# Epistemic Conditions for Iterated Admissibility*

Adam Brandenburger† H. Jerome Keisler‡

First Version 06/16/00
Current Version 05/08/01
Extended Abstract

## 1 Introduction

Iterated admissibility (weak dominance) is a long-standing and attractive solution concept, making strong predictions in many games, e.g., the forward-induction path in signalling games and the backward-induction path in perfect-information trees. Yet its logical basis has remained unclear. The difficulty appears to be this. The 'philosophy' behind iterated admissibility is that a player, Ann say, should consider everything possible; in particular, she should assign strictly positive probability to each of Bob's strategies. Now turn to Bob. If he assumes that Ann adheres to the foregoing 'philosophy,' then he should rule out the possibility that Ann will choose an inadmissible strategy, i.e., he should assign zero probability to these of Ann's strategies. But if Bob, like Ann, adheres to the everything-is-possible philosophy, he should give all of Ann's strategies positive probability. We seem to have reached some kind of contradiction.

This paper provides a logical basis for iterated admissibility. It provides epistemic conditions–i.e., conditions on the rationality of the players in the game, on what the players assume about one another's rationality, etc.–under which the players will indeed choose iteratively admissible strategies. There are two central features of the analysis:

*This article was published in *Theoretical Aspects of Rationality and Knowledge: Proceedings of the Eighth Conference*, edited by Johan van Benthem, Morgan Kaufmann, 2001, 31-37, and is posted with permission from Elsevier. We are indebted to Bob Aumann and Gus Stuart for many conversations relevant to this paper. Geir Asheim, Pierpaolo Battigalli, Amanda Friedenberg, Joe Halpern, Ben Polak, Michael Schwarz, Marciano Siniscalchi, and participants in the Fourth Conference on Logic and the Foundations of Game and Decision Theory (Torino, June 2000), the Eleventh International Conference on Game Theory (State University of New York at Stony Brook, July 2000), the First World Congress of the Game Theory Society (Bilbao, July 2000), and the Microeconomic Theory Workshop at Yale University provided important input. Financial support from the Harvard Business School, the National Science Foundation, and the Vilas Trust Fund is gratefully acknowledged. ecia-11-17-05

†Address: Harvard Business School, Boston, MA 02163, abrandenburger@hbs.edu

‡Address: Department of Mathematics, University of Wisconsin-Madison, Madison, WI 53706, keisler@math.wisc.edu

i. It is assumed that the players possess *lexicographic probability systems* as in Blume, Brandenburger, and Dekel (1991). Specifically, each (type of each) player possesses a sequence of probability measures satisfying certain conditions, and prefers one strategy to another if the sequence of expected utilities associated with the first is lexicographically greater than the sequence associated with the second.

ii. It is assumed that the players operate in a complete *sequential probability structure*. Fix a two-person finite strategic-form game, and let $S^a$ and $S^b$ be the strategy sets of the two players. To this description, a *sequential probability structure* appends type sets $T^a$ and $T^b$. Each element of $T^a$, i.e., each type of player $a$, is associated with a sequence of probability measures on (an appropriate $\sigma$-field of subsets of) $S^b \times T^b$. Likewise, each element of $T^b$, i.e., each type of player $b$, is associated with a sequence of probability measures on (an appropriate $\sigma$-field of subsets of) $S^a \times T^a$. Such a structure is *complete*[1] if for each sequence of probability measures on $S^b \times T^b$ there is a type of player $a$ with that sequence, and for each sequence of probability measures on $S^a \times T^a$ there is a type of player $b$ with that sequence. Roughly speaking then, a structure is complete if all possible types are present.

To state the main result of the paper, some definitions are needed: Say there is a *first-order assumption of rationality* if each player is rational. Say there is a *second-order assumption of rationality* if each player is rational and assumes the other is rational. And so on. Next, call a strategy of a player 1-*admissible* if it is admissible. Call a strategy 2-*admissible* if it is admissible in the restricted game obtained by deleting all strategies that aren't 1-admissible from the original game. And so on.

We can now state the main theorem: Fix a complete sequential probability structure. Fix also a state of the world at which there is an $m$th-order assumption of rationality, for some integer $m = 0, 1, \ldots$. Then the strategies at that state are $m$-admissible. Conversely, fix an $m$-admissible strategy for each player, for some integer $m$. Then there is a state of the world at which there is an $m$th-order assumption of rationality, and these strategies are chosen.

At the conceptual level, the theorem reveals the true 'philosophy' behind iterated admissibility. It is not enough, as we said earlier, that Ann should consider each of Bob's strategies possible. Rather, she considers possible both every strategy that Bob might play and every type that Bob might be. (Likewise, Bob considers possible both every strategy that Ann might play and every type that Ann might be.) The paper shows that it is this notion, when properly formalized, that yields the conclusion that the players will choose iteratively admissible strategies.

The present paper is related to Asheim (1999), Battigalli and Siniscalchi (1999), Board (1998), and Ewerhart (2000) inter alia; also relevant is Brandenburger and Keisler (1999).

[1] The terminology is from Brandenburger and Keisler (1999).

## 2 Lexicographic Probability Systems

By a Polish space we will mean a complete separable metrizable space. Given a Polish space $\Omega$, let $\mathcal{M}(\Omega)$ be the space of all Borel probability measures on $\Omega$ with the topology of weak convergence, so that $\mathcal{M}(\Omega)$ is again a Polish space. By the *empty measure* on $\Omega$ we will mean the function that assigns 0 to every Borel subset of $\Omega$. Let $\mathcal{M}'(\Omega)$ be $\mathcal{M}(\Omega)$ with the empty measure added as a new point at distance 1 from every point of $\Omega$. Then $\mathcal{M}'(\Omega)$ is again a Polish space, with the empty measure as an isolated point, and so $(\mathcal{M}'(\Omega))^N$ is Polish with the product topology. Now define $\mathcal{M}^\infty(\Omega)$ to be the subspace of $(\mathcal{M}'(\Omega))^N$ consisting of all infinite sequences from $\mathcal{M}(\Omega)$ and all finite sequences from $\mathcal{M}(\Omega)$ followed by infinitely many empty measures. That is: $\sigma = (\mu_0, \ldots) \in \mathcal{M}^\infty(\Omega)$ if and only if for each $n$, if $\mu_n$ is the empty measure then $\mu_{n+1}$ is the empty measure. Then $\mathcal{M}^\infty(\Omega)$ with the induced metric is a closed subset of $(\mathcal{M}'(\Omega))^N$, and hence is a Polish space.

The length $l(\sigma)$ of a sequence $\sigma = (\mu_0, \ldots) \in \mathcal{M}^\infty(\Omega)$ is defined as the least $n$ such that $\mu_n$ is the empty measure, and is $\infty$ if there is no such $n$. (We allow the sequence $\sigma$ of length 0 where $\mu_n = 0$ for all $n$.)

**Definition 2.1** *The set $L(\Omega)$ of* ***lexicographic probability systems*** *is the set of all $\sigma = (\mu_0, \ldots) \in M^\infty(\Omega)$ such that:*

*(i) $0 < l(\sigma) < \infty$;*

*(ii) there are Borel sets $U_i$, for $i < l(\sigma)$, such that $\mu_i(U_i) = 1$ and $\mu_i(U_j) = 0$ for $i \neq j$;*

*(iii) $\sigma$ has* ***full support****, that is, $\Omega = \cup_{i<l(\sigma)} \operatorname{Supp} \mu_i$.*

(In the case of finite $\Omega$, Definition 2.1 reduces to Definition 5.2 in Blume, Brandenburger, and Dekel (1991), where an axiomatic derivation is provided in terms of the player's preferences over acts.) In the interpretation of a given $\sigma = (\mu_0, ..., \mu_{n-1}) \in \mathcal{L}(\Omega)$, the set $U_{n-1}$ is to be thought of as player's primary 'hypothesis' about the true state. But the player recognizes that his primary hypothesis might be mistaken, and so he also form a secondary hypothesis, as represented by set $U_{n-2}$. And so on, until the $n$th hypothesis $U_0$.[2] Condition (iii) says that the player 'considers everything possible.'

**Lemma 2.1** *In Definition 2.1, if condition (ii) holds then the sets $U_i$ may be chosen so that they form a partition of $\Omega$.*

**Lemma 2.2** *In Definition 2.1, condition (iii) is equivalent to:*

*(iii′) for each nonempty open set $U$ there is an $i$ such that $\mu_i(U) > 0$.*

**Definition 2.2** *A set $E$ is* ***assumed under*** *$\sigma$ (at* ***level*** *$j$) if $\sigma \in \mathcal{L}(\Omega)$ and there is a $j < l(\sigma)$ such that:*

*(i) $\mu_i(E) = 1$ for all $j \leq i < l(\sigma)$,*

*(ii) $\mu_i(E) = 0$ for all $i < j$,*

*(iii) $E \subseteq \cup_{i \geq j} \operatorname{Supp} \mu_i$.*

[2] Note, then, that the sequence $(\mu_0, \ldots, \mu_{n-1})$ begins with the least significant measure.

One can provide a characterization of conditions (i)-(iii) in this definition in terms of the player's preferences over acts. In these terms, to say that the player assumes $E$ is to say that he acts as he would if he were to know $E$. (Technically, conditions (i) and (ii) say that the consequences in the states in $E$ are determining for strict preference. Condition (iii) is kind of regularity condition, and says that 'all of $E$ should matter.')

# 3 Epistemics

This section develops a formalism, involving sequences of measures, with which to talk about the rationality of the players in a game, what the players assume about one another's rationality, etc.

**Definition 3.1** *Fix nonempty finite sets $S^a$ and $S^b$. An $(S^a, S^b)$-**based (interactive) sequential probability structure** is a structure*

$$(S^a, S^b, T^a, T^b, \lambda^a, \lambda^b)$$

*where $T^a$ and $T^b$ are nonempty Polish spaces, $\lambda^a$ is a continuous mapping from $T^a$ to $\mathcal{M}^{\infty}(S^b \times T^b)$, and similarly for $\lambda^b$. The structure is **complete** if $\lambda^a$ and $\lambda^b$ are onto.*

**Proposition 3.1** *For any nonempty Polish spaces $S^a$, $S^b$ there is a complete $(S^a, S^b)$-based sequential probability structure.*

Now fix a two-person finite strategic-form game $\langle S^a, S^b, \pi^a, \pi^b \rangle$ where $S^a$, $S^b$ denote the strategy sets and $\pi^a$, $\pi^b$ denote the payoff functions of the players $a$, $b$. As usual, all definitions and results have counterparts with $a$ and $b$ reversed.[3]

**Definition 3.2** *A pair $(s^a, t^a) \in S^a \times T^a$ is **rational** (with respect to the payoff function $\pi^a$) if $\lambda^a(t^a) \in \mathcal{L}(S^b \times T^b)$ and for every $r^a \in S^a$,*

$$\left(\int_{S^b \times T^b} \pi^a(s^a, s^b) d\mu_i(s^b, t^b)\right)_{i=n-1}^{0} \geq^L \left(\int_{S^b \times T^b} \pi^a(r^a, s^b) d\mu_i(s^b, t^b)\right)_{i=n-1}^{0},$$

*where we write $\lambda^a(t^a) = (\mu_0, \ldots, \mu_{n-1})$.*

**Remark 3.1** *Note that $(s^a, t^a)$ is rational if and only if for every $r^a \in S^a$,*

$$\left(\sum_{s^b \in S^b} \pi^a(s^a, s^b) \operatorname{marg}_{S^b} \mu_i(s^b)\right)_{i=n-1}^{0} \geq^L \left(\sum_{s^b \in S^b} \pi^a(r^a, s^b) \operatorname{marg}_{S^b} \mu_i(s^b)\right)_{i=n-1}^{0},$$

*where $\operatorname{marg}_{S^b} \mu_i$ denotes the marginal on $S^b$ of $\mu_i$.*

[3] We restrict attention to two players, for the sake of notational simplicity. Our analysis readily extends to three or more players.

For $E \subseteq S^b \times T^b$, we let

$$B^a(E) = \{t^a \in T^a : E \text{ is assumed under } \lambda^a(t^a)\}.$$

Define $R_1^a$ to be the set of all rational pairs $(s^a, t^a)$, and define $R_m^a$ inductively by

$$R_{m+1}^a = R_m^a \cap (S^a \times B^a(R_m^b)).$$

**Definition 3.3** *If $(s^a, t^a, s^b, t^b) \in R_m^a \times R_m^b$, say there is an $m$**th-order assumption of rationality** at this state.*[4]

# 4 Main Result

To state our main result, some standard definitions on admissibility are needed.

**Definition 4.1** *A strategy $s^a \in S^a$ is **admissible** if there is no $\mu \in \mathcal{M}(S^a)$ such that*

$$\begin{array}{ll} \sum_{r^a \in S^a} \pi^a(r^a, s^b)\mu(r^a) \geq \pi^a(s^a, s^b) & \text{for every } s^b \in S^b, \\ \sum_{r^a \in S^a} \pi^a(r^a, s^b)\mu(r^a) > \pi^a(s^a, s^b) & \text{for some } s^b \in S^b. \end{array}$$

Given $X \subseteq S^a$ and $Y \subseteq S^b$, let $\pi^a|(X \times Y)$ denote the restriction of $\pi^a$ to $X \times Y$ and $\pi^b|(Y \times X)$ the restriction of $\pi^b$ to $Y \times X$. Now, let $S_0^a = S^a$ and define $S_m^a$ inductively by

$$S_{m+1}^a = \{s^a \in S_m^a : s^a \text{ is admissible in the game } \langle S_m^a, S_m^b, \pi^a|(S_m^a \times S_m^b), \pi^b|(S_m^b \times S_m^a)\rangle\}.$$

**Definition 4.2** *A strategy $s^a \in S_m^a$ is called $m$-**admissible**.*

**Remark 4.1** *Note that since $S^a$ and $S^b$ are finite, there is an $M$ such that $S_m^a = S_M^a$ and $S_m^b = S_M^b$ for all $m \geq M$.*

The main result of the paper is:

**Theorem 4.1** *Fix a complete $(S^a, S^b)$-based sequential probability structure*

$$(S^a, S^b, T^a, T^b, \lambda^a, \lambda^b)$$

*and payoff functions $\pi^a$ and $\pi^b$.*

*(i) If there is an mth-order assumption of rationality at a state $(s^a, t^a, s^b, t^b)$, then the strategies $s^a$ and $s^b$ are $m$-admissible.*

*(ii) If $s^a$ and $s^b$ are $m$-admissible strategies, then there is a state $(s^a, t^a, s^b, t^b)$ at which there is an mth-order assumption of rationality.*

[4] Thus, there is a first-order assumption of rationality if each player is rational at the state. There is a second-order assumption of rationality if each player is rational and assumes that the other is rational. And so on.

# References


[1] Asheim, G., "Common Knowledge of Proper Consistency," unpublished, Department of Economics, University of Oslo, 1999.

[2] Battigalli, P., and M. Siniscalchi, "Interactive Beliefs and Forward Induction," Working Paper ECO 99/15, European University Institute, 1999.

[3] Blume, L., A. Brandenburger, and E. Dekel, "Lexicographic Probabilities and Choice under Uncertainty," *Econometrica*, 59, 1991, 61-79.

[4] Board, O., "Algorithmic Characterization of Rationalizability in Extensive Form Games," unpublished, Brasenose College, Oxford, 1998.

[5] Brandenburger, A., and H.J. Keisler, "An Impossibility Theorem on Beliefs in Games," unpublished, 1999.

[6] Ewerhart, C., "A Decision-Theoretic Characterization of Iterated Weak Dominance," unpublished, University of Mannheim, 2000.