# Rational Behavior under Correlated Uncertainty*

Mira Frick Assaf Romm

November 4, 2014

**Abstract**

In complete information games, Dekel and Fudenberg (1990) and Börgers (1994) have proposed the solution concept $S^\infty W$ (one round of elimination of weakly dominated strategies followed by iterated elimination of strongly dominated strategies), motivating it by a characterization in terms of "approximate common knowledge" of admissibility. We examine the validity of this characterization of $S^\infty W$ in an *incomplete* information setting. We argue that in Bayesian games with a nontrivial state space, the characterization is very sensitive to the way in which uncertainty in the form of approximate common knowledge of admissibility is taken to interact with the uncertainty already captured by players' beliefs about the states of nature: We show that $S^\infty W$ corresponds to approximate common knowledge of admissibility when this is not allowed to coincide with any changes to players' beliefs about states. If approximate common knowledge of admissibility is accompanied by vanishingly small perturbations to beliefs, then $S^\infty W$ is a (generally strict) subset of the predicted behavior, which we characterize in terms of a generalization of Hu's (2007) perfect $p$-rationalizable set.

*JEL classification:* C70, C72.

*Keywords:* Rationality; admissibility; approximate common knowledge; common $p$-belief; incomplete information; robustness.

# 1 Introduction

In complete information games, Dekel and Fudenberg (1990) and Börgers (1994) have proposed the solution concept $S^\infty W$ (one round of elimination of weakly dominated strategies followed by iterated elimination of strongly dominated strategies), motivating it via its connection with "approximate common knowledge" of admissibility. Admissibility (expected

*Department of Economics, Harvard University and HBS. 1805 Cambridge St, Cambridge, MA 02138. Tel: 510 388 0275. Emails: mfrick@fas.harvard.edu; aromm@hbs.edu. We are extremely grateful to Drew Fudenberg for generous advice and guidance. For helpful discussions, we thank Elchanan Ben-Porath, Eddie Dekel, Sergiu Hart, Ryota Iijima, Yuhta Ishii, Déborah Marciano-Romm, Eric Maskin, Morgan McClellon, Stephen Morris, and Tomasz Strzalecki. We are indebted to two anonymous referees and the co-editor, Marciano Siniscalchi, for thorough and insightful comments which greatly improved the paper.

utility maximization with respect to some full-support conjecture about opponents' behavior) and iterated admissibility are commonly used refinements of Bayesian rationality (e.g. Luce and Raiffa, 1957; Kohlberg and Mertens, 1986).

Börgers' interest in approximate common knowledge of admissibility is driven by epistemic considerations, namely the aim to establish an analog of Tan and da Costa Werlang's (1988) well-known result that the behavioral implications of common knowledge of rationality are given by $S^\infty$ (iterated strong dominance). There is a logical tension between admissibility (holding full-support beliefs about opponents' behavior) and common knowledge of admissibility (which in general rules out some opponent strategies). This tension disappears when common knowledge is relaxed to approximate common knowledge in the sense of common $p$-belief for $p$ close to 1. Börgers (formalized by Hu (2007)) shows that $S^\infty W$ encapsulates the behavioral implications of the latter notion. Dekel and Fudenberg are motivated by robustness considerations, focusing on the special case where approximate common knowledge of admissibility is the result of small amounts of payoff uncertainty. They ask which strategies can arise if players behave according to iterated admissibility, but there is vanishingly small payoff uncertainty, which they model via sequences of elaborations converging to a game. Once again the answer to this question is given by $S^\infty W$.

This paper examines the connection between approximate common knowledge of admissibility and $S^\infty W$ in an *incomplete* information setting. Consider a Bayesian game $G$ with state space $\Theta$ in which each player $i$ has first-order belief $\phi_i$ over $\Theta$. We obtain extensions of Börger's and Dekel and Fudenberg's characterizations of $S^\infty W$, but show that these are very sensitive to the way in which uncertainty in the form of approximate common knowledge of admissibility is taken to interact with the uncertainty (represented by each player $i$'s belief $\phi_i$ on $\Theta$) that is already present in $G$.

Interpreting $S^\infty W$ in the interim-correlated sense of Dekel et al. (2007), Section 3.1 extends Börgers' characterization: We show that if there is common $p$-belief of admissibility *and* of the fact that each player $i$'s first-order belief over $\Theta$ is *exactly* $\phi_i$, then for $p$ close enough to 1, $S^\infty W$ once again emerges as the set of behavioral implications (Theorem 3.1). In Appendix A, we provide an analogous extension of Dekel and Fudenberg's result: Proposition A.2 shows that $S^\infty W$ is the robust extension of $W^\infty$ under elaborations in which "sane" types may assign vanishingly small probability to opponents being "crazy" (i.e. having very different payoffs and beliefs as in the original game), but must themselves have *exactly* the same beliefs (and payoffs) as in the original game.

However, these results break down when approximate common knowledge of admissibility is accompanied by vanishingly small perturbations to players' beliefs about states. In Section 3.2 we show that if there is common $p$-belief of admissibility and of the fact that each player $i$'s first-order belief about $\Theta$ is "approximately" $\phi_i$, then the behavioral implications are given by a generalization of Hu's (2007) perfect $p$-rationalizable set (Theorem 3.5). But even in the limit as $p$ goes to 1 and as the uncertainty about $\phi_i$ becomes vanishingly small, this set is in general a strict superset of $S^\infty W$.[1]

In addition to Börgers (1994), Dekel and Fudenberg (1990), and Hu (2007), our paper connects more broadly with the growing literatures on epistemic conditions related to

[1]Similarly, under elaborations in which sane types' beliefs over $\Theta$ are allowed to be vanishingly small perturbations of the beliefs in $G$, $S^\infty W$ is in general a strict subset of the robust closure of $W^\infty$.

admissibility and on the robustness of solution concepts to small amounts of uncertainty. Within the former literature, an alternative epistemic characterization of $S^\infty W$ is obtained by Schuhmacher (1999): Instead of relaxing common knowledge to common $p$-belief, he relaxes the rationality requirement implied by admissibility to "$\varepsilon$-rationality". Brandenburger et al. (2008) replace probabilistic beliefs with lexicographic probability systems to obtain epistemic foundations for iterated admissibility.[2] Both these papers are set in a complete information environment. Within the latter literature, which was pioneered by Fudenberg et al. (1988) and Kajii and Morris (1997), our paper relates most closely to Weinstein and Yildiz (2007): They show that in incomplete information games satisfying a richness assumption, any action in $S^\infty$ can be made *uniquely* rationalizable if we allow for *arbitrary* perturbations to players' entire hierarchy of beliefs. Put differently, this means that any refinement of $S^\infty$ must be derived from restrictions on the class of perturbations that we are considering. One of the contributions of our paper lies in clarifying the nature of restrictions needed to obtain $S^\infty W$.

The rest of the paper is organized as follows. Section 2 introduces notation and definitions. Section 3 contains our main results. Section 4 concludes. Appendix A extends Dekel and Fudenberg's characterization to our setting, and Appendix B contains all proofs.

# 2 Preliminaries

Throughout the paper we fix an incomplete information normal-form game $G = (I, \Theta, (A_i, u_i, \phi_i)_{i\in I})$, where $\Theta$ is a finite set of states of nature, $I$ is a finite set of players, and for every $i \in I$, $A_i$ is a finite set of strategies, $u_i\colon A_i \times A_{-i} \times \Theta \to \mathbb{R}$ denotes $i$'s state-dependent payoffs, and $\phi_i \in \Delta^\circ(\Theta)$ is $i$'s full-support belief about $\Theta$.[3] We work with the following notions of admissibility, iterated dominance, and approximate common knowledge of admissibility:

## 2.1 Interim Correlated Dominance

We interpret strong and weak dominance in the "interim correlated" sense of Dekel et al. (2007).[4][5] This allows any player's conjectures about her opponents' behavior to be correlated with her beliefs about the state of nature. As argued by Dekel et al. (2007), allowing for this kind of correlation is the natural approach in Bayesian games, because imposing independence on conjectures about states and opponents' behavior produces solution concepts which are very sensitive to "redundant" aspects of the type space.[6] Nevertheless, as we discuss in Section 4, imposing the latter kind of independence on our definition of iterated dominance would not fundamentally alter our results in Section 3.

[2]See Dekel and Siniscalchi (2014) for a more comprehensive discussion of this literature.

[3]For expositional clarity, we assume that in $G$ there is a single type for each player. With slight adjustments, our results extend readily to games with finite type spaces.

[4]See also Battigalli and Siniscalchi (2003).

[5]Our notion of interim correlated weak dominance, which does not appear in Dekel et al. (2007), is derived from their concept of strong dominance in the natural way.

[6]See Dekel et al. (2007) for details and examples.

**Definition 2.1.** Consider non-empty subsets $\hat{A}_j \subseteq A_j$ for every $j$, and let $\hat{A} := \prod_j \hat{A}$ and $\hat{A}_{-j} := \prod_{i \neq j} \hat{A}_i$. We say that $\alpha_i \in \Delta(\hat{A}_j)$ **strongly dominates** $a_i \in \hat{A}_i$ on $\hat{A}$ if for all beliefs $\mu_i \colon \Theta \to \Delta(\hat{A}_{-i})$,

$$\sum_{\theta, a_{-i}} \phi_i(\theta)\mu_i(\theta)[a_{-i}](u_i(\alpha_i, a_{-i}, \theta) - u_i(a_i, a_{-i}, \theta)) > 0.$$

$\alpha_i \in \Delta(\hat{A}_i)$ **weakly dominates** $a_i \in \hat{A}_i$ on $\hat{A}$ if for all beliefs $\mu_i \colon \Theta \to \Delta(\hat{A}_{-i})$,

$$\sum_{\theta, a_{-i}} \phi_i(\theta)\mu_i(\theta)[a_{-i}](u_i(\alpha_i, a_{-i}, \theta) - u_i(a_i, a_{-i}, \theta)) \geq 0,$$

with strict inequality for at least one such $\mu_i$.

Denote by $S(\hat{A})_i$ the set of all strategies $a_i \in \hat{A}_i$ which are not strongly dominated on $\hat{A}$ by any $\alpha_i \in \Delta(\hat{A}_i)$, and let $S(\hat{A}) := \prod_j S(\hat{A})_j$. Let $S^1(\hat{A}) := S(\hat{A})$, $S^{k+1}(\hat{A}) := S(S^k(\hat{A}))$ and $S^\infty(\hat{A}) := \cap_k S^k(\hat{A})$. Similarly, denote by $W(\hat{A})_i$ the set of all **admissible** strategies for player $i$ on $\hat{A}$, i.e. all $a_i \in \hat{A}_i$ which are not weakly dominated on $\hat{A}$ by any $\alpha_i \in \Delta(\hat{A}_i)$, and let $W(\hat{A}) := \prod_j W(\hat{A})_j$. Let $W^1(\hat{A}) := W(\hat{A})$, $W^{k+1}(\hat{A}) := W(W^k(\hat{A}))$ and $W^\infty(\hat{A}) := \cap_k W^k(\hat{A})$. If $\hat{A}_j = A_j$ for all $j$, we write $S^k(G)$, $W^k(G)$ and $S^l W^k(G)$ for $S^k(\hat{A})$, $W^k(\hat{A})$ and $S^l W^k(\hat{A})$, respectively.

The following lemma is an incomplete-information extension of Pearce's well-known result (Pearce, 1984): A strategy is not strongly dominated if and only if it is a best response to some state-dependent belief about opponent strategies, and it is admissible if and only if it is a best response to a state-dependent belief having *full support* on the relevant set of opponent strategies in every state:

**Lemma 2.2** (Equivalence of undominance and best response formulations)**.** *Let $\hat{A}_i$ and $\hat{A}_{-i}$ be nonempty subsets of $A_i$ and $A_{-i}$, respectively. Suppose $a_i \in \hat{A}_i$. Then:*

(i). *$a_i \in S(\hat{A})_i$ if and only if $a_i$ is a best response in $\hat{A}_i$ to some $\lambda_i \colon \Theta \to \Delta(\hat{A}_{-i})$.*[7]

(ii). *$a_i \in W(\hat{A})_i$ if and only if $a_i$ is a best response in $\hat{A}_i$ to some $\lambda_i \colon \Theta \to \Delta^\circ(\hat{A}_{-i})$.*

## 2.2 Approximate Common Knowledge of Admissibility

To model approximate common knowledge of admissibility, we associate with $G$ an epistemic type structure: We let $\mathcal{T} = \left(I, (\Theta \times A_{-i}, T_i, \beta_i)_{i \in I}\right)$ be the universal type space in which each player $i$'s basic space of uncertainty is $\Theta \times A_{-i}$.[8] Then each set of types $T_i$ is a compact metric space and for each $i$, the belief $\beta_i \colon T_i \to \Delta\left(\Theta \times A_{-i} \times T_{-i}\right)$ is a homeomorphism.[9] We will use $\mathcal{T}$ to model small amounts of uncertainty about the fact that players choose admissible

[7] I.e., $a_i \in \operatorname{argmax}_{a'_i \in \hat{A}_i} \sum_{\theta, a_{-i}} \phi_i(\theta)\lambda_i(\theta)[a_{-i}]u_i(a'_i, a_{-i}, \theta)$.

[8] Cf. Brandenburger and Dekel (1993) for details regarding the construction of this space.

[9] For any compact metric space $M$, we equip the set $\Delta(M)$ of probability measures on the Borel $\sigma$-algebra of $M$ with the weak* topology, under which $\Delta(M)$ is itself compact metric.

strategies. In our incomplete information setting, it is natural to allow this uncertainty to interact with the uncertainty already present in $G$: Player $i$'s uncertainty about opponents' behavior might be accompanied by small doubts about opponents' beliefs about $\Theta$ and might be correlated with $i$'s own beliefs about $\Theta$. This is captured in $\mathcal{T}$ by the fact that each $\beta_i(t_i)$ is a joint probability distribution on $\Theta \times A_{-i} \times T_{-i}$ and that $\text{marg}_\Theta \beta_j(t_j)$ is not necessarily equal to $\phi_j$. Let $f_i := \text{marg}_{\Theta \times A_{-i}} \beta_i$, let $\Omega := \Theta \times \Pi_{i \in I}(A_i \times T_i)$, and denote by $\theta$, $a_i$ and $t_i$ the projections from $\Omega$ onto $\Theta$, $A_i$ and $T_i$, respectively.

**Common $p$-belief:** For any event $E \subseteq \Omega$,[10] and any $\omega \in \Omega$, let $E^\omega := \{(\theta, a_{-i}, t_{-i}) \in \Theta \times A_{-i} \times T_{-i} : (\theta, a_i(\omega), t_i(\omega), a_{-i}, t_{-i}) \in E\}$. For any $p \in (0, 1]$, we define the following events:

- $i$ $p$-believes $E$: $B_i^p(E) := \{\omega \in \Omega : \beta_i(t_i(\omega))(E^\omega) \geq p\}$
- mutual $p$-belief of $E$: $B^p(E) := \bigcap_{i \in I} B_i^p(E)$
- $B^{p,1}(E) := B^p(E)$ and inductively $B^{p,n+1}(E) := B^p(B^{p,n}(E))$ for $n \geq 1$
- common $p$-belief of $E$: $\text{CB}^p(E) := \bigcap_{n \in \mathbb{N}} B^{p,n}(E)$.

**Approximate common knowledge (ACK) of admissibility:** Define the event that **player $i$ is rational** given utility $u_i$[11] by

$$\mathcal{R}_i := \left\{ \omega \in \Omega : a_i(\omega) \in \text{argmax}_{a_i' \in A_i} \sum_{\theta \in \Theta, a_{-i} \in A_{-i}} f_i(t_i(\omega))(\theta, a_{-i}) u_i(a_i', a_{-i}, \theta) \right\},$$

and the event that all players are rational by $\mathcal{R} := \bigcap_{i \in I} \mathcal{R}_i$. Define the event that **player $i$'s first-order belief has full support** by $\mathcal{P}_i := \{\omega \in \Omega : f_i(t_i(\omega)) \in \Delta^\circ(\Theta \times A_{-i})\}$, and that all players have full-support first-order beliefs by $\mathcal{P} := \bigcap_{i \in I} \mathcal{P}_i$. Then $\mathcal{R} \cap \mathcal{P}$ is the event that all players play admissible strategies.

In the complete information setting where $|\Theta| = 1$, Börgers (1994) and Hu (2007) model approximate common knowledge (henceforth ACK) of admissibility by $CB^p(\mathcal{R} \cap \mathcal{P})$ as $p \to 1$, capturing the idea that players' uncertainty (and higher-order uncertainty) about opponents' choice of admissible strategies becomes vanishingly small. In the case where $|\Theta| > 1$, we allow player's doubts about opponents' behavior to be accompanied by doubts about opponents' beliefs about $\Theta$ (in the sense that we do not impose that $\text{marg}_\Theta \beta_j(t_j) = \phi_j$). Correspondingly, a natural definition of ACK should require both types of doubts to vanish.

We consider two ways of modeling this: Define the event that **player $i$'s first-order beliefs on $\Theta$ are $\phi_i$** by $[\phi_i] := \{\omega \in \Omega : \text{marg}_\Theta \beta_i(t_i(\omega)) = \phi_i\}$, and let $[\phi] := \bigcap_{i \in I} [\phi_i]$. For any $\varepsilon > 0$, define the event that **player $i$'s first-order beliefs on $\Theta$ are $\varepsilon$-close to $\phi_i$** by $[\phi_i, \varepsilon] := \{\omega \in \Omega : ||\text{marg}_\Theta \beta_i(t_i(\omega)) - \phi_i||_\infty \leq \varepsilon\}$, and let $[\phi, \varepsilon] := \bigcap_{i \in I} [\phi_i, \varepsilon]$.

In Section 3.1 we consider the behavioral implications of **strong ACK of admissibility**, defined as $\text{SACKA}(G) := \bigcap_{p \in (0,1)} \text{Proj}_A CB^p([\phi] \cap \mathcal{R} \cap \mathcal{P})$, which imposes common $p$-belief

[10] An event is a measurable set w.r.t. the Borel $\sigma$-algebra on $\Omega$.

[11] In Section 4 we briefly discuss the implications of allowing for perturbations of the utility.

of the *exact* profile of priors $\phi$. Section 3.2 considers the behavioral implications of **ACK of admissibility with perturbed priors**, defined as

$$\mathrm{PACKA}(G) := \lim_{\bar{p}\to 1^-,\bar{\varepsilon}\to 0^+} \bigcap_{p\in[\bar{p},1),\varepsilon\in(0,\bar{\varepsilon}]} \mathrm{Proj}_A\, CB^p([\phi,\varepsilon]\cap\mathcal{R}\cap\mathcal{P}),^{12}$$

which only imposes common $p$-belief of the fact that priors on $\Theta$ are perturbations of $\phi$, where we require these perturbations to vanish in the limit as $p \to 1$.

**Conditional $p$-belief:** In the following, we will also need to consider the event that *in every state* $\theta$ there is common $p$-belief in admissibility. For this we introduce the notion of *conditional* $p$-belief: For any family $\mathcal{F} := \{F_1, \ldots, F_N\}$ of events, define the following events:

- $i$ $p$-believes $E$ conditional on $F_k$:

$$B_i^p(E \mid F_k) := \left\{\omega \in \Omega \;\middle|\; \beta_i(t_i(\omega))(F_k^\omega) > 0 \text{ and } \frac{\beta_i(t_i(\omega))(E^\omega \cap F_k^\omega)}{\beta_i(t_i(\omega))(F_k^\omega)} \geq p\right\}$$

- $i$ $p$-believes $E$ conditional on each $F_k$: $B_i^p(E \mid \mathcal{F}) := \bigcap_{n\in\{1,\ldots,N\}} B_i^p(E \mid F_k)$
- mutual $p$-belief conditional on each $F_k$: $B^p(E \mid \mathcal{F}) := \bigcap_{i\in I} B_i^p(E \mid \mathcal{F})$
- $B^{p,1}(E \mid \mathcal{F}) := B^p(E \mid \mathcal{F})$ and $B^{p,n+1}(E \mid \mathcal{F}) := B^p(B^{p,n}(E \mid \mathcal{F}) \mid \mathcal{F})$ for all $n \geq 1$
- common $p$-belief conditional on each $F_k$: $\mathrm{CB}^p(E \mid \mathcal{F}) := \bigcap_{n\in\mathbb{N}} B^{p,n}(E \mid \mathcal{F})$.

Letting $E(\theta) := \{\omega \in \Omega \mid \theta(\omega) = \theta\}$ and $E(\Theta) := \{E(\theta)\}_{\theta\in\Theta}$, the event that $i$ $p$-believes $E$ in every state $\theta$ is then given by $B_i^p(E \mid E(\Theta))$. We will make use of the following relationship between conditional and unconditional $p$-beliefs:

**Lemma 2.3.** *For any $i \in I$, $p \in (0,1]$, partition $\mathcal{F} = \{F_1, \ldots, F_N\}$ of $\Omega$, and $\theta \in \Theta$:*

- *$B_i^p(E \mid \mathcal{F}) \subseteq B_i^p(E)$; and*
- *$B_i^p(E) \cap [\phi_i] \subseteq B_i^{\max\left\{1-\frac{1-p}{\phi_i(\theta)},0\right\}}(E \mid E(\theta))$.*

# 3 $S^\infty W$ and ACK of Admissibility

In a complete information setting, Börgers (1994) (formalized by Hu (2007)) shows that for large enough $p$, we have that $S^\infty W(G) = \mathrm{Proj}_A\, CB^p(\mathcal{R}\cap\mathcal{P})$. In this section we examine the validity of this result under the two incomplete-information notions of ACK of admissibility defined in Section 2.2.

[12] Note that whenever $0 < p \leq p' < 1$ and $\varepsilon \geq \varepsilon' > 0$, then $\mathrm{Proj}_A\, CB^p([\phi,\varepsilon]\cap\mathcal{R}\cap\mathcal{P}) \supseteq \mathrm{Proj}_A\, CB^{p'}([\phi,\varepsilon']\cap\mathcal{R}\cap\mathcal{P})$. Thus, by finiteness of $A$ there is $\bar{p} \in (0,1)$ and $\bar{\varepsilon} > 0$ such that $\mathrm{PACKA}(G) = \mathrm{Proj}_A\, CB^{\bar{p}}([\phi,\bar{\varepsilon}]\cap\mathcal{R}\cap\mathcal{P}) = \mathrm{Proj}_A\, CB^p([\phi,\varepsilon]\cap\mathcal{R}\cap\mathcal{P})$ for all $(p,\varepsilon) \in [\bar{p},1) \times (0,\bar{\varepsilon}]$.

## 3.1 Strong ACK of Admissibility

Consider first the behavioral implications of *strong* ACK of admissibility, as defined by the set SACKA($G$) in Section 2.2. In this case, we obtain an incomplete information extension of Börgers's (1994) characterization:

**Theorem 3.1.** *There exists $\bar{p} \in (0,1)$ such that for all $p \in [\bar{p}, 1)$ we have*

$$S^\infty W(G) = \mathrm{Proj}_A \, \mathrm{CB}^p \left([\phi] \cap \mathcal{R} \cap \mathcal{P}\right) = \mathrm{SACKA}(G). \tag{1}$$

The proof of Theorem 3.1 proceeds in two steps, both of which make use of an incomplete information analog of Hu's (2007) perfect $p$-rationalizable set: For any $A' := \prod_{i \in i} A'_i \subseteq A$ and $i \in I$, let

$$D_i^p(A') := \left\{ a_i \in A'_i : \exists \mu \colon \Theta \to \Delta^\circ(A_{-i}) \text{ s.t. } a_i \in \mathrm{BR}_{A'_i}^{\phi_i}(\mu) \text{ and } \mu(\theta)(A'_{-i}) \geq p \, \forall \theta \in \Theta \right\},$$

where $BR_{A'_i}^{\phi_i}(\mu) := \mathrm{argmax}_{a'_i \in A'_i} \sum_{\theta \in \Theta, a_{-i} \in A_{-i}} \phi_i(\theta)\mu(\theta)(a_{-i}) u_i(a'_i, a_{-i}, \theta)$. Inductively define $\tilde{\Lambda}^{p,0}(A) = D^p(A) := \prod_{i \in I} D_i^p(A)$ and $\tilde{\Lambda}^{p,n+1} := D^p(A_i, \tilde{\Lambda}_{-i}^{p,n}(A))$ for all $n \geq 0$. Then the **perfect $p$-rationalizable set** of $G$ is given by $\tilde{R}^p(G) := \bigcap_{n \in \mathbb{N}} \tilde{\Lambda}^{p,n}(A)$.

The first step in the proof of Theorem 3.1 shows that for large enough $p$, the perfect $p$-rationalizable set $\tilde{R}^p(G)$ is precisely $S^\infty W(G)$. This follows from an extension of the key lemma from Börgers (1994) to our incomplete information setting:

**Lemma 3.2.** *There exists $\pi \in (0,1)$ such that for all $i \in I$, $p \in [\pi, 1)$, and non-empty $\hat{A}_i \subseteq A_i$ and $\hat{A}_{-i} \subseteq A_{-i}$, we have that $a_i \in \hat{A}_i$ is in $D_i^p(\hat{A}_i, \hat{A}_{-i})$ if and only if the following two conditions hold:*

(i). *$a_i \in W(\hat{A}_i \times A_{-i})_i$, i.e. there is no $\alpha_i \in \Delta(\hat{A}_i)$ which weakly dominates $a_i$ on $A_{-i}$.*

(ii). *$a_i \in S(\hat{A}_i \times \hat{A}_{-i})_i$, i.e. there is no $\alpha_i \in \Delta(\hat{A}_i)$ which strongly dominates $a_i$ on $\hat{A}_{-i}$.*

**Corollary 3.3.** *There exists $\pi \in (0,1)$ such that for all $i \in I$ and $p \in [\pi, 1)$, we have $\tilde{R}_i^p(G) = S^\infty W(G)_i$.*

The second step in proving Theorem 3.1 is an extension of Theorem 5.1 from Hu (2007) to our incomplete information setting. We show that for any $p$, $\tilde{R}^p(G)$ is the set of strategy profiles which are played when *in every state* there is common $p$-belief of the event $[\phi] \cap \mathcal{R} \cap \mathcal{P}$:

**Proposition 3.4.** *For all $p \in (0,1)$, we have $\tilde{R}^p(G) = \mathrm{Proj}_A \, \mathrm{CB}^p \left([\phi] \cap \mathcal{R} \cap \mathcal{P} \mid E(\Theta)\right)$.*

For any $p \in (0,1)$, Proposition 3.4 combined with the first part of Lemma 2.3 implies $\tilde{R}^p(G) \subseteq \mathrm{Proj}_A \, \mathrm{CB}^p \left([\phi] \cap \mathcal{R} \cap \mathcal{P}\right)$. And for any $p \in (0,1)$ such that $1 - \frac{1-p}{\min_{i \in I, \theta \in \Theta} \phi_i(\theta)} > 0$, Proposition 3.4 along with the second part of Lemma 2.3 implies

$$\mathrm{Proj}_A \, \mathrm{CB}^p \left([\phi] \cap \mathcal{R} \cap \mathcal{P}\right) \subseteq \tilde{R}^{1 - \frac{1-p}{\min_{i \in I, \theta \in \Theta} \phi_i(\theta)}}(G).$$

Letting $\pi \in (0,1)$ be as in Corollary 3.3, there exists $\bar{p} \in (\pi, 1)$ such that for all $p \geq \bar{p}$ we have $1 - \frac{1-p}{\min_{i \in I, \theta \in \Theta} \phi_i(\theta)} \geq \pi$. Then for all $p \geq \bar{p}$, $\tilde{R}^{1 - \frac{1-p}{\min_{i \in I, \theta \in \Theta} \phi_i(\theta)}}(G) = \tilde{R}^p(G) = S^\infty W(G)$. Hence, $\mathrm{Proj}_A \, \mathrm{CB}^p \left([\phi] \cap \mathcal{R} \cap \mathcal{P}\right) = \tilde{R}^p(G) = S^\infty W(G)$, completing the proof of Theorem 3.1.

## 3.2 ACK of Admissibility with Perturbed Priors

We now study the behavioral implications of ACK of admissibility with *perturbed priors*, as defined by the set PACKA$(G)$ in Section 2.2. In this case, Theorem 3.1 breaks down: $S^\infty W(G)$ continues to be a subset of PACKA$(G)$, but in general the containment is strict.

To see this, we first provide a characterization of PACKA$(G)$ in terms of a generalization of the perfect $p$-rationalizable set. Define the **$\varepsilon$-perturbed perfect $p$-rationalizable set** to be $\tilde{R}^{\varepsilon,p}(G) := \bigcap_{n\in\mathbb{N}} \tilde{\Lambda}^{\varepsilon,p,n}(A)$, where for any $A' := \prod_{i\in i} A'_i \subseteq A$ and $i \in I$, we let

$$\tilde{\Lambda}_i^{\varepsilon,p}(A') := \left\{ a_i \in A_i : \begin{array}{l} \exists \phi'_i \in \Delta(\Theta) \ \&\ \mu\colon \Theta \to \Delta^\circ(A_{-i}) \text{ s.t.} \\ a_i \in \mathrm{BR}_{A_i}^{\phi'_i}(\mu) \ \&\ ||\phi_i - \phi'_i||_\infty \le \varepsilon \ \&\ \mu(\theta)(A'_{-i}) \ge p \, \forall \theta \in \Theta \end{array} \right\},$$

and we inductively define $\tilde{\Lambda}^{\varepsilon,p,0}(A) = \tilde{\Lambda}^{\varepsilon,p}(A) := \prod_{i\in I} \tilde{\Lambda}_i^{\varepsilon,p}(A)$ and $\tilde{\Lambda}^{\varepsilon,p,n+1} := \tilde{\Lambda}^{\varepsilon,p}(\tilde{\Lambda}^{p,n}(A))$ for all $n \ge 0$. We have the following analog of Proposition 3.4:

**Theorem 3.5.** *For all $p \in (0,1)$ and $\varepsilon > 0$, $\mathrm{Proj}_A \, \mathrm{CB}^p \left( [\phi, \varepsilon] \cap \mathcal{R} \cap \mathcal{P} \mid E(\Theta) \right) = \tilde{R}^{\varepsilon,p}(G)$.*

By the same logic as in the paragraph following Proposition 3.4, Theorem 3.5 implies that there exists $\bar{p} \in (0,1)$ and $\bar{\varepsilon} > 0$ such that for all $p \in [\bar{p}, 1)$ and $\varepsilon \in (0, \bar{\varepsilon}]$, $\tilde{R}^{\varepsilon,p}(G) = \tilde{R}^{\bar{\varepsilon},\bar{p}}(G) = $ PACKA$(G)$. Since $\tilde{R}^p(G) \subseteq \tilde{R}^{\varepsilon,p}(G)$ for all $p \in (0,1)$ and $\varepsilon > 0$, it follows from Theorem 3.1 that $S^\infty W(G) \subseteq$ PACKA$(G)$. However, as the following example shows, this inclusion is generally strict:

**Example 3.6.** Consider the 2-player game $G$ where $A_1 = \{U, D\}$, $A_2 = \{L, R\}$, $\Theta = \{\theta, \theta'\}$, player 2 is indifferent across all outcomes and states, and 1's payoffs are given by:

| $\theta$ | $L$ | $R$ |
|---|---|---|
| $U$ | 1 | 1 |
| $D$ | 0 | 0 |

| $\theta'$ | $L$ | $R$ |
|---|---|---|
| $U$ | 0 | 1 |
| $D$ | 1 | 0 |

Suppose that $\phi_1 = \phi_2 =: \phi$ assigns equal probability to $\theta$ and $\theta'$. Then $D \notin S^\infty W(G)_1$ because it is weakly dominated by $U$ under $\phi$: Indeed, writing $XY$ (for any $X$, $Y \in A_2$) to denote the belief $\mu\colon \Theta \to A_2$ such that $\mu(\theta) = X$ and $\mu(\theta') = Y$, player 1's expected payoffs from $U$ and $D$ given $\phi$ are summarized by

| | $LL$ | $RL$ | $LR$ | $RR$ |
|---|---|---|---|---|
| $U$ | $\frac{1}{2}$ | $\frac{1}{2}$ | 1 | 1 |
| $D$ | $\frac{1}{2}$ | $\frac{1}{2}$ | 0 | 0 |

On the other hand, consider the sequence of perturbed beliefs $\phi^n \to \phi$ given by $\phi^n(\theta) = \frac{1}{2} - \frac{1}{n}$, $\phi^n(\theta') = \frac{1}{2} + \frac{1}{n}$ for $n \ge 3$, and consider the sequence of conjectures $\mu^n\colon \Theta \to \Delta^\circ(A_2)$ given by $\mu^n(\theta)[L] = \mu^n(\theta')[L] = 1 - \frac{1}{n^2}$, $\mu^n(\theta)[R] = \mu^n(\theta')[R] = \frac{1}{n^2}$. Given $\phi^n$, 1's expected payoffs against $\mu^n$ from $U$ and $D$ are $\frac{1}{2} - \frac{1}{n} + (\frac{1}{2} + \frac{1}{n})\frac{1}{n^2}$ and $(\frac{1}{2} + \frac{1}{n})(1 - \frac{1}{n^2})$, respectively. For large enough $n$, the latter is strictly greater than the former, so $D \in BR^{\phi^n}(\mu^n)$. Thus, $\tilde{\Lambda}_1^{\frac{1}{n},p,0}(A) = \{U, D\} = A_1$. Hence, for all $p \in (0,1)$ and $\varepsilon > 0$, $\tilde{R}^{\varepsilon,p}(G) = A \supsetneq S^\infty W(G)$.

The preceding example has the special feature that $\tilde{R}^{\frac{1}{n},p}(G) = \text{Proj}_{A_1} CB^p([\phi^n] \cap \mathcal{R} \cap \mathcal{P})$ for all $p \in (0,1)$ and large enough $n$. However, in general $\tilde{R}^{\varepsilon,p}(G)$ does not impose common $p$-belief in any single profile $(\phi_1', \phi_2')$ of $\varepsilon$-perturbed priors, instead allowing each player to be uncertain about the way in which his opponents' priors are perturbed, about the way in which his opponents think his priors might be perturbed etc. This additional uncertainty can generate additional behavior outside $S^\infty W(G)$: A given strategy might be played by player 1 only if player 1's belief about $\Theta$ is perturbed in a particular way $\phi_1'$ and if player 1 believes that player 2 plays a particular strategy—but the latter choice of strategy by player 2 might in turn require player 2 to believe that player 1 is willing to play some other strategy which player 1 would only play under a different belief perturbation $\phi_1''$. The following example illustrates this:

**Example 3.7.** Consider the 2-player game $G$ where $A_1 = \{U, I, D\}$, $A_2 = \{L, M, R\}$, $\Theta = \{\theta, \theta'\}$, and payoffs are summarized by the following two tables:

| $\theta$ | $L$ | $M$ | $R$ |
|---|---|---|---|
| $U$ | $1,0$ | $0,\frac{1}{2}$ | $0,1$ |
| $I$ | $\frac{1}{2},0$ | $\frac{1}{2},\frac{1}{2}$ | $\frac{1}{2},0$ |
| $D$ | $0,0$ | $0,\frac{1}{2}$ | $0,0$ |

| $\theta'$ | $L$ | $M$ | $R$ |
|---|---|---|---|
| $U$ | $0,1$ | $0,\frac{1}{2}$ | $0,0$ |
| $I$ | $\frac{1}{2},0$ | $\frac{1}{2},\frac{1}{2}$ | $\frac{1}{2},0$ |
| $D$ | $1,0$ | $0,\frac{1}{2}$ | $0,0$ |

Suppose that $\phi_1 = \phi_2 =: \phi$ assigns equal probability to $\theta$ and $\theta'$. Then for any $p \in (0,1)$ and $\varepsilon > 0$, we have $\tilde{R}^{\varepsilon,p}(G) = A$: Indeed, note first that under the belief $\phi_1' \in \Delta(\Theta)$ with $\phi_1'(\theta) = \alpha$, player 1's expected payoffs from $U$, $I$ and $D$ are summarized by the following table (again $XY$ denotes the belief that $X$ is played in $\theta$ and $Y$ is played in $\theta'$):

| | $LL$ | $LR$ | $LM$ | $RR$ | $RL$ | $RM$ | $MM$ | $MR$ | $ML$ |
|---|---|---|---|---|---|---|---|---|---|
| $U$ | $\alpha$ | $\alpha$ | $\alpha$ | $0$ | $0$ | $0$ | $0$ | $0$ | $0$ |
| $I$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ |
| $D$ | $1-\alpha$ | $0$ | $0$ | $0$ | $1-\alpha$ | $0$ | $0$ | $0$ | $1-\alpha$ |

From this it is easy to see that $\tilde{\Lambda}_1^{\varepsilon,p,0}(A) = \{U, I, D\} = A_1$, because $U$ is rationalizable under any $\phi_1'$ with $\alpha > \frac{1}{2}$ and $D$ is rationalizable under any $\phi_1'$ with $\alpha < \frac{1}{2}$ (and $I$ is a weakly dominant strategy). Similarly, under the belief $\phi_2' \in \Delta(\Theta)$ with $\phi_2'(\theta) = \alpha$, player 2's expected payoffs from $L$, $R$, and $M$ are summarized by:

| | $UU$ | $UI$ | $UD$ | $IU$ | $II$ | $ID$ | $DU$ | $DI$ | $DD$ |
|---|---|---|---|---|---|---|---|---|---|
| $L$ | $1-\alpha$ | $0$ | $0$ | $1-\alpha$ | $0$ | $0$ | $1-\alpha$ | $0$ | $0$ |
| $M$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ |
| $R$ | $\alpha$ | $\alpha$ | $\alpha$ | $0$ | $0$ | $0$ | $0$ | $0$ | $0$ |

Thus, by the same reasoning as above $\tilde{\Lambda}_2^{\varepsilon,p,0}(A) = \{L, M, R\} = A_2$, whence $\tilde{R}^{\varepsilon,p}(G) = A$.

On the other hand, fix $\varepsilon > 0$ sufficiently small and choose $p = p(\varepsilon)$ sufficiently close to 1. We claim that there is no $\phi' := (\phi_1', \phi_2') \in \Delta^\circ(\Theta)$ with $||\phi' - \phi|| \leq \varepsilon$ such that $D \in \text{Proj}_{A_1} CB^p([\phi'] \cap \mathcal{R} \cap \mathcal{P})$. The basic intuition is the following: An easy adaptation of the arguments in Section 3.1 shows that if $\varepsilon$ is sufficiently small and $p$ sufficiently close to 1, then for all $\phi'$ with $||\phi'-\phi|| \leq \varepsilon$, we have $\text{Proj}_{A_1} CB^p([\phi']\cap\mathcal{R}\cap\mathcal{P}) = \tilde{R}^p(G_{\phi'})$, where $G_{\phi'}$ is the incomplete information game with the same states of nature, actions and payoffs as $G$, but with belief profile $\phi'$ rather than $\phi$. Now note that for any $\phi'$ such that $D \in \tilde{\Lambda}_1^{p,0}(G_{\phi'})$, we must have $\phi_1'(\theta) < \frac{1}{2}$ and hence $U \notin \tilde{\Lambda}_1^{p,0}(G_{\phi'})$. But for $p > \frac{1}{2}$, $U \notin \tilde{\Lambda}_1^{p,0}(G_{\phi'})$ implies $L \notin \tilde{\Lambda}_2^{p,1}(G_{\phi'})$: Indeed, for any conjecture $\mu : \Theta \to \Delta^\circ(A_1)$ which in each state puts probability greater than $\frac{1}{2}$ on $\tilde{\Lambda}_1^{p,0}(G_{\phi'})$ (and hence on player 1 not playing $U$), player 2 is strictly better off playing $M$ than $L$. But then, at the next step of the iteration, this means that $D \notin \tilde{\Lambda}_1^{p,2}(G_{\phi'})$: If player 1 puts probability greater than $\frac{1}{2}$ on opponent strategies in $\tilde{\Lambda}_2^{p,1}(G_{\phi'})$ (and hence on $L$ not being played), then he is strictly better off playing $I$ than $D$.

# 4 Discussion

Our positive result in Section 3.1 shows that in an incomplete information setting, $S^\infty W(G)$ continues to admit clean epistemic foundations. However, from a modeler's perspective, the discussion in Section 3.2 suggests caution in making predictions based on $S^\infty W$, except in situations where common $p$-belief of an *exact* profile of priors seems a reasonable assumption—for example, this might be the case if the game is played in an experimental lab setting and prior to the start of the game the experimenter publicly announces an "objective" probability distribution $\phi$ on $\Theta$ to all players. Additional interpretations of ACK of admissibility naturally spring to mind: For instance, if we further weaken ACK of admissibility to allow players to entertain vanishingly small doubts about opponents' state-dependent payoffs,[13] then it is easy to see that any action profile in $S^\infty$ can be predicted.[14] Once again, one might argue that in many situations a modeler would be hard-pressed to rule out uncertainty in the form of such vanishingly small perturbations. If so, this suggests caution in applying any refinement of rationality short of $S^\infty$.[15]

Finally, we note that while we have interpreted $S^\infty W$ in the interim-correlated sense of Dekel et al. (2007), working with an alternative notion of $S^\infty W$ which requires each player's beliefs about opponents' behavior and about the states of nature to be *independent* would not fundamentally alter our results: Let $S^\infty W_I$ denote this independent version of $S^\infty W$[16] and let $\mathcal{H}$ denote the event that each player's first-order beliefs are independent across

[13]That is, we replace the event $\mathcal{R}_i$ for each player with the event $\mathcal{R}_i^\varepsilon$ which requires $a_i(\omega)$ to be a best-response given beliefs $f_i(t_i(\omega))$ and given a state-dependent utility function $u_i'$ such that $||u_i' - u_i||_\infty \leq \varepsilon$.

[14]This notion can be regarded as as an incomplete information analog of the notion of "weak convergence" also considered in Dekel and Fudenberg (1990), but since they study normal forms derived from extensive form games they restrict to payoff perturbations which respect the associated extensive form.

[15]In this sense, our conclusion would be similar in spirit to Weinstein and Yildiz (2007). However, our point is simply that any given action $a$ in $S^\infty$ can arise under ACK of admissibility with perturbed payoffs, not that $a$ will be *uniquely* rationalizable. To make this point, we can work with a more limited subset of perturbations of the type space than Weinstein and Yildiz (2007) and we do not need to impose any richness assumption on the game.

[16]Formally, $S^\infty W_I(G)$ is obtained by imposing on each instance of $\mu_i : \Theta \to \Delta(\hat{A}_{-i})$ in Definition 2.1 the requirement that $\mu_i(\theta) = \mu_i(\theta')$ for all $\theta$, $\theta'$. Note that we do allow for correlation across opponents' actions, because $\mu_i(\theta)$ is not assumed to be a product measure; thus our definition is less stringent than "interim independence" in the sense of Dekel et al. (2007).

opponents' behavior and states of nature.[17] Then analogously to Theorem 3.1 above, we can establish that $S^\infty W_I(G)$ coincides with $\operatorname{Proj}_A CB^p([\phi] \cap \mathcal{R} \cap \mathcal{P} \cap \mathcal{H})$ for $p$ close enough to 1. We can also obtain an analog of Theorem 3.5 under a suitably modified definition of the $\varepsilon$-perturbed perfect $p$-rationalizable set. Finally, it is easy to see that in Example 3.6 we also have $\operatorname{Proj}_A CB^p([\phi, \varepsilon] \cap \mathcal{R} \cap \mathcal{P} \cap \mathcal{H}) = A \supsetneq S^\infty W_I(G)$ for all $p \in (0,1)$ and $\varepsilon > 0$.

# A Extension of Dekel and Fudenberg (1990)

In this section, we briefly discuss Dekel and Fudenberg's (1990) characterization of $S^\infty W$: They show that in a complete information game $G$, $S^\infty W(G)$ represents the set of strategies which survive iterated admissibility in a sequence of *elaborations* $\tilde{G}_n$ converging to $G$, which differ from $G$ in allowing for vanishingly small amounts of payoff uncertainty. We extend this characterization to the incomplete information game $G$ defined in Section 2, but show that this result is once again very sensitive to the way in which the small amounts of additional payoff uncertainty are taken to interact with the uncertainty already present in $G$.

In our incomplete information setting, a sequence of elaborations is a sequence of $I$-player games $\tilde{G}_n$ with the same underlying strategy sets $A_i$ and the same state space $\Theta$ as $G$. There are finite type spaces $T_i$ for each player, and in each $\tilde{G}_n$, payoff functions are given by $\tilde{u}_i^n \colon A_i \times A_{-i} \times \Theta \times T_i \to [0,1]$ and beliefs by $\kappa_i^n : T_i \to \Delta^\circ(\Theta \times T_{-i})$. To capture the idea that payoff uncertainty is vanishingly small, we consider sequences of elaborations $\tilde{G}_n$ which *converge* to $G$, in the sense that in each elaboration there is a "sane" type for each player which has exactly the same payoffs as in $G$ and assigns increasingly large probability to his opponents being sane. Analogously to the two notions of ACK of admissibility we defined in Section 2.2, the question arises whether or not to allow sane types' uncertainty about opponents' payoffs to coincide with small amounts of uncertainty about opponents' beliefs about $\Theta$. Consider first the following strong notion of convergence, which requires the sane types in each elaboration to have *exactly* the same marginal beliefs on $\Theta$ as in $G$:

**Definition A.1.** A sequence of elaborations $\tilde{G}_n$ **converges strongly** to $G$, denoted $\tilde{G}_n \xrightarrow{S} G$, if for every $i$ there exists a "sane" type $\bar{t}_i \in T_i$ such that

(i). $\operatorname{marg}_\Theta \kappa_i^n(\bar{t}_i) = \phi_i$ for all $n$;

(ii). $\tilde{u}_i^n(\cdot, \theta, \bar{t}_i) = u_i(\cdot, \theta)$ for all $\theta \in \Theta$ and for all $n$;

(iii). for all $\theta$, $\kappa_i^n(\bar{t}_i)(\{\theta\} \times \{\bar{t}_{-i}\}) \to \phi_i(\theta)$ as $n \to \infty$.

Then interpreting iterated dominance in both $G$ and in each $\tilde{G}_n$ in the interim correlated sense,[18] we have the following extension of Dekel and Fudenberg's main result:

**Proposition A.2** (Extension of Proposition 3.1 in Dekel and Fudenberg (1990))**.** *Let $\bar{a} \in A$. Then $\bar{a} \in S^\infty W(G)$ if and only if there exists a sequence of elaborations $\tilde{G}_n \xrightarrow{S} G$ such that $\bar{a}_i \in W^\infty(\tilde{G}_n)(\bar{t}_i)$ for the sane type $\bar{t}_i$ of each player $i$ and for all $n$.*

[17] $\mathcal{H}_i := \{\omega : f_i(t_i(\omega))(\theta, a_{-i}) = \operatorname{marg}_\Theta f_i(t_i(\omega))(\theta) \cdot \operatorname{marg}_{A_{-i}} f_i(t_i(\omega))(a_{-i}) \,\forall (\theta, a_{-i})\}$ and $\mathcal{H} := \bigcap_{i \in I} \mathcal{H}_i$.

[18] Interim correlated dominance in $\tilde{G}_n$ allows each type to hold correlated beliefs over states of nature, opponent behavior, and opponent types; cf. Dekel et al. (2007) for the formal definition.

However, if we allow for vanishingly small perturbations to sane types' beliefs about $\Theta$ (i.e. in condition (i) of Definition A.1 we impose equality only in the limit as $n \to \infty$), then $S^\infty W(G)$ is once again only a (generally strict) subset of the behavior predicted by $W^\infty$ in a sequence of elaborations. This is easy to see by considering the game in Example 3.6, along with the sequence of elaborations with a single, sane type $\bar{t}_i$ for each player who has beliefs $\phi^n \in \Delta(\Theta)$ as in Example 3.6—here $D \in W^\infty(\tilde{G}_n)(\bar{t}_1)$ for all $n$ even though $D \notin S^\infty W(G)$.[19]

# B Proofs

***Proof of Lemma 2.2.*** Fix $i$ and consider the 2-player game complete information game $\tilde{G}$ with strategy sets $\tilde{A}_1 = \hat{A}_i$ (with typical elements $a_i$) and $\tilde{S}_2 = \hat{A}_{-i}^{\Theta}$ (with typical elements $\lambda_i \colon \Theta \to \hat{A}_{-i}$), where player 1's payoffs are given by $g_1(a_i, \lambda_i) = \sum_{\theta, a_{-i}} \phi_i(\theta)\lambda_i(\theta)[a_{-i}]u_i(a_i, a_{-i}, \theta)$ and player 2's payoffs are arbitrary.

Then $a_i$ is not strongly (respectively weakly) dominated in $\hat{A}_i$ if and only if $a_i$ is not strongly (respectively weakly) dominated for player 1 in $\tilde{G}$, which by Pearce (1984, Lemma 3) (respectively Pearce (1984, Lemma 4) ) is equivalent to the existence of some $\lambda_i \in \Delta(\hat{A}_{-i}^{\Theta})$ (respectively $\lambda_i \in \Delta^\circ(\hat{A}_{-i}^{\Theta})$) to which $a_i$ is a best response in $\tilde{G}$. But this is equivalent to there being a belief $\lambda_i \colon \Theta \to \Delta(\hat{A}_{-i})$ (respectively $\lambda_i \colon \Theta \to \Delta^\circ(\hat{A}_{-i})$) to which $a_i$ is a best response in $\hat{A}_i$. ■

***Proof of Lemma 2.3.*** Immediate from the definitions. ■

***Proof of Theorem 3.1.*** Given Corollary 3.3 and Proposition 3.4, which are proved below, the proof follows from the discussion in Section 3.1. ■

***Proof of Lemma 3.2.*** We adapt Börgers's (1994) proof to our setting: By finiteness of $I$ and of the strategy sets for each player, it suffices to find $\pi \in (0,1)$ with the required properties for some fixed $i$, $\hat{A}_i \subseteq A_i$ and $\hat{A}_{-i} \subseteq A_{-i}$. Note that by finiteness of the strategy sets and because the correspondence mapping every $p \in (0,1)$ to the set $D_i^p(\tilde{A}_i, \hat{A}_{-i})$ is non-empty valued and decreasing in $p$ (with respect to set inclusion), there exists $\pi \in (0,1)$ such that for all $p \geq \pi$, $D_i^p(\hat{A}_i, \hat{A}_{-i}) = D_i^\pi(\hat{A}_i, \hat{A}_{-i})$. We claim that $\pi$ is as required.

For the "only if" part, suppose that $a_i \in D_i^\pi(\hat{A}_i, \hat{A}_{-i})$. By definition of $D_i^\pi(\hat{A}_i, \hat{A}_{-i})$, $a_i$ is a best response in $\hat{A}_i$ to some $\mu_i \colon \Theta \to \Delta^\circ(A_{-i})$. Hence, by Lemma 2.2, there is no $\alpha_i \in \Delta(\hat{A}_i)$ which weakly dominates $a_i$ on $A_{-i}$, so condition (i) holds. To establish condition (ii), consider a sequence $(p_n)$ such that $p_n \in (\pi, 1)$ for all $n$ and $\lim_{n\to\infty} p_n = 1$. Then $a_i \in D_i^{p_n}(\hat{A}_i, \hat{A}_{-i}) = D_i^\pi(\hat{A}_i, \hat{A}_{-i})$ for all $n$. So for all $n$, $a_i$ is best response in $\hat{A}_i$ to some belief $\mu_i^n \colon \Theta \to \Delta(A_{-i})$ such that $\mu_i^n(\theta)[\hat{A}_{-i}] \geq p_n$ for all $\theta$. Passing to a subsequence if necessary, let $\mu_i$ denote the limit of the $\mu_i^n$. Then $\mu_i \colon \Theta \to \Delta(\hat{A}_{-i})$ and $a_i$ is a best response in $\hat{A}_i$ to $\mu_i$. Hence by Lemma 2.2, there is no $\alpha_i \in \Delta(\hat{A}_i)$ which strongly dominates $a_i$ on $\hat{A}_{-i}$, so condition (ii) is satisfied.

For the "if" part, suppose that $a_i \in \hat{A}_i$ satisfies conditions (i) and (ii). By Lemma 2.2, condition (i) implies that $a_i$ is a best response in $\hat{A}_i$ to some $\lambda_{i,1} \colon \Theta \to \Delta^\circ(A_{-i})$. Condition

[19]See the earlier working paper version of this paper for a general characterization of the robust extension of $W^\infty$ under this notion of "convergence with perturbed priors".

(ii) implies that $a_i$ is a best response in $\hat{A}_i$ to some $\lambda_{i,2}\colon \Theta \to \Delta(\hat{A}_{-i})$. Given any $p \in [\pi, 1)$, $a_i$ is still a best response in $\hat{A}_i$ to the convex combination $\lambda_i^p := (1-p)\lambda_{i,1} + p\lambda_{i,2}$. Moreover, $\lambda_i^p\colon \Theta \to \Delta^\circ(A_{-i})$ and $\lambda_i^p(\theta)[\hat{A}_{-i}] \geq p\lambda_{i,2}(\theta)[\hat{A}_{-i}] = p$ for all $\theta$. Hence, $a_i \in D_i^\pi(\hat{A}_i, \hat{A}_{-i})$. ■

***Proof of Corollary 3.3.*** We again adapt the argument in Börgers (1994): It is easy to see that for any $p \in (0,1)$ and $n \geq 0$, we have $\Lambda_i^{p,n+1}(A) := D_i^p(A_i, \Lambda_{-i}^{p,n}(A)) = D_i^p(\Lambda_i^{p,n}(A), \Lambda_{-i}^{p,n}(A))$. Hence, $\tilde{R}_i^p(G)$ can be determined as follows: First delete all strategies for every player $j$ that are not a best response to any belief $\lambda_j\colon \Theta \to \Delta^\circ(A_{-j})$. Then, *among all the remaining strategies*, delete all strategies for every player $j$ that are not a best response to any belief $\lambda_j\colon \Theta \to \Delta^\circ(A_{-j})$ that in every state assigns probability at least $p$ to the remaining strategies of the opponents. Iterate this procedure until no further strategies can be deleted, which by the finiteness of the strategy sets will happen in some finite number $m$ of steps. The resulting sets of strategies for each player $i$ will be $\tilde{R}_i^p(G) = \Lambda_i^{p,m}(A)$.

Now let $\pi$ be as found in Lemma 3.2. Then Lemma 3.2 implies that for $p \geq \pi$, the above procedure is equivalent to the procedure in which at each step all strategies are eliminated that are either weakly dominated in the original game or strongly dominated in the remaining reduced game. This in turn is equivalent to first deleting all weakly dominated strategies, and then at each later step deleting all strategies that are strongly dominated in the remaining game. This proves that for all $p \geq \pi$, we have $\tilde{R}_i^p(G) = S^\infty W(E)_i$, as claimed. ■

***Proof of Proposition 3.4 and of Theorem 3.5.*** We give a combined proof of Proposition 3.4 and of Theorem 3.5 by showing that for all $p \in (0,1)$ and for all $\varepsilon \geq 0$,

$$\operatorname{Proj}_A CB^p(([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P} \mid E(\Theta)) = \tilde{R}^{\varepsilon,p}(G).$$

Adapting the argument in Hu (2007) to our setting, we note first that for all $n$

$$\operatorname{Proj}_A B^{p,n}\left([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P} \mid E(\Theta)\right) = \tilde{\Lambda}^{\varepsilon,p,n}(A).$$

To see this, we proceed by induction on $n$ and repeatedly invoke Lemma B.1, which is stated and proved below. For $n = 0$, applying Lemma B.1 with $B = A_{-i} \times T_{-i}$ yields $\operatorname{Proj}_{A_i}\left([\phi_i,\varepsilon] \cap \mathcal{R}_i \cap \mathcal{P}_i\right) = \tilde{\Lambda}_i^{\varepsilon,p,0}(A)$. Since $\operatorname{Proj}_A\left([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P}\right) = \prod_i \operatorname{Proj}_{A_i}\left([\phi_i,\varepsilon] \cap \mathcal{R}_i \cap \mathcal{P}_i\right)$, it follows that $\operatorname{Proj}_A B^{p,0}\left([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P} \mid E(\Theta)\right) := \operatorname{Proj}_A\left([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P}\right) = \tilde{\Lambda}^{\varepsilon,p,0}(A)$ as required. Assuming the claim holds for $n$, apply Lemma B.1 with $B = \operatorname{Proj}_{A_{-i}\times T_{-i}} B^{p,n}\left([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P} \mid E(\Theta)\right)$. This yields

$$\begin{aligned}&\operatorname{Proj}_{A_i}\left(B_i^p\left(\Theta \times A_i \times T_i \times \operatorname{Proj}_{A_{-i}\times T_{-i}} B^{p,n}\left([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P} \mid E(\Theta)\right) \mid E(\Theta)\right) \cap \left([\phi_i,\varepsilon] \cap \mathcal{R}_i \cap \mathcal{P}_i\right)\right)\\ &= \Lambda_i^{\varepsilon,p}(A_i \times \operatorname{Proj}_{A_{-i}} B^{p,n}\left([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P} \mid E(\Theta)\right)) = \Lambda_i^{\varepsilon,p,n+1}(A),\end{aligned}$$

where the last equality uses the induction hypothesis. Taking the product across all $i$ yields

$$\begin{aligned}&\operatorname{Proj}_A\left(\bigcap_i B_i^p\left(\Theta \times A_i \times T_i \times \operatorname{Proj}_{A_{-i}\times T_{-i}} B^{p,n}\left([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P} \mid E(\Theta)\right) \mid E(\Theta)\right) \cap \left([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P}\right)\right)\\ &= \Lambda^{\varepsilon,p,n+1}(A)\end{aligned}$$

Finally, since $\Lambda^{\varepsilon,p,n+1}(A) \supseteq \Lambda^{\varepsilon,p,n}(A) = \operatorname{Proj}_A B^{p,n}([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P} \mid E(\Theta))$, we can intersect the LHS of the previous equality with $\operatorname{Proj}_A B^{p,n}([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P} \mid E(\Theta))$ without affecting the equality. Performing this intersection and simplifying yields $\operatorname{Proj}_A B^{p,n+1}([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P} \mid E(\Theta)) = \Lambda^{\varepsilon,p,n+1}(A)$, as required.

Given this it is immediate that $\operatorname{Proj}_A CB^p(([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P} \mid E(\Theta)) \subseteq \tilde{R}^{\varepsilon,p}(G)$. For the converse, suppose $a_i \in \tilde{R}_i^{\varepsilon,p}(G)$. Then by the previous paragraph, there is a sequence $\{t_i^n\}_{n=1}^\infty \subseteq T_i$ such that $(a_i, t_i^n) \in \operatorname{Proj}_{A_i \times T_i} B_i^p(B^{p,n-1}([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P} \mid E(\Theta)) \mid E(\Theta))$ for all $n$. Since $T_i$ is compact metric, there is a subsequence $\{t_i^{n_k}\}_{k=1}^\infty$ of $\{t_i^n\}_{n=1}^\infty$ which converges to some $t_i \in T_i$. By continuity of $\beta_i$, this implies $(a_i, t_i) \in \operatorname{Proj}_{A_i \times T_i} B_i^p(B^{p,n_k-1}([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P} \mid E(\Theta)) \mid E(\Theta))$ for all $k$. It follows that $a_i \in \operatorname{Proj}_{A_i} \mathrm{CB}^p([\phi,\varepsilon] \cap \mathcal{R} \cap \mathcal{P} \mid E(\Theta))$, as required. ■

**Lemma B.1.** *For all $B \subseteq A_{-i} \times T_{-i}$, $p \in (0,1)$ and $\varepsilon \geq 0$,*

$$\operatorname{Proj}_{A_i}(B_i^p(\Theta \times A_i \times T_i \times B \mid E(\Theta)) \cap ([\phi_i,\varepsilon] \cap \mathcal{R}_i \cap \mathcal{P}_i)) = \tilde{\Lambda}_i^{\varepsilon,p}\left(A_i \times \operatorname{Proj}_{A_{-i}} B\right).$$

***Proof of Lemma B.1.*** Suppose that

$$(a_i, t_i) \in \operatorname{Proj}_{A_i}(B_i^p(\Theta \times A_i \times T_i \times B \mid E(\Theta)) \cap ([\phi_i,\varepsilon] \cap \mathcal{R}_i \cap \mathcal{P}_i)).$$

Then $f_i(t_i)$ has full support on $\Theta \times A_{-i}$; for all $\theta \in \Theta$, we have $\frac{f_i(t_i)(\{\theta\} \times \operatorname{Proj}_{A_{-i}} B)}{f_i(t_i)(\{\theta\} \times A_{-i})} \geq p$; $a_i$ maximizes expected payoffs given $f_i(t_i)$; and there exists $\phi_i' \in \Delta(\Theta)$ such that $||\phi_i' - \phi|\,|_\infty \leq \varepsilon$ and such that $\phi_i'(\theta) = \operatorname{marg}_\Theta \beta_i(t_i)(\theta) = f_i(t_i)(\{\theta\} \times A_{-i})$ for all $\theta$. Defining $\mu\colon \Theta \to \Delta^\circ(A_{-i})$ by $\mu(\theta)(a_{-i}) = \frac{f_i(t_i)(\theta, a_{-i})}{f_i(t_i)(\{\theta\} \times A_{-i})}$, we get that $\mu(\theta)(\operatorname{Proj}_{A_{-i}} B) \geq p$ for all $\theta$ and $a_i \in BR^{\phi_i'}(\mu)$. This implies that $a_i \in \tilde{\Lambda}_i^{\varepsilon,p}\left(A_i \times \operatorname{Proj}_{A_{-i}} B\right)$, as required.

For the converse, suppose that $a_i \in \tilde{\Lambda}_i^{\varepsilon,p}\left(A_i \times \operatorname{Proj}_{A_{-i}} B\right)$. Then there exists $\mu\colon \Theta \to \Delta^\circ(A_{-i})$ such that $\mu(\theta)(\operatorname{Proj}_{A_{-i}} B) \geq p$ for all $\theta$ and such that $a_i \in BR_i^{\phi_i'}(\mu)$ for some $\phi_i' \in \Delta(\Theta)$ with $||\phi_i' - \phi_i|\,|_\infty \leq \varepsilon$. Define a map $\Phi_i\colon A_{-i} \to A_{-i} \times T_{-i}$ by $\Phi_i(a_{-i}) = (a_{-i}, t_{-i}[a_{-i}])$, where if $a_{-i} \in \operatorname{Proj}_{A_{-i}} B$ then $t_{-i}[a_{-i}]$ is chosen such that $(a_{-i}, t_{-i}[a_{-i}]) \in B$ and otherwise $t_{-i}[a_{-i}] \in T_{-i}$ is arbitrary. Consider the type $t_i \in T_i$ given by $\beta_i(t_i)(\theta, a_{-i}, t_{-i}) = \phi_i'(\theta) \cdot \mu(\theta)\left(\Phi_i^{-1}(a_{-i}, t_{-i})\right)$ for all $\theta$, $a_{-i}$, and $t_{-i}$. Then note that $\operatorname{marg}_\Theta \beta_i(t_i) = \phi_i'$; $f_i(t_i)(\theta, a_{-i}) = \phi_i'(\theta) \cdot \mu(\theta)(a_{-i}) > 0$ for all $\theta$ and $a_{-i}$; and $\frac{\beta_i(t_i)(\{\theta\} \times B)}{\beta_i(t_i)(\{\theta\} \times A_{-i} \times T_{-i})} = \mu(\theta)(\operatorname{Proj}_{A_{-i}} B) \geq p$ for all $\theta$. Thus, $(a_i, t_i) \in \operatorname{Proj}_{A_i}(B_i^p(\Theta \times A_i \times T_i \times B \mid E(\Theta)) \cap ([\phi_i,\varepsilon] \cap \mathcal{R}_i \cap \mathcal{P}_i))$, as required. ■

***Proof of Proposition A.2.*** We adapt the proof technique of Proposition 3.1 in Dekel and Fudenberg (1990) to our setting: For the "only if" direction, suppose $\bar{a} \in S^\infty W(G)$. We construct a sequence of elaborations $\tilde{G}_n$ converging strongly to $G$ such that for the sane type $\bar{t}_i$ of every player, we have $\bar{a}_i \in W^\infty(\tilde{G}_n)(\bar{t}_i)$. Let $T_i = \{\bar{t}_i, \hat{t}_i\}$, with $\tilde{u}_i^n(\cdot, \theta, \bar{t}_i) = u_i(\cdot, \theta)$ and $\tilde{u}_i^n(\cdot, \theta, \hat{t}_i) := 0$. Define $\kappa_i^n : T_i \to \Delta(\Theta \times T_{-i})$ by $\kappa_i^n(t_i)[\theta, \bar{t}_{-i}] = (1 - \frac{1}{n})\phi_i(\theta)$ and $\kappa_i^n(t_i)[\theta, \hat{t}_{-i}] = \frac{1}{n}\phi_i(\theta)$ for each $t_i$. Clearly, $\tilde{G}_n \xrightarrow{S} G$.

**Step 1:** $\bar{a}_i \in W(\tilde{G}_n)(\bar{t}_i)$.

Since $\bar{a}_i \in W(G)_i$, Lemma 2.2 yields $\bar{\mu}_i\colon \Theta \to \Delta^\circ(A_{-i})$ to which $\bar{a}_i$ is a best response in $G$. We can regard $\bar{\mu}_i$ as a map from $\Theta \times T_{-i}$ to $\Delta^\circ(A_{-i})$ which does not depend on the opponents' types. Since $\operatorname{marg}_\Theta \kappa_i^n(\bar{t}_i) = \phi_i$ and $\tilde{u}_i^n(\cdot, \theta, \bar{t}_i) = u_i(\cdot, \theta)$, it follows that for type

$\bar{t}_i$, $\bar{a}_i$ is a best response to $\bar{\mu}_i$ in $\tilde{G}_n$. Hence, by the analog of Lemma 2.2 for elaborations, $\bar{a}_i \in W(\tilde{G}_n)(\bar{t}_i)$.

**Step 2:** $\bar{a}_i \in W^2(\tilde{G}_n)(\bar{t}_i)$.

It suffices to find a map $\lambda_i \colon \Theta \times T_{-i} \to \Delta^\circ(A_{-i})$ such that $\text{supp}(\lambda_i)(\theta, t_{-i}) = \prod_{j \neq i} W(\tilde{G}_n)(t_j)$ for all $(\theta, t_{-i})$ and such that $\bar{a}_i$ is a best response to $\lambda_i$ in $\tilde{G}_n$. Because $\kappa_i^n(\bar{t}_i)$ assigns positive probability only to opponent type profiles $\bar{t}_{-i}$ and $\hat{t}_{-i}$, it suffices to define $\bar{\lambda}_i = \lambda_i(\cdot, \bar{t}_{-i}) \colon \Theta \to \Delta^\circ(A_{-i})$ and $\hat{\lambda}_i = \lambda_i(\cdot, \hat{t}_{-i}) \colon \Theta \to \Delta^\circ(A_{-i})$, where we need that $\text{supp}(\bar{\lambda}_i)(\theta) = \prod_{j \neq i} W(\tilde{G}_n)(\bar{t}_j)$ and $\text{supp}(\bar{\lambda}_i)(\theta) = \prod_{j \neq i} W(\tilde{G}_n)(\hat{t}_j) = A_{-i}$ for all $\theta$.

Since $\bar{a}_i \in SW(G)_i$, $\bar{a}_i$ is a best response to some $\mu_i \colon \Theta \to \Delta(A_{-i})$ with $\text{supp}(\mu_i)(\theta) \subseteq W(G)_{-i} \subseteq \prod_{j \neq i} W(\tilde{G}_n)(\bar{t}_j)$, where the last inclusion follows from Step 1. Let $\mu_i' \colon \Theta \to \Delta(A_{-i})$ be any map such that $\text{supp}(\mu_i')(\theta) = \prod_{j \neq i} W(\tilde{G}_n)(\bar{t}_j)$ for all $\theta$. Let $\bar{\mu}_i \colon \Theta \to \Delta^\circ(A_{-i})$ be as in Step 1, and let $\delta := \min_{\theta, a_{-i}} \bar{\mu}_i(\theta)[a_{-i}] \in (0,1)$. Let $\beta \in (0,1)$ be sufficiently small (more precisely, we require $\beta < \frac{\delta}{(n-1)}$). Set $\bar{\lambda}_i = (1-\beta)\mu_i + \beta\mu_i'$, so that for all $\theta$, $\text{supp}(\bar{\lambda}_i(\theta)) = \text{supp}(\mu_i'(\theta)) = \prod_{j \neq i} W(\tilde{G}_n)(\bar{t}_j)$, as required. Set $\hat{\lambda}_i = \bar{\mu}_i + \beta(n-1)(\bar{\mu}_i - \mu_i')$; then for all $\theta$, $\hat{\lambda}_i(\theta) \in \Delta^\circ(A_{-i})$, as required, because $\beta$ was chosen sufficiently small and because $\text{supp}(\bar{\mu}_i(\theta)) = A_{-i}$. Finally, by construction, we have for all $\theta$ that

$$\kappa_i^n(\bar{t}_i)[\theta, \bar{t}_{-i}]\bar{\lambda}_i(\theta) + \kappa_i^n(\bar{t}_i)[\theta, \hat{t}_{-i}]\hat{\lambda}_i(\theta) = \phi_i(\theta)\left[\frac{(n-1)(1-\beta)}{n}\mu_i(\theta) + \frac{1+\beta(n-1)}{n}\bar{\mu}_i(\theta)\right].$$

So, since $\tilde{u}_i^n(\cdot, \theta, \bar{t}_i) = u_i(\cdot, \theta)$ and since $\bar{a}_i$ is a best response to both $\bar{\mu}_i$ and $\mu_i$ in $G$, $\bar{a}_i$ is a best response for $\bar{t}_i$ to $\lambda_i = (\bar{\lambda}_i, \hat{\lambda}_i)$ in $\tilde{G}_n$.

Iterating Step 2, we obtain that $\bar{a}_i \in W^\infty(\tilde{G}_n)(\bar{t}_i)$, as claimed.

For the "if" direction, suppose $\tilde{G}_n \xrightarrow{S} G$ and consider $\bar{a} \in A$ such that $\bar{a}_i \in W^\infty(\tilde{G}_n)(\bar{t}_i)$ for the sane type $\bar{t}_i$ of each player $i$ and for all $n$. We show that $\bar{a}_i \in S^\infty W(G)_i$ for all $i$.

**<u>Step 1</u>:** $\bar{a}_i \in W(G)_i$.

Since $\bar{a}_i \in W(\tilde{G}_n)(\bar{t}_i)$ for all $n$, there is a sequence of beliefs $\lambda_i^n \colon \Theta \times T_{-i} \to \Delta^\circ(A_{-i})$ to which $\bar{a}_i$ is a best response for $\bar{t}_i$ in $\tilde{G}_n$. That is,

$$\bar{a}_i \in \text{argmax}_{a_i' \in A_i} \sum_{\theta, t_{-i}, a_{-i}} \kappa_i^n(\bar{t}_i)[\theta, t_{-i}]\lambda_i^n(\theta, t_{-i})[a_{-i}]\tilde{u}_i^n(a_i', a_{-i}, \theta, \bar{t}_i).$$

Using conditions (i) and (ii) of the definition of strong convergence, we can rewrite this as

$$\bar{a}_i \in \text{argmax}_{a_i' \in A_i} \sum_{\theta} \phi_i(\theta) \sum_{a_{-i}} (\sum_{t_{-i}} \kappa_i^n(t_{-i}|\theta, \bar{t}_i)\lambda_i^n(\theta, t_{-i})[a_{-i}]) u_i(a_i', a_{-i}, \theta),$$

where $\kappa_i^n(t_{-i}|\theta, t_i) := \frac{\kappa_i^n(t_i)[\theta, t_{-i}]}{\text{marg}_\Theta \kappa_i^n(t_i)[\theta]}$ (so that by condition (i) of strong convergence, $\kappa_i^n(t_{-i}|\theta, \bar{t}_i) = \frac{\kappa_i^n(\bar{t}_i)[\theta, t_{-i}]}{\phi_i(\theta)}$). So, setting $\hat{\lambda}_i^n(\theta)[a_{-i}] = \sum_{t_{-i}} \kappa_i^n(t_{-i}|\theta, \bar{t}_i)\lambda_i^n(\theta, t_{-i})[a_{-i}]$ for all $\theta$ and $a_{-i}$ yields beliefs $\hat{\lambda}_i^n \colon \Theta \to \Delta^\circ(A_{-i})$ to which $\bar{a}_i$ is a best response in $G$, where $\hat{\lambda}_i^n$ has range $\Delta^\circ(A_{-i})$ because $\lambda_i^n$ does. By Lemma 2.2, $\bar{a}_i \in W(G)_i$, as claimed.

**<u>Step 2</u>:** $\bar{a}_i \in SW(G)_i$.

Since $\bar{a}_i \in W^2(\tilde{G}_n)(\bar{t}_i)$ for all $n$, there are beliefs $\mu_i^n \colon \Theta \times T_{-i} \to \Delta(A_{-i})$ such that $\text{supp}\, \mu_i^n(\theta, t_{-i}) = \prod_{j \neq i} W(\tilde{G}_n)(t_j)$ for all $\theta$ and $t_{-i}$ and such that $\bar{a}_i$ is a best response for $\bar{t}_i$

to $\mu_i^n$ in $\tilde{G}_n$. As in Step 1, setting $\hat{\mu}_i^n(\theta)[a_{-i}] = \sum_{t_{-i}} \kappa_i^n(t_{-i}|\theta, \bar{t}_i)\mu_i^n(\theta, t_{-i})[a_{-i}]$ yields beliefs $\hat{\mu}_i^n : \Theta \to \Delta(A_{-i})$ to which $\bar{a}_i$ is a best response in $G$. Taking limits (passing to a subsequence if necessary) and using condition (iii) of the definition of strong convergence, we have that $\lim_n \hat{\mu}_i^n(\theta)[a_{-i}] = \lim_n \mu_i^n(\theta, \bar{t}_{-i})[a_{-i}]$. But $\operatorname{supp} \mu_i^n(\theta, \bar{t}_{-i}) = \prod_{j \neq i} W(\tilde{G}_n)(\bar{t}_j) \subseteq W(G)_{-i}$, where the last inclusion follows from Step 1. So, $\hat{\mu}_i := \lim_n \hat{\mu}_i^n$ is a map from $\Theta$ into $\Delta(W(G)_{-i})$. Moreover, by continuity $\bar{a}_i$ is a best response to $\hat{\mu}_i$ in $G$. Hence, by Lemma 2.2, $\bar{a}_i \in SW(G)_i$, as claimed.

Iterating the argument of Step 2, we conclude that $\bar{a}_i \in S^\infty W(G)_i$, as required. ■

# References


Battigalli, P., Siniscalchi, M., 2003. Rationalization and incomplete information. Advances in Theoretical Economics 3 (1), 57.

Börgers, T., 1994. Weak dominance and approximate common knowledge. Journal of Economic Theory 64, 265–265.

Brandenburger, A., Dekel, E., 1993. Hierarchies of beliefs and common knowledge. Journal of Economic Theory 59 (1), 189–198.

Brandenburger, A., Friedenberg, A., Keisler, H. J., 2008. Admissibility in games. Econometrica 76 (2), 307–352.

Dekel, E., Fudenberg, D., 1990. Rational behavior with payoff uncertainty. Journal of Economic Theory 52 (2), 243–267.

Dekel, E., Fudenberg, D., Morris, S., 2007. Interim correlated rationalizability. Theoretical Economics 2 (1), 15–40.

Dekel, E., Siniscalchi, M., 2014. Epistemic game theory. In: Young, P., Zamir, S. (Eds.), Handbook of Game Theory, Volume 4. North Holland.

Fudenberg, D., Kreps, D., Levine, D., 1988. On the robustness of equilibrium refinements. Journal of Economic Theory 44 (2), 354–380.

Hu, T.-W., 2007. On $p$-rationalizability and approximate common certainty of rationality. Journal of Economic Theory 136 (1), 379–391.

Kajii, A., Morris, S., 1997. The robustness of equilibria to incomplete information. Econometrica: Journal of the Econometric Society, 1283–1309.

Kohlberg, E., Mertens, J.-F., 1986. On the strategic stability of equilibria. Econometrica: Journal of the Econometric Society 55 (5), 1003–1037.

Luce, R. D., Raiffa, H., 1957. Games and decisions: Introduction and critical survey. Courier Dover Publications.

Pearce, D., 1984. Rationalizable strategic behavior and the problem of perfection. Econometrica: Journal of the Econometric Society, 1029–1050.

Schuhmacher, F., 1999. Proper rationalizability and backward induction. International Journal of Game Theory 28 (4), 599–615.

Tan, T. C.-C., da Costa Werlang, S. R., 1988. The bayesian foundations of solution concepts of games. Journal of Economic Theory 45 (2), 370–391.

Weinstein, J., Yildiz, M., 2007. A structure theorem for rationalizability with application to robust predictions of refinements. Econometrica 75 (2), 365–400.