A Logical Characterization of Iterated Admissibility

Joseph Y. Halpern and Rafael Pass
Computer Science Department, Cornell University, Ithaca, NY, 14853, U.S.A.
e-mail: halpern@cs.cornell.edu, rafael@cs.cornell.edu

June 23, 2009

Abstract

Brandenburger, Friedenberg, and Keisler provide an epistemic characterization of iterated admissibility (i.e., iterated deletion of weakly dominated strategies) where uncertainty is represented using LPSs (lexicographic probability sequences). Their characterization holds in a rich structure called a *complete* structure, where all types are possible. Here, a logical charaacterization of iterated admisibility is given that involves only standard probability and holds in all structures, not just complete structures. A stronger notion of *strong admissibility* is then defined. Roughly speaking, strong admissibility is meant to capture the intuition that "all the agent knows" is that the other agents satisfy the appropriate rationality assumptions. Strong admissibility makes it possible to relate admissibility, *canonical* structures (as typically considered in completeness proofs in modal logic), complete structures, and the notion of "all I know".

1 Introduction

Admissibility is an old criterion in decision making. A strategy for player i is admissible if it is a best response to some belief of player i that puts positive probability on all the strategy profiles for the other players. Part of the interest in admissibility comes from the observation (due to Pearce [1984]) that a strategy σ for player i is admissible iff it is not weakly dominated; that is, there is no strategy σ' for player i that gives i at least as high a payoff as σ no matter what strategy the other players are using, and sometimes gives i a higher payoff.

It seems natural to ignore strategies that are not admissible. But there is a conceptual problem when it comes to dealing with *iterated* admissibility (i.e., iterated deletion of weaklhy dominated strategies). As Mas-Colell, Whinston, and Green [1995, p. 240] put in their textbook when discussing iterated deletion of weakly dominated strategies:

> [T]he argument for deletion of a weakly dominated strategy for player i is that he contemplates the possibility that every strategy combination of his rivals occurs with positive probability. However, this hypothesis clashes with the logic of iterated deletion, which assumes, precisely, that eliminated strategies are not expected to occur.

Brandenburger, Friedenberg, and Keisler [2008] (BFK from now on) resolve this paradox in the context of iterated deletion of weakly dominated strategies by assuming that strategies are not really eliminated. Rather, they assumed that strategies that are weakly dominated occur with infinitesimal (but nonzero) probability. (Formally, this is captured by using an LPS—*lexicographically ordered probability sequence*.) They define a notion of belief (which they call *assumption*) appropriate for their setting, and show that strategies that survive k rounds of iterated deletion are ones that are played in states where there there is kth-order mutual belief in rationality; that is, everyone assume that everyone assumes . . . ($k-1$ times) that everyone is rational. However, they prove only that their characterization of iterated admissibility holds in particularly rich structures called *complete* structures (defined formally in Section 4), where all types are possible.

Here, we provide an alternate logical characterization of iterated admissibility. The characterization simply formalizes the intuition that an agent must consider possible all strategies consistent with the rationality assumptions he is making. Repeated iterations correspond to stronger rationality asumptions. The characterization has the advantage that it holds in all structures, not just complete structures, and assumes that agents represent their uncertainty using standard probability meaures, rather than LPS's or nonstandard probability measures (as is done in a characterization of Rajan [1998]). Moreover, while complete structures must be uncountable, we show that our characterization is always satisfible in a structure with finitely many states.

In an effort to understand better the role of complete structures, we consider *strong admissibility*. Roughly speaking, strong admissibility is meant to capture the intuition that "all the agent knows" is that the other agents satisfy the appropriate rationality assumptions. We are using the phrase "all agent i knows" here in the same sense that it is used by Levesque [1990] and Halpern and Lakemeyer [2001]. We formalize strong admissibility by requiring that the agent ascribe positive probability to all formulas consistent with his rationality assumptions. (This admittedly fuzzy description is made precise in Section 3.) We give a logical characterization of iterated strong admissibility and show that a strategy σ survives iterated deletion of weakly dominated strategies iff there is a structure and a state where σ is played and the formula characterizing iterated strong admissibility holds. While we can take the structure where the formula holds to be countable, perhaps the most natural structure to consider is the *canonical* structure, which has a state corresponding to very satisfiable collection of formulas. The canonical structure is uncountable.

We can show that the canonical structure is complete in the sense of BFK. Moreover, under a technical assumption, every complete structure is essentially canonical (i.e., it has a state corresponding to every satisfiable collection of formulas). This sequence of results allows us to connect (iterated admissibility), complete structures, canonical structures, and the notion of "all I know".

2 Characterizing Iterated Deletion

We consider normal-form games with n players. Given a (normal-form) n-player game Γ, let $\Sigma_i(\Gamma)$ denote the strategies of player i in Γ. We omit the parenthetical Γ when it is clear from context or irrelevant. Let $\vec{\Sigma} = \Sigma_1 \times \cdots \times \Sigma_n$.

Let $\mathcal{L}_1$ be the language where we start with *true* and the special primitive proposition RAT_i and close off under modal operators B_i and $\langle B_i \rangle$, for $i = 1, \ldots, n$, conjunction, and negation. We think of $B_i\varphi$ as saying that φ holds with probability 1, and $\langle B_i \rangle\varphi$ as saying that φ holds with positive probability. As we shall see, $\langle B_i \rangle$ is definable as $\neg B_i \neg$ if we make the appropriate measurability assumptions.

To reason about the game Γ, we consider a class of probability structures corresponding to Γ. A *probability structure* M *appropriate for* Γ is a tuple $(\Omega, \mathbf{s}, \mathcal{F}, \mathcal{PR}_1, \ldots, \mathcal{PR}_n)$, where Ω is a set of states; $\mathbf{s}$ associates with each state $\omega \in \Omega$ a pure strategy profile $\mathbf{s}(\omega)$ in the game Γ; $\mathcal{F}$ is a σ-algebra over Ω; and, for each player i, $\mathcal{PR}_i$ associates with each state ω a probability distribution $\mathcal{PR}_i(\omega)$ on $(\Omega, \mathcal{F})$ such that, (1) for each strategy σ_i for player i, $[\![\sigma_i]\!]_M = \{\omega : \mathbf{s}_i(\omega) = \sigma_i\} \in \mathcal{F}$, where $\mathbf{s}_i(\omega)$ denotes player i's strategy in the strategy profile $\mathbf{s}(\omega)$; (2) $\mathcal{PR}_i(\omega)([\![\mathbf{s}_i(\omega)]\!]_M) = 1$; (3) for each probability measure π on $(\Omega, \mathcal{F})$, and player i, $[\![\pi, i]\!]_M = \{\omega : \Pi_i(\omega) = \pi\} \in \mathcal{F}$; and (4) $\mathcal{PR}_i(\omega)([\![\mathcal{PR}_i(\omega), i]\!]_M) = 1$. These assumptions essentially say that player i knows his strategy and knows his beliefs.

The semantics is given as follows:

- $(M, \omega) \models$ *true* (so *true* is vacuously true).
- $(M, \omega) \models RAT_i$ if $\mathbf{s}_i(\omega)$ is a best response, given player i's beliefs on the strategies of other players induced by $\mathcal{PR}_i(\omega)$. (Because we restrict to appropriate structures, a players expected utility at a state ω is well defined, so we can talk about best responses.)
- $(M, \omega) \models \neg\varphi$ if $(M, \omega) \not\models \varphi$.
- $(M, \omega) \models \varphi \wedge \varphi'$ iff $(M, \omega) \models \varphi$ and $(M, \omega) \models \varphi'$
- $(M, \omega) \models B_i\varphi$ if there exists a set $F \in \mathcal{F}_i$ such that $F \subseteq [\![\varphi]\!]_M$ and $\mathcal{PR}_i(\omega)(F) = 1$, where $[\![\varphi]\!]_M = \{\omega : (M, \omega) \models \varphi\}$.
- $(M, \omega) \models \langle B_i \rangle\varphi$ if there exists a set $F \in \mathcal{F}_i$ such that $F \subseteq [\![\varphi]\!]_M$ and $\mathcal{PR}_i(\omega)(F) > 0$.

Given a language (set of formulas) $\mathcal{L}$, M is *$\mathcal{L}$-measurable* if M is appropriate (for some game Γ) and $[\![\varphi]\!]_M \in \mathcal{F}$ for all formulas $\varphi \in \mathcal{L}$. It is easy to check that in an $\mathcal{L}_1$-measurable structure, $\langle B_i \rangle\varphi$ is equivalent to $\neg B_i \neg\varphi$.

To put our results on iterated admissibility into context, we first consider rationalizability. Pearce [1984] gives two definitions of rationalizability, which give rise to different epistemic characterizations. We repeat the definitions here, using the notation of Osborne and Rubinstein [1994].

Definition 2.1: A strategy σ for player i in game Γ is *rationalizable* if, for each player j, there is a set $\mathcal{Z}_j \subseteq \Sigma_j(\Gamma)$ and, for each strategy $\sigma' \in \mathcal{Z}_j$, a probability measure $\mu_{\sigma'}$ on $\Sigma_{-j}(\Gamma)$ whose support is a subset of $\mathcal{Z}_{-j}$ such that

- $\sigma \in \mathcal{Z}_i$; and
- for each player j and strategy $\sigma' \in \mathcal{Z}_j$, strategy σ' is a best response to (the beliefs) $\mu_{\sigma'}$.

■

The second definition characterizes rationalizability in terms of iterated deletion.

Definition 2.2: A strategy σ for player i in game Γ is *rationalizable*$'$ if, for each player j, there exists a sequence $X_j^0, X_j^1, X_j^2, \ldots$ of sets of strategies for player j such that $X_j^0 = \Sigma_j$ and, for each strategy $\sigma' \in X_j^k$, $k \geq 1$, a probability measure $\mu_{\sigma',k}$ whose support is a subset of $\vec{X}_{-j}^{k-1}$ such that

- $\sigma \in \cap_{j=0}^{\infty} X_i$; and
- for each player j, each strategy $\sigma' \in X_j^k$ is a best response to the beliefs $\mu_{\sigma',k}$.

■

Intuitively, X_j^1 consists of strategies that are best responses to some belief of player j, and X_j^{h+1} consists of strategies in X_j^h that are best responses to some belief of player j with support X_{-j}^h; that is, beliefs that assume that everyone else is best reponding to some beliefs assuming that everyone else is responding to some beliefs assuming …(h times).

Proposition 2.3: [Pearce 1984] *A strategy is rationalizable iff it is rationalizable$'$.*

We now give our epistemic characterizations of rationalizability. Let RAT be an abbreviation for $RAT_1 \wedge \ldots \wedge RAT_n$; let $E\varphi$ be an abbreviation of $B_1\varphi \wedge \ldots \wedge B_n\varphi$; and define $E^k\varphi$ for all k inductively by taking $E^0\varphi$ to be φ and $E^{k+1}\varphi$ to be $E(E^k\varphi)$. Common knowledge of φ holds iff $E^k\varphi$ holds for all $k \geq 0$.

We now give an epistemic characterization of rationalizability. Part of the characterization (the equivalence of (a) and (b) below) is well known [Tan and Werlang 1988]; it just says that a strategy is rationalizable iff it can be played in a state where rationality is common knowledge.

Theorem 2.4: *The following are equivalent:*

(a) *σ is a rationalizable strategy for i in a game Γ;*

(b) *there exists a measurable structure M that is appropriate for Γ and a state ω such that $\mathbf{s}_i(\omega) = \sigma$ and $(M, \omega) \models E^k RAT$ for all $k \geq 0$;*

(c) *there exists a measurable structure M that is appropriate for Γ and a state ω such that $\mathbf{s}_i(\omega) = \sigma$ and $(M, \omega) \models \langle B_i \rangle E^k RAT$ for all $k \geq 0$;*

(d) *there exists a structure M that is appropriate for Γ and a state ω such that $\mathbf{s}_i(\omega) = \sigma$ and $(M, \omega) \models \langle B_i \rangle E^k RAT$ for all $k \geq 0$.*

Proof: Suppose that σ is rationalizable. Choose $\mathcal{Z}_j \subseteq \Sigma_j(\Gamma)$ and measures $\mu_{\sigma'}$ for each strategy $\sigma' \in \mathcal{Z}_j$ guaranteed to exist by Definition 2.1. Define an appropriate structure $M = (\Omega, \mathbf{s}, \mathcal{F}, \mathcal{PR}_1, \ldots, \mathcal{PR}_n)$, where

- $\Omega = \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_n$;
- $\mathbf{s}_i(\vec{\sigma}) = \sigma_i$;
- $\mathcal{F}$ consist of all subsets of Ω;
- $\mathcal{PR}_i(\vec{\sigma})(\vec{\sigma}')$ is 0 if $\sigma'_i \neq \sigma_i$ and is $\mu_{\sigma_i}(\sigma'_{-i})$ otherwise.

Since each player is best responding to his beliefs at every state, it is easy to see that $(M, \vec{\sigma}) \models RAT$ for all states $\vec{\sigma}$. It easily follows (formally, by induction on k), that $(M, \vec{\sigma}) \models E^k RAT$. Clearly M is measurable. This shows that (a) implies (b).

The fact that (b) implies (c) is immediate, since if $E^{k+1}\varphi$ logically implies $B_i E^k \varphi$, which in turn logically implies $\langle B_i \rangle_i E^k \varphi$ for all k and all formulas φ. The fact that (c) implies (d) is also immediate.

Finally, to see that (d) implies (a), suppose that M is a structure appropriate for Γ and ω is a state in M such that $\mathbf{s}_i(\omega) = \sigma$ and $(M, \omega) \models \langle B_i \rangle E^k RAT$ for all $k \geq 0$. For each player j, define the formulas C^k inductively by taking C_j^0 to be *true* and C_j^{k+1} to be $RAT_j \wedge B_j(\wedge_{j' \neq j} C_{j'}^k)$. An easy induction shows that for $k > 1$, C_j^k is equivalent to $RAT_j \wedge B_j(E^0 RAT \wedge \ldots \wedge E^{k-2} RAT)$ in appropriate structures. Define $X_j^k = \{\mathbf{s}_j(\omega') : (M, \omega') \models C_j^k\}$. If $\sigma' \in X_j^k$ for $k \geq 1$, choose some state ω' such that $(M, \omega') \models RAT_j \wedge B_j E^{k-2} RAT$ and $\mathbf{s}_j(\omega') = \sigma'$, and define $\mu_{\sigma',k}$ to be the projection of $\mathcal{PR}_j(\omega')$ onto Σ_{-j}. It easily follows that the support of $\mu_{\sigma',k}$ is X_{-j}^{k-1} and that σ' is a best response with respect to $\mu_{\sigma,k}$. Finally, since $(M, \omega) \models \langle B_i \rangle E^k RAT$ for all $k \geq 0$, it easily follows that $\sigma = \mathbf{s}_i(\omega) \in \cap_{k=0}^{\infty} X_i^k$. Thus, by Definition 2.2, σ is rationalizable$'$ and, by Proposition 2.3, σ is rationalizable. ∎

We now characterize iterated deletion of strongly dominated (resp., weakly dominated) strategies.

Definition 2.5: Strategy σ for player is i *strongly dominated by σ' with respect to $\Sigma'_{-i} \subseteq \Sigma_{-i}$* if $u_i(\sigma, \tau_{-i}) > u_i(\sigma, \tau_{-i})$ for all $\tau_{-i} \in \Sigma'_{-i}$. Strategy σ for player is i *weakly dominated by σ' with respect to* $\Sigma'_{-i} \subseteq \Sigma_{-i}$ if $u_i(\sigma, \tau_{-i}) \geq u_i(\sigma, \tau_{-i})$ for all $\tau_{-i} \in \Sigma'_{-i}$ and $u_i(\sigma, \tau'_{-i}) > u_i(\sigma, \tau'_{-i})$ for some $\tau'_{-i} \in \Sigma'_{-i}$.

Strategy σ for player i survives k rounds of iterated deletion of strongly dominated (resp., weakly dominated) strategies if, for each player j, there exists a sequence $X_j^0, X_j^1, X_j^2, \ldots, X_j^k$ of sets of strategies for player j such that $X_j^0 = \Sigma_j$ and, if $h < k$, then X_j^{h+1} consists of the strategies in X_j^h not strongly (resp., weakly) dominated by any strategy with respect to X_{-j}^h, and $\sigma \in X_i^k$. Strategy σ survives iterated deletion of strongly dominated (resp., weakly dominated) strategies if it survives k rounds of iterated deletion for all k. ∎

The following well-known result connects strong and weak dominance to best responses.

Proposition 2.6: [Pearce 1984]

- *A strategy σ for player i is not strongly dominated by any strategy with respect to Σ'_{-i} iff there is a belief μ_σ of player i whose support is a subset of Σ'_{-i} such that σ is a best response with respect to μ_σ.*
- *A strategy σ for player i is not weakly dominated by any strategy with respect to Σ'_{-i} iff there is a belief μ_σ of player i whose support is all of Σ'_{-i} such that σ is a best response with respect to μ_σ.*

It immediately follows from Propositions 2.3 and 2.6 (and is well known) that a strategy is rationalizable iff it survives iterated deletion of strongly dominated strategies. Thus, the characterization of rationalizability in Theorem 2.4 is also a characterization of strategies that survive iterated deletion of strongly dominated strategies. To characterize iterated deletion of weakly dominated strategies, we need to enrich the langauge $\mathcal{L}_1$ somewhat. Let $\mathcal{L}_2(\Gamma)$ be the extension of $\mathcal{L}_1$ that includes a primitive proposition $play_i(\sigma)$ for each player i and strategy $\sigma \in \Sigma_i$, and is also closed off under the modal operator $\Diamond$. We omit the parenthetical Γ when it is clear from context. We extend the truth relation to $\mathcal{L}_2$ in probability structures appropriate for Γ as follows:

- $(M, \omega) \models play_i(\sigma)$ iff $\omega \in [\![\sigma]\!]_M$.
- $(M, \omega) \models \Diamond\varphi$ iff there is some structure M' appropriate for Γ and state ω' such that $(M', \omega') \models \varphi$.

Intuitively, $\Diamond\varphi$ is true if there is some state and structure where φ is true; that is, if φ_i is satisfiable. Note that if $\Diamond\varphi$ is true at some state, then it is true at all states in all structures.

Let $play(\vec{\sigma})$ be an abbreviation for $\wedge_{j=1}^n play_j(\sigma_j)$, and let $play_{-i}(\sigma_{-i})$ be an abbrevation for $\wedge_{j\neq i} play_j(\sigma_j)$. Intuitively, $(M, \omega) \models play(\vec{\sigma})$ iff $\mathbf{s}(\omega) = \sigma$, and $(M, \omega) \models play_{-i}(\sigma_{-i})$ if, at ω, the players other than i are playing strategy profile σ_{-i}. Define the formulas D_j^k inductively by taking D_j^0 to be the formula *true*, and D_j^{k+1} to be an abbreviation of

$$RAT_j \wedge B_j(\wedge_{j'\neq j} D_{j'}^k) \wedge (\wedge_{\sigma_{-j}\in\Sigma_{-j}} \Diamond(play_{-j}(\sigma_{-j}) \wedge (\wedge_{j'\neq j} D_{j'}^k)) \Rightarrow \langle B_j\rangle(play_{-j}(\sigma_{-j})).$$

It is easy to see that D_j^k implies the formula C_j^k defined in the proof of Theorem 2.4, and hence implies $RAT_j \wedge B_j(E^0 RAT \wedge \ldots \wedge E^{k-2} RAT)$. But D_j^k requires more; it requires that player j assign positive probability to each strategy profile for the other players that is compatible with D_{-j}^{k-1}.

Theorem 2.7: *The following are equivalent:*

(a) the strategy σ for player i survives k rounds of iterated deletion of weakly dominated strategies;

(b) for all $k' \leq k$, there is a measurable structure $M^{k'}$ appropriate for Γ and a state $\omega^{k'}$ in $M^{k'}$ such that $\mathbf{s}_i(\omega^{k'}) = \sigma$ and $(M^{k'}, \omega^{k'}) \models D_i^{k'}$;

(c) for all $k' \leq k$, there is a structure $M^{k'}$ appropriate for Γ and a state $\omega^{k'}$ in $M^{k'}$ such that $\mathbf{s}_i(\omega^{k'}) = \sigma$ and $(M^{k'}, \omega^{k'}) \models D_i^{k'}$.

In addition, there is a finite structure $\overline{M}^k = (\Omega^k, \mathbf{s}, \mathcal{F}, \mathcal{PR}_1, \ldots, \mathcal{PR}_n)$ *such that* $\Omega^k = \{(k', i, \vec{\sigma}) : k' \le k, 1 \le i \le n, \vec{\sigma} \in X_1^{k'} \times \cdots \times X_n^{k'}\}$, $\mathbf{s}(k', i, \vec{\sigma}) = \vec{\sigma}$, $\mathcal{F} = 2^{\Omega^k}$, *where* $X_j^{k'}$ *consists of all strategies for player* j *that survive* k' *rounds of iterated deletion of weakly dominated strategies and, for all states* $(k', i, \vec{\sigma}) \in \Omega^k$, $(\overline{M}^k, (k', i, \vec{\sigma})) \models \wedge_{j \ne i} D_j^{k'}$.

Proof: We proceed by induction on k, proving both the equivalence of (a), (b), and (c) and the existence of a structure $\overline{M}^k$ with the required properties.

The result clearly holds if $k = 0$. Suppose that the result holds for k; we show that it holds for $k+1$. We first show that (c) implies (a). Suppose that $(M^{k'}, \omega^{k'}) \models D_j^{k'}$ and $\mathbf{s}_j(\omega^{k'}) = \sigma_j$ for all $k' \le k+1$. It follows that σ_j is a best response to the belief μ_{σ_j} on the strategies of other players induced by $\mathcal{PR}_j^{k+1}(\omega)$. Since $(M^{k+1}, \omega^{k+1}) \models B_j(\wedge_{j' \ne j} D_{j'}^k)$, it follows from the induction hypothesis that the support of μ_{σ_j} is contained in X_{-j}^k. Since $(M, \omega) \models \wedge_{\sigma_{-j} \in \Sigma_{-j}} (\Diamond(play_{-j}(\sigma_{-j}) \wedge (\wedge_{j \ne i} D_j^k)) \Rightarrow \langle B_j \rangle (play_{-j}(\sigma_{-j})))$, it follows from the induction hypothesis that the support of μ_{σ_j} is all of X_{-j}^k. Since $(M^{k'}, \omega^{k'}) \models D_j^{k'}$ for $k' \le k$, it follows from the induction hypothesis that $\sigma_j \in X_j^k$. Thus, $\sigma_j \in X_j^{k+1}$.

We next construct the structure $\overline{M}^{k+1} = (\Omega^{k+1}, \mathbf{s}, \mathcal{F}, \mathcal{PR}_1, \ldots, \mathcal{PR}_n)$. As required, we define $\Omega^{k+1} = \{(k', i, \vec{\sigma}) : k' \le k+1, 1 \le i \le n, \vec{\sigma} \in X_1^{k'} \times \cdots \times X_n^{k'}\}$, $\mathbf{s}(k', i, \vec{\sigma}) = \vec{\sigma}$, $\mathcal{F} = 2^{\Omega^{k+1}}$. For a state ω of the form $(k', i, \vec{\sigma})$, since $\sigma_j \in X_j^{k'}$, by Proposition 2.6, there exists a distribution μ_{k', σ_j} whose support is all of X_{-j}^{k-1} such that σ_j is a best response to μ_{σ_j}. Extend μ_{k', σ_j} to a distribution μ'_{k', i, σ_j} on Ω^{k+1} as follows:

- for $i \ne j$, let $\mu'_{k', i, \sigma_j}(k'', i', \vec{\tau}) = \mu_{k', \sigma_j}(\vec{\tau}_{-j})$ if $i' = j, k'' = k' - 1$, and $\tau_j = \sigma_j$, and 0 otherwise;
- $\mu'_{k', j, \sigma_j}(k'', i', \vec{\tau}) = \mu_{k', \sigma_j}(\vec{\tau}_{-j})$ if $i' = j, k'' = k'$, and $\tau_j = \sigma_j$, and 0 otherwise.

Let $\mathcal{PR}_j(k', i, \vec{\sigma}) = \sigma'_{k', i, \sigma_j}$. We leave it to the reader to check that this structure is appropriate. An easy induction on k' now shows that $(\overline{M}^{k+1}, (k', i, \vec{\sigma})) \models \wedge_{j \ne i} D_j^{k'}$ for $i = 1, \ldots, n$.

To see that (a) implies (b), suppose that $\sigma_j \in X_j^{k+1}$. Choose a state ω in $\overline{M}^{k+1}$ of the form $(k+1, i, \vec{\sigma})$, where $i \ne j$. As we just showed, $(\overline{M}^{k+1}, (k', i, \vec{\sigma}) \models D_j^{k'}$, and $\mathbf{s}_j(k', i, \vec{\sigma}) = \sigma_j$. Moreover, $\overline{M}^{k+1}$ is measurable (since $\mathcal{F}$ consists of all subsets of Ω^{k+1}).

Clearly (b) implies (c). ∎

Corollary 2.8: *The following are equivalent:*

(a) the strategy σ *for player* i *survives iterated deletion of weakly dominated strategies;*

(b) there is a measurable structure M *that is appropriate for* Γ *and a state* ω *such that* $\mathbf{s}_i(\omega) = \sigma$ *and* $(M, \omega) \models \langle B_i \rangle D_i^k$ *for all* $k \ge 0$*;*

(c) there is a structure M *that is appropriate for* Γ *and a state* ω *such that* $\mathbf{s}_i(\omega) = \sigma$ *and* $(M, \omega) \models \langle B_i \rangle D_i^k$ *for all* $k \ge 0$.

Note that there is no analogue of Theorem 2.4(b) here. This is because there is no state where D_i^k holds for all $k \geq 0$; it cannot be the case that i places positive probability on all strategies (as required by D_1^k) and that i places positive probability only on strategies that survive one round of iterated deletion (as required by D_2^k), unless all strategies survive one round on iterated deletion. We can say something slightly weaker though. There is some k such that the process of iterated deletion converges; that is, $X_j^k = X_j^{k+1}$ for all j (and hence $X_j^k = X_j^{k'}$ for all $k' \geq k$). That means that there is a state where $D_i^{k'}$ holds for all $k' > k$. Thus, we can show that a strategy σ for player i survives iterated deletion of weakly dominated strategies iff there exists a k and a state ω such that $\mathbf{s}_i(\omega) = \sigma$ and $(M, \omega) \models D_i^{k'}$ for all $k' > k$. Since C_j^{k+1} implies C_i^k, an anlagous results holds for iterated deletion of strongly dominated strategies, with $D_i^{k'}$ replaced by $C_i^{k'}$.

It is also worth noting that in a state where D^k holds, an agent does *not* consider all strategies possible, but only the ones consistent with the appropriate level of rationality. We could require the agent to consider all strategies possible by using LPS's or nonstandard probability. The only change that this would make to our characterization is that, if we are using nonstandard probability, we would interpret $B_i\varphi$ to mean that φ holds with probability infinitesimally close to 1, while $\langle B_i \rangle \varphi$ would mean that φ holds with probability whose standard part is positive (i.e., non-infinitesimal probability). We do not pursue this point further.

3 Strong Admissibility

We have formalized iterated admissibility by saying that an agent consider possible all strategies consistent with the appropriate rationality assumption. But why focus just on strategies? We now consider a stronger admissibility requirement that we call, not surprisingly, *strong admissibility*. Here we require, intuitively, that *all* an agent knows about the other agents is that they satisfy the appropriate rationality assumptions. Thus, the agent ascribes positive probability to all beliefs that the other agents could have as well as all the strategies they could be using. By considering strong admissibility, we will be able to relate work on "all I know" [Halpern and Lakemeyer 2001; Levesque 1990], BFK's notion of complete structures, and admisibility.

Roughly speaking, we interpret "all agent i knows is φ" as meaning that agent i believes φ, and considers possible every formula about the other players' strategies and beliefs consistent with φ. Thus, what "all I know" means is very sensitive to the choice of language. Let $\mathcal{L}^0$ be the language whose only formulas are (Boolean combinations of) formulas of the form $play_i(\sigma)$, $i = 1, \ldots, n$, $\sigma \in \Sigma_i$. Let $\mathcal{L}_i^0$ consist of just the formulas of the form $play_i(\sigma)$, and let $\mathcal{L}_{-i}^0 = \cup_{j \neq i} \mathcal{L}_j^0$. Define $O_i^- \varphi$ to be an abbreviation for $B_i\varphi \wedge (\wedge_{\psi \in \mathcal{L}_{-i}^0} \Diamond(\varphi \wedge \psi) \Rightarrow \langle B_i \rangle \psi)$. Then it is easy to see that D_j^{k+1} is just $RAT_j \wedge O_j^-(\wedge_{j' \neq j} D_{j'}^k)$.

We can think of $O_i^- \varphi$ as saying "all agent i knows with respect to the language $\mathcal{L}^0$ is φ." The language $\mathcal{L}^0$ is quite weak. To relate our results to those of BFK, even the language $\mathcal{L}^2$ is too weak, since it does not allow an agent to express probabilistic beliefs. Let $\mathcal{L}^3(\Gamma)$ be the language that extends $\mathcal{L}^2(\Gamma)$ by allowing formulas of the form $pr_i(\varphi) \geq \alpha$ and $pr_i(\varphi) > \alpha$, where α is a rational number in $[0, 1]$; $pr_i(\varphi) \geq \alpha$ can be read as "the probability of φ according to i is at least α", and similarly for $pr_i(\varphi) > \alpha$. We allow nesting here, so that we can have a formula of the form $pr_j(play_i(\sigma) \wedge pr_k(play_i(\sigma')) > 1/3) \geq 1/4$. As we would expect,

- $(M, \omega) \models pr_i(\varphi)$ iff $\mathcal{PR}_i(\omega)([\![\varphi]\!]_M) \geq \alpha$.

The restriction to α being rational allows the language to be countable. However, as we now show, it is not too serious a restriction.

Let $\mathcal{L}^4(\Gamma)$ be the language that extends $\mathcal{L}^2(\Gamma)$ by closing off under countable conjunctions, so that if $\varphi_1, \varphi_2, \ldots$ are formulas, then so is $\wedge_{m=1}^{\infty}\varphi_m$, and formulas of the form $pr_i(\varphi) > \alpha$, where α is a real number in $[0,1]$. (We can express $pr_i(\varphi) \geq \alpha$ as the countable conjunction $\wedge_{\beta<\alpha, \beta\in Q\cap[0,1]} pr_i(\varphi) > \beta$, where Q is the set of rational numbers, so there is no need to include formulas of the form $pr_i(\varphi) \geq \alpha$ explicitly in $\mathcal{L}^4(\Gamma)$.) We omit the parenthetical Γ in $\mathcal{L}^3(\Gamma)$ and $\mathcal{L}^4(\Gamma)$ when the game Γ is clear from context. A subset Φ of $\mathcal{L}^3$ is $\mathcal{L}^3$-*realizable* if there exists an appropriate structure M for Γ and state ω in M such that, for all formulas $\varphi \in \mathcal{L}^3$, $(M, \omega) \models \varphi$ iff $\varphi \in \Phi$.[1] We can similarly define what it means for a subset of $\mathcal{L}^4$ to be $\mathcal{L}^4$-realizable.

Lemma 3.1: *Every $\mathcal{L}^3$-realizable set can be uniquely extended to an $\mathcal{L}^4$-realizable set.*

Proof: It is easy to see that every $\mathcal{L}^3$-realizable set can be extended to an $\mathcal{L}^4$-realizable set. For suppose that Φ is $\mathcal{L}^3$-realizable. Then there is some state ω and structure M such that, for every formula $\varphi \in \mathcal{L}^3$, we have that $(M, \omega) \models \varphi$ iff $\varphi \in \Phi$. Let Φ' consist of the $\mathcal{L}^4$ formulas true at ω. Then clearly Φ' is an $\mathcal{L}^4$-realizable set that extends Φ.

To show that the extension is unique, suppose that there are two $\mathcal{L}^4$-realizable sets, say Φ_1 and Φ_2, that extend Φ. We want to show that $\Phi_1 = \Phi_2$. To do this, we consider two language, $\mathcal{L}^5$ and $\mathcal{L}^6$, intermediate between $\mathcal{L}^3$ and $\mathcal{L}^4$.

Let $\mathcal{L}^5$ be the language that extends $\mathcal{L}^2$ by closing off under countable conjunctions and formulas of the form $pr_i(\varphi) > \alpha$, where α is a rational number in $[0,1]$. Thus, in $\mathcal{L}^5$, we have countable conjunctions and disjunctions, but can talk explicitly only about rational probabilities. Nevertheless, it is easy to see that for every formula $\varphi \in \mathcal{L}^4$, there is an formula equivalent formula $\varphi' \in \mathcal{L}^5$, since if α is a real number, then $pr_i(\varphi) > \alpha$ is equivalent to $\vee_{\beta>\alpha, \beta\in[0,1]\cap Q}\, pr_i(\varphi) > \beta$ (an infinite disjunction $\vee_{i=1}^{\infty}\varphi_i$ can be viewed as an abbreviation for $\neg \wedge_{i=1}^{\infty} \neg\varphi_i$).

Next, let $\mathcal{L}^6$ be the result of closing off formulas in $\mathcal{L}^3$ under countable conjunction and disjunction. Thus, in $\mathcal{L}^6$, we can apply countable conjunction and disjunction only at the outermost level, not inside the scope of pr_i. We claim that for every formula $\varphi \in \mathcal{L}^5$, there is an equivalent formula in $\mathcal{L}^6$. More precisely, for every formula $\varphi \in \mathcal{L}^5$, there exist formulas $\varphi_{ij} \in \mathcal{L}^3$, $1 \leq i, j < \infty$ such that φ is equivalent to $\wedge_{m=1}^{\infty} \vee_{n=1}^{\infty} \varphi_{mn}$. We prove this by induction on the structure of φ. If φ is RAT_i, $play_i(\sigma)$, or *true*, then the statement is clearly true. The result is immediate from the induction hypothesis if φ is a countable conjunction. If φ has the form $\neg\varphi'$, we apply the induction hypothesis, and observe that $\neg(\wedge_{m=1}^{\infty} \vee_{n=1}^{\infty} \varphi_{mn})$ is equivalent to $\vee_{m=1}^{\infty} \wedge_{n=1}^{\infty} \neg\varphi_{mn}$. We can convert this to a conjunction of disjunctions by distributing the disjunctions over the conjunctions in the standard way (just as $(E_1 \cap E_2) \cup (E_3 \cap E_4)$ is equivalent to $(E_1 \cup E_3) \cap (E_1 \cup E_4) \cap (E_2 \cup E_3) \cap (E_2 \cup E_4)$). Finally, if φ has the form $pr_i(\varphi') > \alpha$, we apply the induction hypothesis, and observe that $pr_i(\wedge_{m=1}^{\infty} \vee_{n=1}^{\infty} \varphi_{mn}) > \alpha$ is equivalent to

$$\vee_{\alpha'>\alpha, \alpha'\in Q\cap[0,1]} \wedge_{M=1}^{\infty} \vee_{N=1}^{\infty} pr_i(\wedge_{m=1}^{M} \vee_{n=1}^{N} \varphi_{mn}) > \alpha'.$$

The desired result follows, since if two states agree on all formulas in $\mathcal{L}^3$, they must agree on all formulas in $\mathcal{L}^6$, and hence on all formulas in $\mathcal{L}^5$ and $\mathcal{L}^4$. ∎

[1]For readers familiar with standard completeness proofs in modal logic, if we had axiomatized the logic we are implicitly using here, the $\mathcal{L}^3$-realizable sets would just be the maximal consistent sets in the logic.

The choice of language turns out to be significant for a number of our results; we return to this issue at various points below.

With this background, we can define strong admissibility. Let $\mathcal{L}_i^3$ consist of all formulas in $\mathcal{L}^3$ of the form $pr_i(\varphi) \geq \alpha$ and $pr_i(\varphi) > \alpha$ (φ can mention pr_i; it is only the outermost modal operator that must be i). Intuitively, $\mathcal{L}_i^3$ consists of the formulas describing i's beliefs. Let $\mathcal{L}_{i+}^3$ consist of $\mathcal{L}_i^3$ together with formulas of the form *true*, RAT_i, and $play_i(\sigma)$, for $\sigma \in \Sigma_i$. Let $\mathcal{L}_{(-i)+}^3$ be an abbreviation for $\cup_{j \neq i} \mathcal{L}_{j+}^3$. We can similarly define $\mathcal{L}_i^4$ and $\mathcal{L}_{i+}^4$.

If $\varphi \in \mathcal{L}_{(-i)+}^3$, define $O_i\varphi$, read "all agent i knows (with respect to $\mathcal{L}^3$) is φ," as an abbreviation for the $\mathcal{L}^4$ formula

$$B_i\varphi \wedge (\wedge_{\psi \in \mathcal{L}_{(-i)+}^3} \Diamond(\varphi \wedge \psi) \Rightarrow \langle B_j \rangle \psi).$$

Thus, $O_i\varphi$ holds if agent i believes φ but does not know anything beyond that; he ascribes positive probability to all formulas in $\mathcal{L}_{(-i)+}^3$ consistent with φ. This is very much in the spirit of the Halpern-Lakemeyer [2001] definition of O_i in the context of epistemic logic.

Of course, we could go further and define a notion of "all i knows" for the language $\mathcal{L}^4$. Doing this would give a definition that is even closer to that of Halpern and Lakemeyer. Unfortunately, we cannot require than agent i ascribe positive probability to all the formulas in $\mathcal{L}_{(-i)+}^4$ consistent with φ; in general, there will be an uncountable number of distinct and mutually exclusive formulas consistent with φ, so they cannot all be assigned positive probability. This problem does not arise with $\mathcal{L}^3$, since it is a countable language. Halpern and Lakemeyer could allow an agent to consider an uncountable set of worlds possible, since they were not dealing with probabilistic systems. This stresses the point that the notion of "all I know" is quite sensitive to the choice of language.

Define the formulas F_i^k inductively by taking F_i^0 to be the formula *true*, and F_i^{k+1} to an abbreviation of $RAT_i \wedge O_i(\wedge_{j \neq i} F_j^k)$. Thus, F_j^{k+1} says that i is rational, believes that all the other players satisfy level-k rationality (i.e., F_j^k), and that is all that i knows. An easy induction shows that F_j^{k+1} implies that j is rational and j believes that everyone believes (k times) that everyone is rational. Moreover, it is easy to see that F_j^{k+1} implies D_j^{k+1}. The difference is that instead of requiring just that j assign positive probability to all strategy profiles compatible with F_{-j}^k, it requires that j assign positive probability to all formulas compatible with F_{-j}^k.

A strategy σ_i for player i is *kth-level strongly admissible* if it is consistent with F_i^k; that is, if $play_i(\sigma_i) \wedge F_i^k$ is satisfied in some state. The next result shows that strong admissibility characterizes iterated deletion, just as admissibility does.

Theorem 3.2: *The following are equivalent:*

(a) the strategy σ for player i survives k rounds of iterated deletion of weakly dominated strategies;

(b) for all $k' \leq k$, there is a measurable structure $M^{k'}$ appropriate for Γ and a state $\omega^{k'}$ in $M^{k'}$ such that $\mathbf{s}_i(\omega^{k'}) = \sigma$ and $(M^{k'}, \omega^{k'}) \models F_i^{k'}$;

(c) for all $k' \leq k$, there is a structure $M^{k'}$ appropriate for Γ and a state $\omega^{k'}$ in $M^{k'}$ such that $\mathbf{s}_i(\omega^{k'}) = \sigma$ and $(M^{k'}, \omega^{k'}) \models F_i^{k'}$;

Proof: The proof is similar in spirit to the proof of Theorem 2.7. We again proceed by induction on k. The result clearly holds for $k = 0$. If $k = 1$, the proof that (c) implies (a) is essentially identical to that of Theorem 2.7; we do not repeat it here.

To prove that (a) implies (b), we need the following three lemmas; the first shows that a formula is always satisfied in a state that has probability 0; the second shows that that we can get a new structure with a world where agent i ascribes positive probability to each of a countable collection of satisfiable formulas in $\mathcal{L}^3_{-i}$; and the third shows that formulas in $\mathcal{L}^4_{i+}$ for different players i are independent; that is, if $\varphi_i \in \mathcal{L}^4_{i+}$ is satisfiable, then so is $\varphi_1 \wedge \ldots \wedge \varphi_n$.

Lemma 3.3: *If $\varphi \in \mathcal{L}^4$ is satisfiable in a measurable structure, then there exists a measurable structure M and state ω such that $(M, \omega) \models \varphi$, $\{\omega\}$ is measurable, $\mathcal{PR}_j(\omega)(\{\omega\}) = 0$ for $j = 1, \ldots, n$.*

Proof: Suppose that $(M', \omega') \models \varphi$, where $M' = (\Omega', \mathbf{s}', \mathcal{F}', \mathcal{PR}'_1, \ldots, \mathcal{PR}'_n)$. Let $\Omega = \Omega' \cup \{\omega\}$, where where ω is a fresh state; let $\mathcal{F}$ be the smallest σ-algebra that contains $\mathcal{F}$ and $\{\omega\}$; let $\mathbf{s}$ and $\mathcal{PR}_j$ agree with $\mathbf{s}'$ and $\mathcal{PR}'_j$ when restricted to states in Ω' (more precisely, if $\omega'' \in \Omega'$, then $\mathcal{PR}_j(\omega'')(A) = \mathcal{PR}'_j(\omega'')(A \cap \Omega')$ for $j = 1, \ldots, n$). Finally, define $\mathbf{s}_i(\omega) = \mathbf{s}_i(\omega')$, and take $\mathcal{PR}_j(\omega)(A) = \mathcal{PR}'_j(\omega')(A \cap \Omega')$ for $j = 1, \ldots, n$. Clearly $\{\omega\}$ is measurable, and $\mathcal{PR}_j(\omega)(\{\omega\}) = 0$ for $j = 1, \ldots, n$. An easy induction on structure shows that for all formulas ψ, (a) $(M, \omega) \models \psi$ iff $(M, \omega') \models \psi$, and (b) for all states $\omega'' \in \Omega'$, we have that $(M, \omega'') \models \psi$ iff $(M', \omega'') \models \psi$. It follows that $(M, \omega) \models \varphi$, and that M is measurable. ∎

Lemma 3.4: *Suppose that $\vec{\sigma} \in \vec{\Sigma}$, Φ' is a countable collection of formulas in $\mathcal{L}^4_{-i}$, $\varphi \in \mathcal{L}^4_{-i}$, and Σ'_{-i} is a set of strategy profiles in Σ_{-i} such that (a) for each formula $\varphi' \in \Phi'$, there exists some profile $\sigma_{-i} \in \Sigma'_{-i}$ such that $\varphi \wedge \varphi' \wedge play_{-i}(\sigma_{-i})$ is satisfied in a measurable structure, and (b) for each profile $\sigma_{-i} \in \Sigma'_{-i}$, $play_{-i}(\sigma_{-i})$ is one of the formulas in Φ'. Then there exists a measurable structure M and state ω such that $\mathbf{s}(\omega) = \vec{\sigma}$, $(M, \omega) \models play_{-j}(\sigma_{-i}) \geq \alpha$ iff $\mu_j(\sigma_{-i}) \geq \alpha$ (that is, μ_{-i} agrees with $\mathcal{PR}_i(\omega)$ when marginalized to strategy profiles in Σ'_{-i}), and $(M, \omega) \models B_i\varphi \wedge \langle B_i \rangle \varphi'$ for all $\varphi' \in \Phi'$.*

Proof: Let Φ' and Σ'_{-i} be as in the statement of the lemma. Suppose that $\Phi' = \{\varphi_1, \varphi_2, \ldots, \ldots\}$. By assumption, for each formula $\varphi_k \in \Phi'$, there exists some strategy profile $\sigma'_{-i} \in \Sigma'_{-i}$, measurable structure $M^k = (\Omega^k, \mathbf{s}^k, \mathcal{F}^k, \mathcal{PR}^k_1, \ldots, \mathcal{PR}^k_n)$, and $\omega^k \in \Omega^k$ such that $(M^k, \omega^k) \models \varphi \wedge \varphi_k \wedge play_{-i}(\sigma'_{-i})$, for $k = 1, 2, \ldots$. By Lemma 3.3, we can assume without loss of generality that $\{\omega^k\} \in \mathcal{F}^k$ and $\mathcal{PR}^k_j(\omega^k)(\{\omega^i\}) = 0$. Define $M^\infty = (\Omega^\infty, \mathbf{s}^\infty, \mathcal{F}^\infty, \mathcal{PR}^\infty_1, \ldots, \mathcal{PR}^\infty_n)$ as follows:

- $\Omega^\infty = \cup_{k=0}^\infty \Omega^k \cup \{\omega\}$, where ω is a fresh state;
- $\mathcal{F}^\infty$ is the smallest σ-algebra that contains $\{\omega\} \cup \mathcal{F}_1 \cup \mathcal{F}_2 \cup \ldots$;
- $\mathbf{s}^\infty$ agrees with $\mathcal{PR}^k_j$ when restricted to states in Ω^k, except that $\mathbf{s}^\infty_i(\omega^k) = \sigma_i$ and $\mathbf{s}^\infty(\omega) = \vec{\sigma}$;
- $\mathcal{PR}^\infty_j$ agrees with $\mathcal{PR}^k_j$ when restricted to states in Ω^k (more precisely, if $\omega' \in \Omega^k$, then $\mathcal{PR}^\infty_j(\omega')(A) = \mathcal{PR}^k_j(\omega')(A \cap \Omega^k)$, except that $\mathcal{PR}^\infty_i(\omega) = \mathcal{PR}^\infty_i(\omega^1) = \mathcal{PR}_i(\omega^2) = \cdots$ is defined to be a distribution with support $\{\omega^1, \omega^2, \ldots\}$ (so that all these states are given positive probability) such that $\mathcal{PR}^\infty_i(\omega)$ agrees μ when marginalized to profiles in Σ_{-i}, and $\mathcal{PR}^\infty_j(\omega)(\{\omega\}) = 1$ for $j \neq i$. It is easy to see that our assumptions guarantee that this can be done.

We can now prove by a straightforward induction on the structure of ψ that (a) for all formulas ψ, $k = 1, 2, 3, \ldots$, and states $\omega' \in \Omega^k - \{\omega^k\}$, we have that $(M^k, \omega') \models \psi$ iff $(M^\infty, \omega') \models \psi$; and

(b) for all formulas $\psi \in \mathcal{L}^4_{(-i)+}$, $k = 1, 2, 3, \ldots$, and $(M^k, \omega^k) \models \psi$ iff $(M^\infty, \omega^k) \models \psi$. (Here it is important that $\mathcal{PR}^\infty_j(\omega^k) = \mathcal{PR}^k_j(\omega) = 0$ for $j \neq i$; this ensures that j's beliefs about i's strategies and beliefs unaffected by the fact that $\mathbf{s}^k_i(\omega^k) \neq \mathbf{s}^\infty_i(\omega^k)$ and $\mathcal{PR}^k_i(\omega^k) \neq \mathcal{PR}^\infty_i(\omega^k)$.) It easily follows that $(M^\infty, \omega) \models B_i\varphi \wedge \langle B_i \rangle \varphi'$ for all $\varphi' \in \Phi'$. ∎

Lemma 3.5: *If $\varphi_i \in \mathcal{L}^4_{i+}$ is satisfiable for $i = 1, \ldots, n$, then $\varphi_1 \wedge \ldots \wedge \varphi_n$ is satisfiable.*

Proof: Suppose that $(M^i, \omega^i) \models \varphi_i$, where $M^i = (\Omega^i, \mathbf{s}^i, \mathcal{F}^i, \mathcal{PR}^i_1, \ldots, \mathcal{PR}^i_n)$ and $\varphi_i \in \mathcal{L}^4_{i+}$. By Lemma 3.3, we again assume without loss of generality that $\{\omega^i\} \in \mathcal{F}^i$ and $\mathcal{PR}_j(\omega^i)(\{\omega^i\}) = 0$. Let $M^* = (\Omega^*, \mathbf{s}^*, \mathcal{F}^*, \mathcal{PR}^*_1, \ldots, \mathcal{PR}^*_n)$, where

- $\Omega^* = \cup_{i=1}^n \Omega^i$;
- $\mathcal{F}^*$ is the smallest σ-algebra containing $\mathcal{F}^1 \cup \ldots \cup \mathcal{F}^n$;
- $\mathbf{s}^*$ agrees with $\mathbf{s}^j$ on states in Ω^j except that $\mathbf{s}^*_i(\omega^j) = \mathbf{s}^i_i(\omega^i)$ (so that $\mathbf{s}^*(\omega^1) = \cdots = \mathbf{s}^*(\omega^n)$);
- $\mathcal{PR}^*_i$ agrees with $\mathcal{PR}^j_i$ on states in Ω^j except that $\mathcal{PR}^*_i(\omega^j) = \mathcal{PR}^i_i(\omega^i)$ (so that $\mathcal{PR}^*_i(\omega^1) = \cdots = \mathcal{PR}^*_i(\omega^n) = \mathcal{PR}^i_i(\omega^i)$).

We can now prove by induction on the structure of ψ that (a) for all formulas ψ, $i = 1, \ldots, n$, and states $\omega' \in \Omega^i$, we have that $(M^i, \omega') \models \psi$ iff $(M^*, \omega') \models \psi$; (b) for all formulas $\psi \in \mathcal{L}^4_{i+}$, $1 \leq i, j \leq n$, $(M^i, \omega^i) \models \psi$ iff $(M^*, \omega^j) \models \psi$ (again, here it is important that $\mathcal{PR}^*_i(\omega^j) = 0$ for $j = 1, \ldots, n$). Note that part (b) implies that the states $\omega^1, \ldots, \omega^n$ satisfy the same formulas in M^*. It easily follows that $(M^*, \omega^i) \models \varphi_1 \wedge \ldots \wedge \varphi_n$ for $i = 1, \ldots, n$. ∎

We can now prove the theorem. Again, let X^k_j be the strategies for player j that survive k rounds of iterated deletion of weakly dominated strategies. To see that (a) implies (b), suppose that $\sigma_i \in X^{k+1}_i$. By Proposition 2.6, there exists a distribution μ_i whose support is X^k_{-i} such that σ_i is a best response to μ_i. By the induction hypothesis, for each strategy profile $\tau_{-i} \in X^k_{-i}$, and all $j \neq i$, the formula $play_j(\tau_j) \wedge F^0_j$ is satisfied in a measurable structure. By Lemma 3.5, $play_{-j}(\tau_{-j}) \wedge (\wedge_{j \neq i} F^k_j)$ is satisfied in a measurable structure. Taking φ to be $\wedge_{j \neq i} F^k_j$, by Lemma 3.4, there exists a measurable structure M and state ω in M such that the marginal of $\mathcal{PR}_i(\omega)$ on X^k_{-i} is μ_i, $\mathbf{s}_i(\omega)$ is σ_i, and $(M, \omega) \models B_i(\wedge_{j \neq i} F^k_j) \wedge (\wedge_{\psi \in \mathcal{L}^3_{(-j)+}} \Diamond(\psi \wedge (\wedge_{j \neq i} F^k_j)) \Rightarrow \langle B_j \rangle \psi)$. It follows that $(M, \omega) \models RAT_i$, and hence that $(M, \omega) \models F^{k+1}_i$, as desired.

It is immediate that (b) implies (c). ∎

Corollary 3.6: *The following are equivalent:*

(a) the strategy σ for player i survives iterated deletion of weakly dominated strategies;

(b) there is a measurable structure M that is appropriate for Γ and a state ω such that $\mathbf{s}_i(\omega) = \sigma$ and $(M, \omega) \models \langle B_i \rangle F^k_i$ for all $k \geq 0$;

(c) there is a structure M that is appropriate for Γ and a state ω such that $\mathbf{s}_i(\omega) = \sigma$ and $(M, \omega) \models \langle B_i \rangle F^k_i$ for all $k \geq 0$.

Proof: The proof is essentially identical to that of Corollary 2.8, so is omitted here. ∎

4 Complete and Canonical Structures

4.1 Canonical Structures

Intuitively, to check whether a formula is strongly admissible, and, more generally, to check if all agent i knows is φ, we want to start with a very rich structure M that contains all possible consistent sets of formulas, so that if $\varphi \wedge \psi$ is satisfied at all, it is satisfied in that structure. Motivated by this intuition, Halpern and Lakemeyer [2001] worked in the *canonical* structure for their language, which contains a state corresponding to every consistet set of formulas. We do the same thing here.

Define the canonical structure $M^c = (\Omega^c, \mathbf{s}^c, \mathcal{F}^c, \mathcal{PR}_1^c, \ldots, \mathcal{PR}_n^c)$ for $\mathcal{L}^4$ as follows:

- $\Omega^c = \{\omega_\Phi : \Phi$ is a realizable subset of $\mathcal{L}^4(\Gamma\}$;
- $\mathbf{s}^c(\omega_\Phi) = \vec{\sigma}$ iff $play(\sigma) \in \Phi$;
- $\mathcal{F}^c = \{F_\varphi : \varphi \in \mathcal{L}^4\}$, where $F_\varphi = \{\omega_\Phi : \varphi \in \Phi\}$;
- $\Pr_i^c(\omega_\Phi)(F_\varphi) = \inf\{\alpha : pr_i(\varphi) > \alpha \in \Phi\}$.

Lemma 4.1: *M^c is an appropriate measurable structure for Γ.*

Proof: It is easy to see that $\mathcal{F}^c$ is a σ-algebra, since the complement of F_φ is $F_{\neg\varphi}$ and $\cap_{m=1}^\infty F_{\varphi_i} = F_{\wedge_{m=1}^\infty \varphi_m}$. Given a strategy σ for player i, $[\![\sigma]\!]_{M^c} = F_{play_i(\sigma)} \in \mathcal{F}$. Moreover, each realizable set Φ that includes $play_i(\sigma)$ must also include $pr_i(play_i(\sigma)) = 1$, so that $\mathcal{PR}_i(\omega_\Phi)(\mathbf{s}_i(\omega_\Phi)) = \mathcal{PR}_i(\omega_\Phi)(F_{play_i(\mathbf{s}_i(\omega_\Phi))}) = 1$. Similarly, suppose that $\mathcal{PR}_i(\omega_\Phi) = \pi$. Then $\{\omega \in \Omega^c : \mathcal{PR}_i(\omega) = \pi\} = \cap_{\varphi \in \mathcal{L}^3} \cap_{\{\alpha \in Q \cap [0,1] : \pi([\![\varphi]\!]_{M^c}) \geq \alpha\}} F_{\varphi \geq \alpha} \in \mathcal{F}^c$. Moreover, if $\alpha \in Q \cap [0,1]$, then $\pi([\![\varphi]\!]_{M^c}) \geq \alpha$ iff $pr_i(\varphi) \geq \alpha \in \Phi$. But if $pr_i(\varphi) \geq \alpha \in \Phi$, then $pr_i(pr_i(\varphi) \geq \alpha) = 1 \in \Phi$. It easily follows that $\mathcal{PR}_i(\omega_\Phi)(\{\omega : \mathcal{PR}_i(\omega) = \pi\}) = 1$. Finally, the definition of $\mathcal{F}^c$ guarantees that every set $[\![\varphi]\!]_{M^c}$ is measurable and that $\mathcal{PR}_i(\omega_\Phi)$ is indeed a probability distribution on $(\Omega^c, \mathcal{F}^c)$. ∎

The following result is the analogue of the standard "truth lemma" in completeness proofs in modal logic.

Proposition 4.2: *For $\psi \in \mathcal{L}^4$, $(M^c, \omega_\Phi) \models \psi$ iff $\psi \in \Phi$.*

Proof: A straightforward induction on the structure of ψ. ∎

We have constructed a canonical structure for $\mathcal{L}^4$. It follows easily from Lemma 3.1 that the canonical structure for $\mathcal{L}^3$ (where the states are realizable $\mathcal{L}^3$ sets) is isomorphic to M^c. (In this case, the set $\mathcal{F}^c$ of measurable sets would be the smallest σ-algebra containing $[\![\varphi]\!]_M$ for $\varphi \in \mathcal{L}^3$.) Thus, the choice of $\mathcal{L}^3$ vs. $\mathcal{L}^4$ does not play an important role when constructing a canonical structure.

A strategy σ_i for player i survives iterated deletion of weakly dominated strategies iff the $\mathcal{L}^4$ formula $undominated(\sigma_i) = play_i(\sigma_i) \wedge (\wedge_{k=1}^\infty \langle B_i \rangle F_i^k)$ is satisfied at some state in the canonical structure. But there are other structures in which $undominated(\sigma_i)$ is satisfied. One way to get such a struture is by essentially "duplicating" states in the canonical structure. The canonical structure can be *embedded* in a structure M if, for all $\mathcal{L}^3$-realizable sets Φ, there is a state ω_Φ in M such that $(M, \omega_\Phi) \models \varphi$ iff

$\varphi \in \Phi$. Clearly $undominated(\sigma_i)$ is satisfied in any structure in which the canonical structure can be embedded.

A structure in which the canonical structure can be embedded is in a sense larger than the canonical structure. But $undominated(\sigma_i)$ can be satisfied in structures smaller than the canonical structure. (Indeed, with some effort, we can show that it is satisfiable in a structure with countably many states.) There are two reasons for this. The first is that to satisfy $undominated(\sigma_i)$, there is no need to consider a structure with states where all the players are irrational. It suffices to restrict to states where at least one player is using a strategy that survives at least one round of iterated deletion. This is because players know their strategy; thus, in a state where a strategy σ_j for player j is admissible, player j must ascribe positive probability to all other strategies; however, in those states, player j still plays σ_j.

A perhaps more interesting reason that we do not need the canonical structure is our use of the language $\mathcal{L}_3$. Strong admissibility guarantees that player j will ascribe positive probability to all formulas φ consistent with rationality. Since a finite conjunction of formulas in $\mathcal{L}^3$ is also a formula in $\mathcal{L}^3$, player j will ascribe positive probability to all finite conjunctions of formulas consistent with rationality. But a state is characterized by a *countable* conjunction of formulas. Since $\mathcal{L}^3$ is not closed under countable conjunctions, a structure that satisfies $undominated(\sigma_i)$ may not have states corresponding to all L^3-realizable sets of formulas. If we had used $\mathcal{L}^4$ instead of $\mathcal{L}^3$ in the definition of strong admissibility (ignoring the issues raised earlier with using $\mathcal{L}^4$), then there would be a state corresponding to every $\mathcal{L}^4$-realizable (equivalently, $\mathcal{L}^3$-realizable) set of formulas. Alternatively, if we consider appropriate structures that are compact in a topology where all sets definable by formulas (i.e., sets of the form $[\![\varphi]\!]_M$, for $\varphi \in \mathcal{L}^3$) are closed (in which case they are also open, since $[\![\neg\varphi]\!]_M$ is the complement of $[\![\varphi]\!]_M$), then all states where at least one player is using a strategy that survives at least one round of iterated deletion will be in the structure.

Although, as this discussion makes clear, the formula that characterizes strong admissibility can be satisfied in structures quite different from the canonical structure, the canonical structure does seem to be the most appropriate setting for reasoning about statements involving "all agent i knows", which is at the heart of strong admissibility. Moreover, as we now show, canonical structures allow us to relate our approach to that of BFK.

4.2 Complete Structures

BFK worked with complete structures. We now want to show that M^c is complete, in the sense of BFK. To make this precise, we need to recall some notions from BFK (with some minor changes to be more consistent with our notation).

BFK considered what they called *interactive probability structures*. These can be viewed as a special case of probability structures. A *BFK-like structure* (for a game Γ) is a probability structure $M = (\Omega, \mathbf{s}, \mathcal{F}, \mathcal{PR}_1, \ldots, \mathcal{PR}_n)$ such that there exist spaces $T_1, \ldots, T_n$ (where T_i can be thought of as the *type space* for player i) such that

- Ω is isomorphic to $\vec{\Sigma} \times \vec{T}$, via some isomorphism h;
- if $h(\omega) = \vec{\sigma} \times \vec{t}$, then
  - $\mathbf{s}(\omega) = \vec{\sigma}$,

– taking $T_i(\omega) = t_i$ (i.e., the type of player i in $h(\omega)$ is t_i); the support of $\mathcal{PR}_i(\omega)$ is contained in $\{\omega' : \mathbf{s}_i(\omega') = \sigma', T_i(\omega') = t_i\}$, so that $\mathcal{PR}_i(\omega)$ induces a probability on $\Sigma_{-i} \times T_{-i}$;
– $\mathcal{PR}_i(\omega)$ depends only on $T_i(\omega)$, in the sense that if $T_i(\omega) = T_i(\omega')$, then $\mathcal{PR}_i(\omega)$ and $\mathcal{PR}_i(\omega')$ induce the same probability distribution on $\Sigma_{-i} \times T_{-i}$.

A BFK-like structure M whose state space is isomorphic to $\vec{\Sigma} \times \vec{T}$ is *complete* if, for every for each distribution μ_i over $\Sigma_{-i} \times T_{-i}$, there is a state ω in M such that the probability distribution on $\Sigma_{-i} \times T_{-i}$ induced by $\mathcal{PR}_i(\omega)$ is μ_i.

Proposition 4.3: *M^c is complete BFK-like structure.*

Proof: A set $\Phi \subseteq \mathcal{L}_i^3$ is *$\mathcal{L}_i^3$-realizable* if there exists an appropriate structure M for Γ and state ω in M such that, for all formulas $\varphi \in \mathcal{L}^3$, $(M, \omega) \models \varphi$ iff $\varphi \in \Phi$. Take the type space T_i to consist of all $\mathcal{L}_i^3$-realizable sets of formulas. There is an isomorphism h between Ω^c and $\vec{\Sigma} \times \vec{T}$, where $T_i(\omega)$ is the i-realizable type of formulas of the form $pr_i(\varphi) \geq \alpha$ that are true at ω; that is, $h(\omega) = \mathbf{s}(\omega) \times T_1(\omega) \times \cdots \times T_n(\omega)$. It follows easily from Lemma 3.5 that h is a surjection. we can identify Ω^c, the state space in the canonical structure, with $\vec{\Sigma} \times \vec{T}$.

To prove that M^c is complete, given a probability μ on $\Sigma_{-i} \times T_{-i}$, we must show that there is some state ω in M^c such that the probability induced by $\mathcal{PR}_i(\omega)$ on $\Sigma_{-i} \times T_{-i}$ is μ. Let $M^\mu = (\Omega^{sigma,\mu}, \mathcal{F}^\mu, \mathbf{s}^\mu, \mathcal{PR}_1^\mu, \ldots, \mathcal{PR}_n^\mu)$, where M^μ are defined as follows:

- $\Omega^\mu = \Omega^c \cup \Sigma \times \{\mu\} \times T_{-i}$;
- $\mathcal{F}^\mu$ is the smallest σ-algebra that contains $\mathcal{F}^c$ and all sets of the form $\vec{\sigma} \times \{\mu\} \times [\![\varphi]\!]'_{M^c}$, and $[\![\varphi]\!]'_{M^c}$ consists of the all type profiles t_{-i} such that, for some state ω in M^c, $(M^c, \varphi) \models \varphi$ and $T_{-i}(\varphi) = t_{-i}$;
- $\mathbf{s}^\mu(\omega) = \mathbf{s}^c(\omega)$ for $\omega \in \Omega^c$, and $\mathbf{s}^\mu(\vec{\sigma} \times \{\mu\} \times \vec{t}) = \vec{\sigma}$;
- $\mathcal{PR}_j^\mu(\omega) = \mathcal{PR}_j^c(\omega)$ for $\omega \in \Omega^c$, $j = 1, \ldots, n$; for $j \neq i$, $\mathcal{PR}_j^\mu(\vec{\sigma} \times \mu \times t_{-i}) = \mathcal{PR}_j(\omega)$, where $\mathbf{s}_j(\omega) = \sigma_j$ and $T_j(\omega) = t_j$ (this is well defined, since if $\mathbf{s}_j(\omega') = \sigma_j$ and $T_j(\omega') = t_j$, then $\mathcal{PR}_j(\omega) = \mathcal{PR}_j(\omega')$; finally, $\mathcal{PR}_i^\mu(\vec{\sigma} \times \mu \times t_{-i})$ is a distribution whose support is contained in $\{\sigma_i\} \times \Sigma_{-i} \times \{\mu\} \times T_{-i}$, and $\mathcal{PR}_i^\mu(\vec{\sigma} \times \mu \times t_{-i})(\vec{\sigma} \times \mu \times [\![\varphi]\!]'_{M^c}) = \mu([\![\varphi]\!]'_{M^c})$.

Choose an arbitrary state $\omega \in \vec{\Sigma} \times \{\mu\} \times T_{-i}$. The construction of M^μ guarantees that for $\varphi \in \mathcal{L}^4_{(-i)+}$, $(M^\mu, \omega) \models pr_i(\varphi) > \alpha$ iff $\mu([\![\varphi]\!]'_{M^c}) > \alpha$. By the construction of M^c, there exists a state $\omega' \in \Omega^c$ such that $(M^c, \omega') \models \psi$ iff $(M^\mu, \omega) \models \psi$. Thus, the distribution on $\Sigma_{-i} \times T_{-i}$ induced by $\mathcal{PR}_i(\omega)$ is μ, as desired. This shows that M^c is complete. ∎

We now would like to show that every measurable complete BFK-like structure is the canonical model. This is not quite true because states can be duplicated in an interactive structure. This suggests that we should try to show that the canonical structure can be embedded in every measurable complete structure. We can essentially show this, except that we need to restrict to *strongly measurable* complete structures, where a structure is strongly measurable if it is measurable and the only measurable sets are those defined by $\mathcal{L}_4$ formulas (or, equivalently, the set of measurable sets is the smallest set that contains the sets defined by $\mathcal{L}_3$ formulas). We explain where strong measurability is needed at the end of the proof of the following theorem.

Theorem 4.4: *If M is a strongly measurable complete BFK-like structure, then the canonical structure can be embedded in M.*

Proof: Suppose that M is a strongly measurable complete BFK-like structure. We can assume without loss of generality that the state space of M has the form $\vec{\Sigma} \times \vec{T}$. To prove the result, we need the following lemmas.

Lemma 4.5: *If M is BFK-like, the truth of a formula $\varphi \in \mathcal{L}_i^4$ at a state ω in M depends only on i's type; That is, if $T_i(\omega) = T_i(\omega')$, then $(M, \omega) \models \varphi$ iff $(M, \omega') \models \varphi$. Similarly, the truth of a formula in $\mathcal{L}_{i+}$ in ω depends only on $\mathbf{s}_i(\omega)$ and $T_i(\omega)$, and the truth of a formula in $\mathcal{L}_{i+}^4$ in ω depends only on $T_{-i}(\omega)$.*

Proof: A straightforward induction on structure. ∎

Define a *basic formula* to be one of the form $\psi_1 \wedge \ldots \wedge \psi_n$, where $\psi_i \in \mathcal{L}_{i+}^3$ for $i = 1, \ldots, n$.

Lemma 4.6: *Every formula in $\mathcal{L}^3$ is equivalent to a finite disjunction of basic formulas.*

Proof: A straightforward induction on structure. ∎

Lemma 4.7: *Every formula in $\mathcal{L}_{i+}^3$ is equivalent to a disjunction of formulas of the form*

$$play_i(\sigma) \wedge (\neg) RAT_i \wedge (\neg) pr_i(\varphi_1) > \alpha_1 \wedge \ldots \wedge (\neg) pr_i(\varphi_m) > \alpha_m \wedge (\neg) pr_i(\psi_1) \geq \beta_1 \wedge \ldots \wedge (\neg) pr_i(\psi_{m'}) \geq \beta_{m'}, \tag{1}$$

where $\varphi_1, \ldots, \varphi_m, \psi_1, \ldots, \psi_{m'} \in \mathcal{L}_{(-i)+}^3$ and the "$(\neg)$" indicates that the presence of negation is optional.

Proof: A straightforward induction on the structure of formulas, using the observation that $\neg play_i(\sigma)$ is equivalent to $\vee_{\{sigma' \in \Sigma_i : \sigma' \neq \sigma\}} play_i(\sigma')$. ∎

Lemma 4.8: *If $\varphi \in \mathcal{L}^3$ is satisfiable, then $[\![\varphi]\!]_M \neq \emptyset$.*

Proof: By Lemma 4.6, it suffices to prove the result for the case that φ is a basic formula. By Lemma 4.7, it suffices to assume that the the "i-component" of the basic formula is a conjunction. We now prove the result by induction on the depth of nesting of the modal operator pr_i in φ. (Formally, define $D(\psi)$, the depth of nesting of pr_i's in ψ, by induction on the structure of ψ. if ψ has the form $play_j(\sigma)$, RAT_j, or *true*, then $D(\psi) = 0$; $D(\neg\psi) = D(\psi)$; $D(\psi_1 \wedge \psi_2) = \max(D(\psi_1), D(\psi_2))$; and $D(pr_i(\psi) > \alpha) = D(pr_i(\psi) \geq \alpha) = 1 + D(\psi)$.) Because the state space Ω of M is essentially a product space, by Lemma 4.5, it suffices to prove the result for formulas in $\mathcal{L}_{(i)+}^3$. It is clear that φ possibly puts constraints on what strategy i is using, the probability of strategy profiles in Σ_{-i}, and the probability of formulas that appear in the scope of pr_i in φ. If $M' = (\Omega', \mathbf{s}', \mathcal{F}', \mathcal{PR}_1', \ldots, \mathcal{PR}_n')$ is a structure and $\omega' \in \Omega'$, then $(M', \omega') \models \varphi$ iff $\mathbf{s}_i'(\omega')$ and $\mathcal{PR}_i'(\omega')$ satisfies these constraints. (We leave it to the reader to formalize this somewhat informal claim.) By the induction hypothesis, each formula in the scope of pr_i in φ that is assigned positive probability by $\mathcal{PR}_i(\omega')$ is satisfied in M. Since M is complete and measurable, there is a state ω in M such that $\mathbf{s}_i(\omega) = \mathbf{s}'_i(\omega')$ and $\mathcal{PR}_i(\omega)$ places the same constraints on formulas that appear in φ as $\mathcal{PR}_i$. We must have $(M, \omega) \models \varphi$. ∎

Returning to the proof of the theorem, suppose that $M = (\Omega, \mathbf{s}, \mathcal{F}, \mathcal{PR}_1, \ldots, \mathcal{PR}_n)$. Given a state $\omega \in \Omega^c$, we claim that there must be a state ω' in M such that $\mathbf{s}(\omega') = \mathbf{s}^c(\omega)$ and, for all $i = 1, \ldots, n$, $\mathcal{PR}_i^c(\omega)([\![\psi]\!]_{M^c}) = \mathcal{PR}_i(\omega')([\![\psi]\!]_M)$. to show this, because of Ω is a product space, and $\mathcal{PR}_i(\omega')$ depends only on $T_i(\omega')$, it suffices to show that, for each i, there exists a state ω_i in M such that, for each i, $\mathcal{PR}_i^c(\omega)([\![\psi]\!]_{M^c}) = \mathcal{PR}_i(\omega_i)([\![\psi]\!]_M)$. By Lemma 4.8, if $[\![\psi]\!]_{M^c} \neq \emptyset$, then $[\![\psi]\!]_M \neq \emptyset$. Thus, the existence of ω_i follows from the assumption that M is complete and strongly measurable.

Roughly speaking, To understand the need for strong measurability here, note that even without strong measurability, the argument above tells us that there exists an appropriate measure defined on sets of the form $[\![\varphi]\!]_M$ for φ in $\mathcal{L}^3_{(-i)+}$. We can easily extend μ to a measure μ' on sets of the form $[\![\varphi]\!]_M$ for φ in $\mathcal{L}^4_{(-i)+}$. However, if the set $\mathcal{F}$ of measurable sets in M is much richer than the sets definable by $\mathcal{L}^4$ formulas, it is not clear that we can extend μ' to a measure on all of $\mathcal{F}$. In general, a countably additive measure defined on a subalgebra of a set $\mathcal{F}$ of measurable sets cannot be extended to $\mathcal{F}$. For example, it is known that, under the continuum hypothesis, Lebesgue measure defined on the Borel sets cannot be extended to all subsets of $[0, 1]$ [Ulam 1930]; see [Keisler and Tarski 1964] for further discussion). Strong measurability allows us to avoid this problem. ∎

References

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

Halpern, J. Y. and G. Lakemeyer (2001). Multi-agent only knowing. *Journal of Logic and Computation 11*(1), 41–70.

Keisler, J. and A. Tarski (1964). From accessible to inaccessible cardinals. *Fundamenta Mathematica 53*, 225–308.

Levesque, H. J. (1990). All I know: a study in autoepistemic logic. *Artificial Intelligence 42*(3), 263–309.

Mas-Colell, A., M. Whinston, and J. Green (1995). *Microeconomic Thoery*. Oxford, U.K.: Oxford University Press.

Osborne, M. J. and A. Rubinstein (1994). *A Course in Game Theory*. Cambridge, Mass.: MIT Press.

Pearce, D. G. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica 52*(4), 1029–1050.

Rajan, U. (1998). Trembles in the Bayesian foundation of solution concepts. *Journal of Economic Theory 82*, 248–266.

Tan, T. and S. Werlang (1988). The Bayesian foundation of solution concepts of games. *Journal of Economic Theory 45*(45), 370–391.

Ulam, S. (1930). Zur masstheorie in der allgemeinen mengenlehre. *Fundamenta Mathematicae 16*, 140–150.