$ be type structures such that there exists a bimeasurable type morphism $(\varphi_i)_{i\in I} : T \to T^*$ from $\mathcal{T}$ to $\mathcal{T}^*$. Then, for each non-empty event $E_{-i} \subseteq S_{-i} \times T_{-i}$, it holds that $(\mathrm{Id}_{S_i}, \varphi_i)(\mathbf{WB}_i(E_{-i})) \subseteq \mathbf{WB}_i^*\left(\left(\mathrm{Id}_{S_{-i}}, \varphi_{-i}\right)(E_{-i})\right)$.*

Proof: Fix $(s_i, t_i) \in \mathbf{WB}_i(E_{-i})$, and write $\beta_i(t_i) = (\mu_1, ..., \mu_n)$ and $\beta_i^*(\varphi_i(t_i)) = (\mu_1^*, ..., \mu_n^*)$. By bimeasurability of $(\varphi_i)_{i\in I}$, the set $\left(\mathrm{Id}_{S_{-i}}, \varphi_{-i}\right)(E_{-i})$ is Borel in $S_{-i} \times T_{-i}^*$. Since E_{-i} is weakly believed under $\beta_i(t_i)$, we obtain

$$\begin{aligned}\mu_1^*(\left(\mathrm{Id}_{S_{-i}}, \varphi_{-i}\right)(E_{-i})) &= \mu_1(\left(\mathrm{Id}_{S_{-i}}, \varphi_{-i}\right)^{-1}\left(\mathrm{Id}_{S_{-i}}, \varphi_{-i}\right)(E_{-i})) \\ &\geq \mu_1(E_{-i}) = 1,\end{aligned}$$

where the first equality follows from the fact that $(\varphi_i)_{i\in I}$ is a type morphism (eq. (4.1)). Therefore, $\left(\mathrm{Id}_{S_{-i}}, \varphi_{-i}\right)(E_{-i})$ is weakly believed under $\beta_i^*(\varphi_i(t_i))$, i.e.,

$$(s_i, \varphi_i(t_i)) \in \mathbf{WB}_i^*\left(\left(\mathrm{Id}_{S_{-i}}, \varphi_{-i}\right)(E_{-i})\right).$$

■

Proof of Lemma 1: The proof of part (i) is by induction on m. For $m = 1$ the result follows from Lemma C.6. Suppose that the result holds for some $m \geq 1$; we show that it also holds for $m+1$. Fix a player $i \in I$, and let $(s_i, t_i) \in R_i^{m+1}$. So $(s_i, t_i) \in R_i^m$, and, by the induction hypothesis, $(s_i, \varphi_i(t_i)) \in R_i^{*,m}$. Hence, in order to show that $(s_i, \varphi_i(t_i)) \in R_i^{*,m+1}$, it is enough to show that $(s_i, \varphi_i(t_i)) \in \mathbf{WB}_i^*\left(R_{-i}^{*,m}\right)$, i.e., $R_{-i}^{*,m}$ is weakly believed under $\beta_i^*(\varphi_i(t_i))$. To this end, first note that, by the induction hypothesis, $\left(\mathrm{Id}_{S_{-i}}, \varphi_{-i}\right)\left(R_{-i}^m\right) \subseteq R_{-i}^{*,m}$. Moreover, by bimeasurability of $(\varphi_i)_{i\in I}$, the set $\left(\mathrm{Id}_{S_{-i}}, \varphi_{-i}\right)\left(R_{-i}^m\right)$ is Borel in $S_{-i} \times T_{-i}$. Therefore, Lemma C.7 yields that $\left(\mathrm{Id}_{S_{-i}}, \varphi_{-i}\right)\left(R_{-i}^m\right)$ is weakly believed under $\beta_i^*(\varphi_i(t_i))$. The conclusion of the proof of part (i) follows from the monotonicity property of weak belief (Property 3).

Part (ii) follows immediately from part (i). ■

Finally, we provide the proofs of the two lemmas of Section 4.3.

Proof of Lemma 2: The proof of part (i) is by induction on m.

$(m = 1)$ Fix $i \in I$. Let $s_i \in \mathrm{Proj}_{S_i}(R_i^1)$, so that $(s_i, t_i) \in R_i^1$ for some $t_i \in T_i$. Then s_i is optimal under $\overline{\mathrm{marg}}_{S_{-i}}\beta_i(t_i) \in \mathcal{N}^+(S_{-i})$, that is, $s_i \in S_i^1$. So $\mathrm{Proj}_{S_i}(R_i^1) \subseteq S_i^1$ for each $i \in I$.

$(m \geq 2)$ Suppose that the statement has been shown to hold for all $l = 1, ..., m-1$. We show that the statement is true for $l = m$.

Fix a player $i \in I$, and let $s_i \in \mathrm{Proj}_{S_i}(R_i^m)$, so that $(s_i, t_i) \in R_i^m$ for some $t_i \in T_i$. It follows from the definition of R_i^m that $(s_i, t_i) \in R_i^{m-1}$, so, by the induction hypothesis, $s_i \in S_i^{m-1}$. Also, R_{-i}^{m-1} is weakly believed under $\beta_i(t_i) = (\mu_i^1, ..., \mu_i^n)$, hence

$$\mathrm{marg}_{S_{-i}}\mu_i^1\left(S_{-i}^{m-1}\right) \geq \mathrm{marg}_{S_{-i}}\mu_i^1\left(\mathrm{Proj}_{S_{-i}}\left(R_{-i}^{m-1}\right)\right) \geq \mu_i^1\left(R_{-i}^{m-1}\right) = 1,$$

where the first inequality follows from the induction hypothesis. Hence $s_i \in S_i^{m-1}$ is optimal under $\overline{\mathrm{marg}}_{S_{-i}}\beta_i(t_i) \in \mathcal{N}^+(S_{-i})$ with $\mathrm{marg}_{S_{-i}}\mu_i^1\left(S_{-i}^{m-1}\right) = 1$, that is, $s_i \in S_i^m$. So $\mathrm{Proj}_{S_i}(R_i^m) \subseteq S_i^m$ for each $i \in I$.

This concludes the proof of part (i). Part (ii) follows immediately from part (i). ■

Proof of Lemma 3: If $Q \in \mathcal{Q}$ is a full WBRS, then, for each $s_i \in Q_i$, there exists $\overline{\nu}_{s_i} \in \mathcal{W}^+(Q_{-i}) \subseteq \mathcal{N}^+(S_{-i})$ such that $s_i \in r_i(\overline{\nu}_{s_i})$ and $r_i(\overline{\nu}_{s_i}) \subseteq Q_i$. So, we fix some $\overline{\nu}_{s_i}$ satisfying the above conditions for every $s_i \in Q_i$, and we construct a finite type structure

$\mathcal{T}^* = \langle S_i, T_i^*, \beta_i^* \rangle_{i \in I}$ as follows. For each player $i \in I$, let T_i^* be a homeomorphic copy of Q_i. So, for every $s_i \in Q_i$, we will denote the corresponding type as $t_{s_i}^*$. For each $i \in I$, let $\psi_i : Q_i \to (S_i \times T_i^*)$ be the continuous map that associates each strategy $s_i \in Q_i$ with the pair $(s_i, t_{s_i}^*)$. Thus, we define each belief map $\beta_i^* : T_i^* \to \mathcal{N}(S_{-i} \times T_{-i}^*)$ by

$$\beta_i^*(t_{s_i}^*) = \widehat{\psi}_{-i}(\overline{\nu}_{s_i}), \; s_i \in Q_i.$$

Finiteness of each type set T_i^* guarantees that each belief map is measurable (in fact, continuous). Note that, for each $s_i \in Q_i$, $\overline{\mathrm{marg}}_{S_{-i}} \beta_i^*(t_{s_i}^*) = \overline{\nu}_{s_i}$.

We show that $\mathcal{T}^*$ satisfies the required properties. We prove, by induction on $m \geq 1$, that the following properties hold for each $i \in I$:

(a) if $s_i \in Q_i$ then $(s_i, t_{s_i}^*) \in R_i^{*,m}$; and

(b) if $(s_i', t_{s_i}^*) \in R_i^{*,m}$ then $s_i' \in Q_i$.

Part (a) yields $Q_i \subseteq \mathrm{Proj}_{S_i}(R_i^{*,\infty})$, while part (b) yields $\mathrm{Proj}_{S_i}(R_i^{*,\infty}) \subseteq Q_i$.

($m = 1$) If $s_i \in Q_i$, then $(s_i, t_{s_i}^*) \in R_i^{*,1}$ because s_i is optimal under $\overline{\mathrm{marg}}_{S_{-i}} \beta_i^*(t_{s_i}^*) \in \mathcal{N}^+(S_{-i})$. For property (b): If $(s_i', t_{s_i}^*) \in R_i^{*,1}$, then s_i' is optimal under $\overline{\mathrm{marg}}_{S_{-i}} \beta_i^*(t_{s_i}^*) \in \mathcal{N}^+(S_{-i})$, and, since Q is a full WBRS, then $s_i' \in Q_i$.

($m \geq 2$) Assume that property (a) holds for $m \geq 1$. Equivalently, property (a) says that $\psi_i(Q_i) \subseteq R_i^{*,m}$ for each $i \in I$, and so

$$\forall i \in I, \; Q_i \subseteq \psi_i^{-1}\left(R_i^{*,m}\right). \tag{C.1}$$

Let $s_i \in Q_i$. Then, by the induction hypothesis, $(s_i, t_{s_i}^*) \in R_i^{*,m}$, so we need to show that $R_{-i}^{*,m}$ is weakly believed under $\beta_i^*(t_{s_i}^*)$, i.e., the first component probability of $\beta_i^*(t_{s_i}^*)$, viz. $\widetilde{\psi}_{-i}(\nu_{s_i}^1)$, assigns probability 1 to $R_{-i}^{*,m}$. To see this, note that

$$\widetilde{\psi}_{-i}(\nu_{s_i}^1)\left(R_{-i}^{*,m}\right) = \nu_{s_i}^1\left(\psi_{-i}^{-1}\left(R_{-i}^{*,m}\right)\right) \geq \nu_{s_i}^1(Q_{-i}) = 1,$$

where the second inequality follows from (C.1). Therefore $(s_i, t_{s_i}^*) \in R_i^{*,m+1}$.

Assume now that property (b) holds for $m \geq 1$. If $(s_i', t_{s_i}^*) \in R_i^{*,m+1}$ then $(s_i', t_{s_i}^*) \in R_i^{*,m}$, and so, by the induction hypothesis, $s_i' \in Q_i$. ■

References

[1] Asheim, G., and M. Dufwenberg (2003): "Admissibility and common belief," *Games and Economic Behavior*, 42, 208–234.

[2] Battigalli, P., and A. Friedenberg (2012): "Context-Dependent Forward Induction Reasoning," IGIER Working Paper 351, Bocconi University.

[3] Battigalli, P., and M. Siniscalchi (1999): "Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games," *Journal of Economic Theory*, 88, 188-230.

[4] Blume, L., A. Brandenburger, and E. Dekel (1991): "Lexicographic Probabilities and Choice Under Uncertainty," *Econometrica,* 59, 61-79.

[5] Blume, L., A. Brandenburger, and E. Dekel (1991): "Lexicographic Probabilities and Equilibrium Refinements," *Econometrica,* 59, 81-98.

[6] Brandenburger, A. (1992): "Lexicographic Probabilities and Iterated Admissibility," in *Economic Analysis of Markets and Games*, ed. by P. Dasgupta, D. Gale, O. Hart, and E. Maskin. Cambridge MA: MIT Press, 282-290.

[7] Brandenburger, A. (2003): "On the Existence of a 'Complete' Possibility Structure," in *Cognitive Processes and Economic Behavior*, ed. by M. Basili, N. Dimitri and I. Gilboa. New York: Routledge, 30-34.

[8] Brandenburger, A., and A. Friedenberg (2010): "Self-Admissible Sets," *Journal of Economic Theory*, 145, 785-811.

[9] Brandenburger, A., A. Friedenberg, and H.J. Keisler (2008): "Admissibility in Games," *Econometrica*, 76, 307-352.

[10] Brandenburger, A., A. Friedenberg, and H.J. Keisler (2012): "Fixed Points in Epistemic Game Theory," in *Proceedings of the 2008 Clifford Lectures, AMS Proceedings of Symposia in Applied Mathematics.*

[11] Borgers, T. (1994): "Weak Dominance and Approximate Common Knowledge," *Journal of Economic Theory*, 64, 265-276.

[12] Catonini, E., and N. De Vito (2016a): "Hierarchies of Lexicographic Beliefs," working paper.

[13] Catonini, E., and N. De Vito (2018): "Cautious Belief and Iterated Admissibility," working paper.

[14] Dekel, E., and D. Fudenberg (1990): "Rational Behavior with Payoff Uncertainty," *Journal of Economic Theory,* 52, 243-67.

[15] Dekel, E., and M. Siniscalchi (2015): "Epistemic Game Theory," in *Handbook of Game Theory with Economic Applications, Volume 4*, ed. by P. Young and S. Zamir. Amsterdam: North-Holland, 619-702.

[16] Dekel, E., A. Friedenberg, and M. Siniscalchi (2016): "Lexicographic Beliefs and Assumption," *Journal of Economic Theory*, 163, 955-985.

[17] Friedenberg, A. (2010): "When Do Type Structures Contain All Hierarchies of Beliefs?," *Games and Economic Behavior*, 68, 108-129.

[18] Friedenberg, A., and H.J. Keisler (2011): "Iterated Dominance Revisited," working paper.

[19] Halpern, J.Y. (2010): "Lexicographic Probability, Conditional Probability, and Nonstandard Probability," *Games and Economic Behavior*, 68, 155-179.

[20] Heifetz, A., and D. Samet (1998): "Topology-Free Typology of Beliefs," *Journal of Economic Theory*, 82, 324-341.

[21] Hu, T. (2007): "On p-rationalizability and Approximate Common Certainty of Rationality," *Journal of Economic Theory*, 136, 379-391.

[22] Kechris, A. (1995): *Classical Descriptive Set Theory.* Berlin: Springer Verlag.

[23] Lo, K.C. (1999): "Nash Equilibrium without Mutual Knowledge of Rationality," *Economic Theory*, 14, 621-633.

[24] Mertens, J.F., and S. Zamir (1985): "Formulation of Bayesian Analysis for Games With Incomplete Information," *International Journal of Game Theory*, 14, 1-29.

[25] Milgrom, P. and J. Roberts (1991): "Adaptive and Sophisticated Learning in Normal-Form Games," *Games and Economic Behavior*, 3, 82-100.

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

[27] Perea A. (2012): *Epistemic Game Theory: Reasoning and Choice*, CUP Press.

[28] Savage, L. (1972): *The Foundations of Statistics*. New York: Wiley.

[29] Stahl, D.O. (1995): "Lexicographic Rationalizability and Iterated Admissibility," *Economic Letters*, 47, 155-159.