A Conceptually Well-Founded Characterization of Iterated Admissibility Using an "All I Know" Operator

Joseph Y. Halpern and Rafael Pass
Department of Computer Science
Cornell University
halpern@cs.cornell.edu rafael@cs.cornell.edu

Brandenburger, Friedenberg, and Keisler provide an epistemic characterization of *iterated admissibility* (IA), also known as *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. In earlier work, we gave a characterization of iterated admissibility using an "all I know" operator, that captures the intuition that "all the agent knows" is that agents satisfy the appropriate rationality assumptions. That characterization did not need complete structures and used probability structures, not LPSs. However, that characterization did not deal with Samuelson's conceptual concern regarding IA, namely, that at higher levels, players do not consider possible strategies that were used to justify their choice of strategy at lower levels. In this paper, we give a characterization of IA using the all I know operator that does deal with Samuelson's concern. However, it uses LPSs. We then show how to modify the characterization using notions of "approximate belief" and "approximately all I know" so as to deal with Samuelson's concern while still working with probability structures.

1 Introduction

A strategy σ_i for player i is *admissible* with respect to a set $\Sigma = \Sigma_1 \times \cdots \times \Sigma_n$ of strategy profiles if it is a best response to some belief of player i that puts positive probability on all the strategy profiles in Σ. That is, there is some probability μ on Σ_{-i} such that no strategy in Σ_i gives player i a higher expected utility than σ_i with respect to the beliefs μ. As Pearce [18] has shown, a strategy σ_i for player i is admissible with respect to Σ iff it is not weakly dominated; that is, there is no strategy σ_i' for player i that gives i at least as high a payoff as σ_i no matter what strategy in Σ_{-i} the other players are using, and sometimes gives i a higher payoff. It seems natural for a rational player not to play an inadmissible strategy. If we delete all strategy profiles from Σ that involve an inadmissible strategy, we get a new set Σ' of strategy profiles. We can then consider which strategies are inadmissible with respect to Σ', and iterate this process. This leads to the solution concept of *iterated admissibility* (IA) (also known as *iterated deletion of weakly dominated strategies*), one of the most studied solution concepts in normal-form games.

As Samuelson [20] pointed out, there is a conceptual problem when it comes to dealing with IA. As he says, "the process appears to initially call for agents to assume that opponents may play any of their strategies but to subsequently assume that opponents will certainly not play some strategies." Brandenburger, Friedenberg, and Keisler [7] (BFK from now on) resolve this paradox by assuming that strategies are not really eliminated. Rather, they assume that strategies that are weakly dominated occur with infinitesimal (but nonzero) probability. Formally, they capture this by using what they call a *full-support* LPS—*lexicographically ordered probability sequence* [4, 5]. Recall that a *lexicograpic probability space* is a tuple $(\Omega, \mathscr{F}, (\mu_0, \mu_1, \ldots, \mu_k))$, where $\mathscr{F}$ is a σ-algebra over Ω and $\mu_0, \ldots, \mu_k$ are probability distributions on $(\Omega, \mathscr{F})$; $\vec{\mu} = (\mu_0, \ldots, \mu_k)$ is an LPS. Intuitively, the first measure in the sequence $\vec{\mu}$, μ_0, is the

Submitted to:

most important one, followed by μ_1, μ_2, and so on. The full-support requirement says that the union of the supports of $\mu_0,\dots,\mu_k$ is Ω. In this paper, for simplicity, we assume that all sets are measurable and keep Ω implicit when we speak of an LPS.

BFK define a notion of belief that they call *assumption*, where an event E is assumed in an LPS $\vec{\mu}=(\mu_0,\dots,\mu_k)$ if E is *infinitely more likely* than $\overline{E}$ under $\vec{\mu}$, and E is infinitely more likely than F for events E and F if, for all $\omega \in E$ and $\omega' \in F$, there is some i such that $\mu_i(\omega) > 0$ and if there exists j such that $\mu_j(\omega') > 0$, then there exists $j' < j$ such that $\mu_{j'}(\omega) > 0$.[1] They then show that strategies that survive k rounds of iterated deletion are exactly the ones played in states in a *complete* type structure where there is a kth-order assumption of rationality; that is, everyone assumes that everyone assumes ... ($k-1$ times) that everyone is rational. Complete type structures are particularly rich structures, where all types are possible. By considering LPSs with full support, BFK guarantee that strategies are not really eliminated; that is, no strategies are ever assigned probability 0. But full support in complete type structures also forces agents to ascribe positive probability to many other events; in particular, they must consider possible all beliefs that other agents could have about beliefs that other agents could have ... about strategies that an agent is using. The use of complete type structures also leads to other technical problems. For example, although *common assumption of rationality* (RCAR) (kth-order assumption of rationality for all k) is consistent, BFK show that it cannot hold in a complete and continuous type structure.

There has been a great deal of follow-on work on IA. We briefly discuss some of the results here. With regard to the latter point, Keisler and Lee [13] show that RCAR is satisfiable in complete (but not continuous) type structure. In their construction, the structure depends on the game; Lee [14] provides a general game-independent construction. Yang [25] defines a notion of *weak assumption* that, as the name suggests, is weaker than assumption, and shows that common weak assumption of rationality is satisfiable in continuous type structures. Catonini and de Vito [8] point out that the full-support condition depends crucially on the topology of the type space; they replace the full-support condition by what they call *cautiousness*, which requires only that all strategy profiles are considered possible, and provide a characterization of IA in complete type spaces using a notion of common cautious belief in rationality. Finally, Perea [19], using the same notion of cautiousness as Catonini and de Vito, provides a characterization of IA using his version of common assumption of rationality. Perea does not need to consider complete type spaces; indeed, he shows that his notion of common assumption of rationality is satisfied even in finite spaces. However, his construction only works if the space of types is finite.

In earlier work [12], we provided a characterization of IA using an "all I know" operator. Roughly speaking, instead of assuming only that agents know (or assume) that all other agents satisfy appropriate levels of rationality, we assume that "all the agents know" is that the other agents satisfy the appropriate rationality assumptions. We formalized this notion by requiring that the agent ascribes positive probability to all formulas of some language $\mathcal{L}$ that are consistent with his rationality assumptions. (This admittedly fuzzy description is made precise in Section 5.) We show that the formula $\psi_k^{\mathcal{L}}$ that, roughly speaking, says that "all that players know with respect to language $\mathcal{L}$ is that all that players know with respect to $\mathcal{L}$ ... (k times) is that all players are rational" characterizes k levels of iterated deletion of weakly dominated strategies, both in the case $\mathcal{L}$ that just describes the set of possible strategies played and in the case that $\mathcal{L}$ describes, not only players' strategies, but also players' beliefs about what strate-

[1] This definition of infinitely more likely is due to Blume, Brandenburger, and Dekel [4]. BFK give a somewhat general definition that applies even if not all sets are meaurable. BFK also require that the measures in an LPS have disjoint supports. While we do not require this, our results would continue to hold with essentially no change in proof if we also imposed this requirement. We remark that the idea of requiring strategies that survive $k+1$ rounds of interated deletion to be infinitely more likely than strategies that survive only k rounds of iterated deletion, used by BFK, goes back to Stahl [23].

gies other players are using (including higher-order beliefs about other players beliefs). That is, we show that if the formula $\psi_k^{\mathscr{L}}$ (for these two choices of $\mathscr{L}$) holds at some state in an arbitrary model, then the strategies used at that state survive k rounds of iterated deletion. Conversely, if a strategy σ survives k rounds of iterated deletion, then there is a state in some model M where $\psi_k^{\mathscr{L}}$ holds and strategy σ is played. If the language $\mathscr{L}$ just talks about strategies, then we can take M to be finite; we do not need to work with complete type structures to characterize IA. When we consider the language of strategies, "all I know" can be viewed as roughly analogous to Catonini and de Vito's [8] notion of cautiousness. On the other hand, if $\mathscr{L}$ talks about strategies and beliefs, then the structures we use are essentially complete type structures.

The problem with this characterization of IA is that the formula $\psi_k^{\mathscr{L}}$ says that players ascribe positive probability to all and only the strategies of other players that survive $k-1$ rounds of iterated deletion. Thus, while it does provide an elegant characterization of IA, it does not deal with Samuelson's concern that, at higher levels, players do not consider possible strategies that were used to justify their choice of strategy at lower levels. In this paper we give a characterization using an "all I know" operator in the spirit of our earlier characterization, but that does enforce a full-support condition, at least for strategies.

To do this, we use a generalized belief operator. Roughly speaking, $B_i(\varphi_1,\ldots,\varphi_\ell)$ says that agent i believes that φ_1 is true, but if it not, then φ_2 is true, and if neither φ_1 nor φ_2 is true, then φ_3 is true, and so on. Thus, with this generalized belief operator, i describes not only his beliefs, but his "plan of retreat" in case his beliefs turn out to be false. There is an "all I know" operator that corresponds to this generalized belief operator in a natural way. The combination of the generalized belief operator and the corresponding "all I know" operator leads to a characterization of IA with a full-support requirement on strategies.

In our earlier work, we were able to use standard beliefs and represent uncertainty using standard probability. However, to give semantics to the generalized belief operator, we need LPSs (we could equally well use other approaches that can represent infinitesimal probability, like conditional probability spaces or nonstandard probability spaces. Just as with our earlier work (and unlike BFK), we do not need to use complete structures; indeed, it suffices to work with finite structures to get our characterization. Nor do we assume *a priori* that the LPS is a full-support LPS; the formula that characterizes IA forces any LPS that satisfies it to be a full-support LPS (at least with respect to strategies).

Although our new approach requires LPSs, it does lead to an arguably more elegant epistemic characterization that more directly deals with Samuelson's concern (in much the same way that BFK's approach does). That said, LPSs require agents to make very fine probability distinctions. In Section 4, we show how we can modify the new approach using notions of "approximate belief" and "approximately all I know" so as to deal with Samuelson's concern while still allowing us to work with probability structures, rather than LPSs. Roughly speaking, our result says that a strategy for agent i survives k rounds of iterated deletion if it is played at a state where all agent i approximately knows is that all other agents are k-level rational, but if i were to find out that they are not, then all i approximately knows is that they are $(k-1)$-level rational, and so on.

2 Probability Structures, Rationalizability, and Admissibility

The material in this section is taken almost verbatim from our earlier paper [12].

We consider normal-form games with n players. Given a (normal-form) n-player game Γ, let Σ_i^Γ denote the strategies of player i in Γ, and let u_i^Γ denote the utility function of player i in Γ. We omit the superscript Γ when it is clear from context or irrelevant. Let $\vec{\Sigma} = \Sigma_1 \times \cdots \times \Sigma_n$. We restrict to finite games, so we assume that $\vec{\Sigma}$ is finite. We further assume, without loss of generality (since the game is finite), that for each player i, the range of u_i is $[0,1]$. Let $\mathscr{L}^1$ 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, according to player i, φ holds with probability 1, and $\langle B_i \rangle\varphi$ as saying, accoding to i, 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 $(\Omega, \mathbf{s}, \mathscr{F}, \mathscr{PR}_1, \ldots, \mathscr{PR}_n)$, 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; $\mathscr{F}$ is a σ-algebra over Ω; and, for each player i, $\mathscr{PR}_i$ associates with each state ω a probability distribution $\mathscr{PR}_i(\omega)$ on $(\Omega, \mathscr{F})$. Intuitively, $\mathbf{s}(\omega)$ is the strategy profile used at state ω and $\mathscr{PR}_i(\omega)$ is player i's probability distribution at state ω. As is standard, we require that each player knows his strategy and his beliefs. Formally, we require that

1. for each strategy σ_i for player i, $[\![\sigma_i]\!]_M = \{\omega : \mathbf{s}_i(\omega) = \sigma_i\} \in \mathscr{F}$, where $\mathbf{s}_i(\omega)$ denotes player i's strategy in the strategy profile $\mathbf{s}(\omega)$;
2. $\mathscr{PR}_i(\omega)([\![\mathbf{s}_i(\omega)]\!]_M) = 1$;
3. for each probability measure π on $(\Omega, \mathscr{F})$ and player i, $[\![\pi, i]\!]_M = \{\omega : \mathscr{PR}_i(\omega) = \pi\} \in \mathscr{F}$; and
4. $\mathscr{PR}_i(\omega)([\![\mathscr{PR}_i(\omega), i]\!]_M) = 1$.

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 $\mathscr{PR}_i(\omega)$. That is, i's expected utility with $\mathbf{s}_i(\omega)$ is at least as high as with any other strategy in Σ_i, given i's beliefs. (Because we restrict to appropriate structures, a player's 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 $F \in \mathscr{F}$ such that $F \subseteq [\![\varphi]\!]_M$ and $\mathscr{PR}_i(\omega)(F) = 1$, where $[\![\varphi]\!]_M = \{\omega : (M,\omega) \models \varphi\}$.
- $(M,\omega) \models \langle B_i \rangle\varphi$ if there exists a set $F \in \mathscr{F}$ such that $F \subseteq [\![\varphi]\!]_M$ and $\mathscr{PR}_i(\omega)(F) > 0$.

We say that φ is *valid (for game* Γ*)* if $(M,\omega) \models \varphi$ for all structures M appropriate for game Γ and all states ω in M. We say that φ is *satisfiable (for game* Γ*)* if $(M,\omega) \models \varphi$ for some state ω in some structure M appropriate for Γ.

Note that here we do not assume that $[\![\varphi]\!]_M$ is measurable. Thus, we cannot take $B_i\varphi$ to mean that agent i ascribes probability 1 to $[\![\varphi]\!]_M$. Rather, we take it to mean that there is a set of probability 1 contained in $[\![\varphi]\!]_M$. Put another way, we are requiring that the *inner measure* of $[\![\varphi]\!]_M$ is 1. Similarly, $\langle B_i \rangle\varphi$ does not quite say that i ascribes $[\![\varphi]\!]_M$ positive probability; rather, it says that the inner measure of $[\![\varphi]\!]_M$ is positive. Given a language (set of formulas) $\mathscr{L}$, M is *$\mathscr{L}$-measurable* if M is appropriate (for some game Γ) and $[\![\varphi]\!]_M \in \mathscr{F}$ for all formulas $\varphi \in \mathscr{L}$. It is easy to check that in an $\mathscr{L}^1$-measurable structure, $B_i\varphi$ means that i ascribes probability 1 to $[\![\varphi]\!]_M$, $\langle B_i \rangle\varphi$ means that i ascribes positive probability to $[\![\varphi]\!]_M$, and $\langle B_i \rangle\varphi$ is equivalent to $\neg B_i \neg\varphi$.

Definition 2.1: Strategy σ for player is i *weakly dominated by mixed strategy* 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 weakly dominated strategies if, for each player j, there exists a sequence $NWD_j^0, NWD_j^1, NWD_j^2, \ldots, NWD_j^k$ of sets of strategies for player j such that $NWD_j^0 = \Sigma_j$ and, if $h < k$, then NWD_j^{h+1} consists of the strategies in NWD_j^h not weakly dominated by any mixed strategy with respect to NWD_{-j}^h, and $\sigma \in NWD_i^k$. Strategy σ for player j survives iterated deletion of weakly dominated strategies if it survives k rounds of iterated deletion of weakly dominated strategies for all k, that is, if $\sigma \in NWD_j^\infty = \cap_k NWD_j^k$. ∎

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

Proposition 2.2: [18] *A strategy* σ *for player* i *is not weakly dominated by any mixed 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*.*

3 The Earlier Characterization of IA

In this section, we review our earlier characterization of iterated admissibility, to set the stage for the new results. Again, the exposition is taken almost verbatim from our earlier paper.

For each player i, define the formulas RAT_i^k inductively by taking RAT_i^0 to be *true* and RAT_i^{k+1} to be an abbreviation of

$$RAT_i \wedge B_i(RAT_{-i}^k),$$

where RAT_{-i}^k is an abbreviation of $\wedge_{j \neq i} RAT_j^k$.[2] That is, RAT_i^{k+1} holds (i.e., player i is $(k+1)$-level rational) iff player i is playing a best response to his beliefs, and he knows that all players are k-level rational. Thus, with these definitions, player i is taken to be $(k+1)$-level rational iff player i is rational (i.e., playing a best response to his beliefs), and knows that all other player are k-level rational.[3] But what else do players know?

We want to consider a situation where, intuitively, *all* an agent knows about the other agents is that they satisfy the appropriate rationality assumptions. More precisely, we modify the formula RAT_i^{k+1} to require that not only does player i know that the players are k-level rational, but this is the *only* thing that he knows about the other players. That is, we say that agent i is $(k+1)$-level rational if player i is rational, he knows that the players are k-level rational, and this is all player i knows about the other players. We here use the phrase "all agent i knows" in essentially the same sense that it is used by Levesque [15] and Halpern and Lakemeyer [11], but formalize it a bit differently. 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 that is consistent with φ. Thus, what "all I know" means is very sensitive to the choice of the language. To stress this point, we talk about "all I knows *with respect to language* $\mathscr{L}$".

To define the "all I know" operator, we use a modal operator $\Diamond$ that characterizes consistency, which is defined as follows:

- $(M, \omega) \models \Diamond\varphi$ iff there is some structure M' appropriate for Γ and state ω' such that $(M', \omega') \models \varphi$.

[2]We use similar abbreviations in the sequel without comment.

[3]We should perhaps say "believes" here rather than "knows", since a player can be mistaken. We are deliberately blurring the subtle distinctions between "knowledge" and "belief" here.

Intuitively, $\Diamond\varphi$ is true if there is some state and structure where φ is true; that is, if φ is satisfiable. Note that if $\Diamond\varphi$ is true at some state, then it is true at all states in all structures. Define $O_i^{\mathscr{L}}\varphi$ (read "all agent i knows with respect to the language $\mathscr{L}$") to be an abbreviation of

$$B_i\varphi \wedge (\wedge_{\psi\in\mathscr{L}}(\Diamond(\varphi\wedge\psi) \Rightarrow \langle B_i\rangle\psi)).$$

In this paper, we focus on just one of the languages considered in our earlier paper, whose formulas can talk about strategies (but not beliefs) of the players. Define the primitive proposition $play_i(\sigma)$ as follows:

- $(M,\omega) \models play_i(\sigma)$ iff $\omega \in [\![\sigma]\!]_M$.

Let $play(\vec{\sigma})$ be an abbreviation of $\wedge_{j=1}^n play_j(\sigma_j)$, and let $play_{-i}(\sigma_{-i})$ be an abbreviation of $\wedge_{j\neq i} play_j(\sigma_j)$. 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}. Let $\mathscr{L}^0(\Gamma)$ 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 $\mathscr{L}_i^0(\Gamma)$ consist of just the formulas of the form $play_i(\sigma)$, and let $\mathscr{L}_{-i}^0(\Gamma) = \cup_{j\neq i}\mathscr{L}_j^0(\Gamma)$. Again, we omit the parenthetical Γ when it is clear from context or irrelevant.

The sense in which a player is rational when playing a strategy that survives iterated deletion is captured by the formulas $\mathscr{L}^0\text{-}RAT_i^k$, which are defined inductively by taking $\mathscr{L}^0\text{-}RAT_i^0$ to be *true* and $\mathscr{L}^0\text{-}RAT_i^{k+1}$ to be an abbreviation of

$$RAT_i \wedge B_i(PLAYCON_i^k) \wedge O_i^{\mathscr{L}_{-i}^0}(\mathscr{L}^0\text{-}RAT_{-i}^k),$$

where $PLAYCON_i^k$ (read "player i plays a strategy consistent with k-level rationality") is an abbreviation of $\wedge_{\sigma'\in\Sigma_i(\Gamma)}(play_i(\sigma') \Rightarrow \Diamond(play_i(\sigma') \wedge \mathscr{L}^0\text{-}RAT_i^k))$. That is, $\mathscr{L}^0\text{-}RAT_i^{k+1}$ holds (i.e., player i is $(k+1)$-level rational) iff player i is rational, believes that he is playing a strategy that is consistent with k-level rationality, knows that other players are k-level rational, and that is all player i knows about the *strategies* of the other players.

By expanding the modal operator O, it easily follows that $\mathscr{L}^0\text{-}RAT_i^{k+1}$ implies $RAT_i \wedge B_i(\mathscr{L}^0\text{-}RAT_{-i}^k)$. an easy induction on k then shows that $\mathscr{L}^0\text{-}RAT_i^{k+1}$ implies RAT_i^{k+1}. But $\mathscr{L}^0\text{-}RAT_i^{k+1}$ requires more; it requires player i to assign positive probability to each strategy profile for the other players that is compatible with $\mathscr{L}^0\text{-}RAT_{-i}^k$ (i.e., with level-k rationality). As shown in our earlier paper [12], the formula $\mathscr{L}^0\text{-}RAT_i^k$ characterizes strategies that survive iterated deletion of weakly dominated strategies.

Theorem 3.1: *The following are equivalent:*

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

(b) there exists an $\mathscr{L}^0$-measurable structure M^k appropriate for Γ and a state ω^k in M^k such that $\mathbf{s}_i(\omega^k) = \sigma$ and $(M^k,\omega^k) \models \mathscr{L}^0\text{-}RAT_i^k$;

(c) there exists a structure M^k appropriate for Γ and a state ω^k in M^k such that $\mathbf{s}_i(\omega^k) = \sigma$ and $(M^k,\omega^k) \models \mathscr{L}^0\text{-}RAT_i^k$.

In addition, if $\vec{\sigma} \in NWD^k$, then there is a finite structure $\bar{M}^k = (\Omega^k, \mathbf{s}^k, \mathscr{F}^k, \mathscr{PR}_1^k, \ldots, \mathscr{PR}_n^k)$ such that $\Omega^k = \{(k',i,\vec{\sigma}) : 0 \le k' \le k, 1 \le i \le n, \vec{\sigma} \in NWD^{k'}\}$, $\mathbf{s}^k(k',i,\vec{\sigma}) = \vec{\sigma}$, $\mathscr{F}^k = 2^{\Omega^k}$, and for all states $(k',i,\vec{\sigma}) \in \Omega^k$, $(\bar{M}^k,(k',i,\vec{\sigma})) \models \mathscr{L}^0\text{-}RAT_{-i}^{k'}$.

4 The new characterization of IA

In our earlier characterization of IA, in a state where $\mathscr{L}^0$-RAT^k holds, player i does *not* consider all strategies possible, but only the ones consistent with the appropriate level of rationality. That is, because of the $B_i(PLAYCON_i^{k-1})$ conjunct in $\mathscr{L}^0$-RAT_i^k, player i ascribes positive probability only to strategies consistent with $(k-1)$-level rationality. This means that the characterization of the earlier paper does not address Samuelson's concern. More specifically, it does not provide an epistemic explanation for *why*, at higher levels, players do not consider possible strategies that were used to justify their choice of strategy at lower levels; it just assumes that they do. We deal with this problem in our new characterization of IA. The new characterization forces the agent to ascribe positive probability to all strategies, and thus can be viewed as forcing a full-support requirement, at the level of strategies.

As a first step to getting this characterization, we introduce a notion of *generalized belief*, which may be of independent interest. Specifically, we consider formulas of the form $B_i(\varphi_1,\ldots,\varphi_\ell)$. As we said in the introduction, this formula can be read "agent i believes that φ_1 is true, but if it not, then φ_2 is true, and if neither φ_1 nor φ_2 is true, then φ_3 is true, ..., and if none of $\varphi_1,\ldots,\varphi_{\ell-1}$ is true, then φ_ℓ is true. We give semantics to such formulas in an LPS $\vec{\mu}=(\mu_0,\ldots,\mu_k)$.[4]

To give semantics to generalized belief, we use *LPS structures*, that is, structures of the form $M=(\Omega,\mathbf{s},\mathscr{F},\mathscr{PR}_1,\ldots,\mathscr{PR}_n)$, where now $\mathscr{PR}_i$ associates with each state an LPS. To define the semantics of the generalized belief operator, we need to recall the definition of conditioning in LPSs [4]. For simplicity, we restrict our attention to structures M where Ω is finite and $\mathscr{F}$ consists of all the subsets of Ω; that is, every set is measurable; we refer to such structures as *fully measurable*. Given a measurable set U and $\vec{\mu}=(\mu_0,\ldots,\mu_k)$, define

$$\vec{\mu}|U=(\mu_{k_0}(\cdot\mid U),\mu_{k_1}(\cdot\mid U),\ldots),$$

where $(k_0,k_1,\ldots)$ is the subsequence of all indices for which the probability of U is positive. Formally, $k_0=\min\{k:\mu_k(U)>0\}$ and, if μ_{k_h} has been defined and there exists an index h' such that $k_h<h'\leq k$ and $\mu_{h'}(U)>0$, then $k_{h+1}=\min\{h':\mu_{h'}(U)>0,\ k_h<h'\leq k\}$. Note that $\vec{\mu}|U$ is undefined if $\vec{\mu}(U)=\vec{0}$ (i.e., $\mu_j(U)=0$ for $j=0,\ldots,k$) and that the length of the sequence $\vec{\mu}|U$ depends on U. If $(\vec{\mu}|U)=(\mu_{k_0},\ldots)$, then we write $\vec{\mu}(V\mid U)_0$ to denote $\mu_{k_0}(V\mid U)$, the conditional probability according to the first probability measure in the LPS $\vec{\mu}|U$.

If $M=(\Omega,\mathbf{s},\mathscr{F},\mathscr{PR}_1,\ldots,\mathscr{PR}_n)$ and $\mathscr{PR}_i(\omega)=\vec{\mu}=(\mu_0,\ldots,\mu_k)$, then

$$\begin{array}{c}(M,\omega)\models B_i(\varphi_1,\ldots,\varphi_\ell) \text{ if } \vec{\mu}(\neg\varphi_1\wedge\ldots\wedge\neg\varphi_{\ell-1})\neq\vec{0},\ \mu_0([\![\varphi_1]\!]_M)=1,\ (\vec{\mu}([\![\varphi_2]\!]_M\mid[\![\neg\varphi_1]\!]_M)_0=1,\\ \ldots,\ (\vec{\mu}([\![\varphi_\ell]\!]_M\mid[\![\neg\varphi_1\wedge\ldots\wedge\neg\varphi_{\ell-1}]\!]_M)_0=1.\end{array}$$

That is, φ_1 gets probability 1 at the top level, φ_2 get probability 1 at the top level conditional on φ_1 being false, and so on. (The first requirement, that $\vec{\mu}(\neg\varphi_1\wedge\ldots\wedge\neg\varphi_{\ell-1})\neq\vec{0}$, ensures that all the conditional probabilities are well defined.)

There is also a corresponding "all I know" operator, $O_i^{\mathscr{L}}(\varphi_1,\ldots,\varphi_\ell)$, which again is taken with

[4]There is nothing special about the use of LPSs here. Battigalli and Sinischalchi use *conditional probability systems* to define their notion of strong belief, and we could equally well use conditional probability systems here. We could also easily use *nonstandard probability measures*. Readers familiar with these representations of uncertainty (see [10] for an overview) should have no difficulty giving analogues of our semantic definitions using these alternative approaches.

respect to a language $\mathscr{L}$, defined as follows:

$$\begin{array}{ll}(M,\omega)\models O_i^{\mathscr{L}}(\varphi_1,\ldots,\varphi_\ell) \text{ if } & (M,\omega)\models B_i(\varphi_1,\ldots,\varphi_\ell) \text{ and,}\\ & \text{for all } \psi\in\mathscr{L}, \text{ if } (M,\omega)\models\Diamond(\varphi_1\wedge\psi) \text{ then } \mu_0([\![\psi]\!]_M)\neq\vec{0} \text{ and,}\\ & \text{for all } h \text{ with } 2\leq h\leq\ell, \text{ if } (M,\omega)\models\Diamond(\neg\varphi_1\wedge\ldots\wedge\neg\varphi_{h-1}\wedge\varphi_h\wedge\psi)\\ & \quad\text{then } (\vec{\mu}([\![\psi]\!]_M \mid [\![\neg\varphi_1\wedge\ldots\wedge\neg\varphi_{h-1}]\!]_M)_0\neq\vec{0}.\end{array}$$

It is easy to see that the new definition $O_i^{\mathscr{L}}(\varphi_1)$ is identical to the earlier definition. The generalized version $O_i^{\mathscr{L}}(\varphi_1,\ldots,\varphi_k)$ requires all formulas ψ consistent with φ to have positive probability at the top level and, in addition, for $h\geq 2$, all formulas ψ consistent with $\neg\varphi_1\wedge\ldots\wedge\neg\varphi_{h-1}\wedge\varphi_h$ must have positive probability at the top level conditional on $\neg\varphi_1\wedge\ldots\wedge\neg\varphi_{h-1}$.

Before going on, we briefly review how best response is defined in LPS structures. Since player i's beliefs at a state ω are defined by an LPS $(\mu_0,\ldots,\mu_k)$, we take the expected utility associated with i's strategy $\mathbf{s}_i(\omega)$ at ω to be a tuple $(u_0,\ldots,u_k)$, where u_j is the expected utility of $\mathbf{s}_i(\omega)$ with respect to probability μ_j. We can then compare two expected utilities lexicographically: $(u_0,\ldots,u_k)>(u'_0,\ldots,u'_k)$ if there exists a $j\leq k$ such that $u_0=u'_0$, …, $u_{j-1}=u'_{j-1}$, and $u_j>u'_j$. With this definition, we can still take RAT_i to hold at ω if $\mathbf{s}_i(\omega)$ is a best response, given i's about the strategies of other players at ω.

We can now define the formulas $\mathscr{L}^0$-$GRAT_i^k$ (the G stands for "generalized") inductively by taking $\mathscr{L}^0$-$GRAT_i^0$ to be *true* and $\mathscr{L}^0$-$GRAT_i^{k+1}$ to be an abbreviation of

$$RAT_i\wedge O_i^{\mathscr{L}^0_{-i}}(\mathscr{L}^0\text{-}GRAT^k_{-i},\ldots,\mathscr{L}^0\text{-}GRAT^0_{-i}).$$

That is, all agent i knows is that the other agents are k-level rational, but if they are not, then are $(k-1)$-level rational, and if they are not, they are $(k-2)$-level rational, and so on.

Theorem 4.1: *The following are equivalent:*

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

(b) there exists a fully measurable $\mathbf{s}_i(\omega^k)=\sigma$ and $(M^k,\omega^k)\models\mathscr{L}^0$-$GRAT_i^k$.[5]

In addition, if $\vec{\sigma}\in NWD^k$, then there is a fully measurable LPS structure $\bar{M}^k=(\Omega^k,\mathbf{s},\mathscr{F},\mathscr{PR}_1^k,\ldots,\mathscr{PR}_n^k)$ such that $\Omega^k=\{(k',i,\vec{\sigma}):0\leq k'\leq k, 1\leq i\leq n,\vec{\sigma}\in NWD^{k'}\}$, $\mathbf{s}^k(k',i,\vec{\sigma})=\vec{\sigma}$, $\mathscr{F}^k=2^{\Omega^k}$, and for all states $(k',i,\vec{\sigma})\in\Omega^k$, $(\bar{M}^k,(k',i,\vec{\sigma}))\models\mathscr{L}^0$-$GRAT_{-i}^{k'}$.

We defer the proof of Theorem 3.2 to the appendix, but we mention here one of the key propositions used in proving the theorem, since it also gives some intuition for the $\mathscr{L}^0$-$GRAT_j$ operator and will allow us to compare our results to those of others. Suppose that M is a model appropriate for a game Γ, $(M,\omega)\models\mathscr{L}^0$-$GRAT_i^{k+1}$, and $\mathscr{PR}_i(\omega)=\vec{\mu}$. Part (a) of the proposition says that player i satisfies cautiousness under $\vec{\mu}$ in the sense of Catonini and de Vito [8] and Perea [19]: for all strategy profiles $\vec{\tau}_{-i}\in\Sigma_{-i}$, we have $\vec{\mu}([\![play(\tau_{-i})]\!]_M)\neq\vec{0}$. Part (b) says that, if there are at least two players not all of whose strategies survive iterated deletion, then the formulas $GRAT_i^h$ for $h=1,2,3,\ldots$, are mutually exclusive. Part (c) says that for all $h\leq k$, strategy profiles compatible with $\mathscr{L}^0$-$GRAT^h_{-i}$ are infinitely more likely those not compatible with $\mathscr{L}^0$-$GRAT^h_{-i}\vee\ldots\vee\mathscr{L}^0$-$GRAT^k_{-i}$ under $\vec{\mu}$. But our sense of "infinitely more likely than" is weaker than that of Blume, Brandenburger, and Dekel [4], and closer in spirit to that of Lo [16]. Formally, we use the notion of domination, where event E μ*-dominates* F, written $E\gg_{\vec{\mu}}F$, if $\min\{\ell:\mu_\ell(E)>0\}<\min\{\ell:\mu_\ell(F)>0\}$ (where we take $\min(\emptyset)=\infty$), rather than BFK's notion of assumption. However, as we discuss below, we could have used assumption here as well.

[5] It follows from the proof that we can take all the LPSs in M^k to have length $k+1$.

Proposition 4.2: *Suppose that M is an appropriate model for game Γ, $(M,\omega) \models \mathscr{L}^0$-$GRAT_i^{k+1}$, and $\mathscr{PR}_i(\omega) = \vec{\mu} = (\mu_0,\ldots,\mu_m)$.*

(a) *For all strategy profiles $\vec{\tau}_{-i} \in \Sigma_{-i}$, we have $\vec{\mu}([\![play(\tau_{-i})]\!]_M) \neq \vec{0}$.*

(b) *If $NWD_j^1 \neq NWD_j^0$ for at least two players j, then $(M,\omega) \models \neg\mathscr{L}^0$-$GRAT_i^1 \wedge \ldots \wedge \neg\mathscr{L}^0$-$GRAT_i^k$.*

(c) *If $h < h' \leq k$, then $[\![H_{-i}^{h'}]\!]_M \gg_\mu [\![H_{-i}^h]\!]_M$, where $H_{-i}^{k'}$ is an abbreviation of the formula $GRAT_{-i}^{k'} \wedge \neg GRAT_{-i}^{k'+1} \wedge \ldots \wedge \neg GRAT_{-i}^k$ (so H_{-i}^k is $GRAT_{-i}^k$). Moreover, for all $h \leq k$ and strategy profiles $\vec{\tau}_{-i}$ and $\vec{\tau}'_{-i}$, if $[\![GRAT_{-i}^h \wedge play(\vec{\tau}_{-i})]\!]_M \neq \emptyset$, and $[\![(GRAT_{-i}^h \vee \ldots \vee GRAT_{-i}^k) \wedge play(\vec{\tau}'_{-i})]\!]_M = \emptyset$, then $[\![play(\vec{\tau}_{-i})]\!]_M \gg_{\vec{\mu}} [\![play(\vec{\tau}'_{-i})]\!]_M$.*

It follows from part (b) that if some strategies of at least two players are weakly dominated, then the analogue of common assumption of rationality cannot hold. There is no state where $GRAT_i^k$ holds for all k; indeed, there is not even a state where $GRAT_i^k$ holds for all sufficiently large k. (The same comment applies to the $\mathscr{L}^0$-RAT_i^k operators used in Theorem 3.1.) By way of contrast, Catonini and de Vito [8] and Perea [19] show that their variants of common assumption do hold, while for BFK, k-level assumption for all k larger that some k^* holds (but which k^* it is depends on the game Γ). Unlike Catonini and DeVito and BFK, but like Perea, we are able to characterize IA using only finite structures. However, as we mentioned earlier, Perea's result holds only if there are only finitely many types. By way of contrast, our result holds for arbitrary models. Specifically, part (b) of Theorem 3.2 applies to arbitrary models; as long as $\mathscr{L}^0$-$GRAT_i^k$ holds in state ω of model M, then $\mathbf{s}_i(\omega)$ survives k rounds of iterated deletion. This is true even if M is a model that ascribes positive probability to all possible beliefs of other agents (see Section 5 for further discussion of this issue); in particular, it is true of *canonical model* for the full language of beliefs, which is the analogue of the complete structures considered by BFK. (See [12] for more discussion of this issue.)

Another difference between Perea's characterization and ours is that we have different notions of caution. For Perea's notion of $(k+1)$*-fold assumption of rationality* to hold for player i at a state ω, each strategy profile $\vec{\tau}_{-i}$ compatible with a k-fold assumption of rationality must get positive probability (i.e., if $\Pr_i(\omega) = \vec{\mu} = (\mu_0,\ldots,\mu_m)$, then $\mu_h(\vec{\tau}_{-i}) > \vec{0}$ for some h). On the other hand, if $GRAT_i^{k+1}$ holds at ω, then for each strategy profile $\vec{\tau}_{-i}$ compatible with $GRAT_{-i}^k$ we have $\mu_0(\vec{\tau}_{-i}) > 0$. The fact that we require $\mu_0(\vec{\tau}_{-i}) > 0$ rather than just $\mu_h(\vec{\tau}_{-i}) > 0$ for some h will play an important role in the characterization of IA given in the next section that uses only standard probability. Perea's approach does not lead to an obvious analogue of that result.

Finally, we return to the issue of assumption. As we observed earlier, the notion of belief that we use is weaker than assumption. However, we could replace belief everywhere by assumption to get a notion of "all i assumes", and a generalization of it corresponding to our $B_i(\varphi_1,\ldots,\varphi_k)$. Our results we continue to hold if we used this (generalized) "all i assumes" rather than all i knows. In particular, it is easy to check that the finite model $\bar{M}^k$ constructed in the proof of Theorem 3.2 satisfies the "all i assumes" analogue of $\mathscr{L}^0$-$GRAT_i^k$ at states where the strategy used by i satisfies k round of iterated deletion.

5 Using approximate belief and probability structures

While the approach described in Section 3.2 deals with Samuelson's concern, it does so by assuming that the agents' beliefs are characterized by LPSs. Our earlier approach characterized IA using (standard) probability structures, but did not deal with Samuelson's concerns. We now show that we can characterize IA using standard probability structures, while still dealing with Samuelson's concern, by considering approximate belief in an appropriate sense.

We start with a quantitative analogues of the belief operators $B_i, \langle B_i \rangle$ and also define a conditional belief operators. Just as we did in the previous section, for simplicity, we restrict our attention to fully measurable structures. If $M = (\Omega, \mathbf{s}, \mathscr{F}, \mathscr{PR}_1, \ldots, \mathscr{PR}_n)$ is a fully measurable probability structure, then

- $(M,\omega) \models B_i^\delta \varphi$ if $\mathscr{PR}_i(\omega)([\![\varphi]\!]_M) \geq 1-\delta$, and $(M,\omega) \models \langle B_i \rangle^\delta \varphi$ if $\mathscr{PR}_i(\omega)([\![\varphi]\!]_M) \geq \delta$,
- $(M,\omega) \models B_i^\delta(\varphi \mid \theta)$ if $\mathscr{PR}_i(\omega)([\![\theta]\!]_M) > 0$ and $\mathscr{PR}_i(\omega)([\![\varphi]\!]_M \mid [\![\theta]\!]_M) \geq 1-\delta$, and analogously for $\langle B_i \rangle^\delta(\varphi \mid \theta)$.

That is, $B_i^\delta \varphi$ means that player i is "almost certain" that φ holds—i assigns probability at least $1-\delta$ to φ holding—and $B_i^\delta(\varphi \mid \theta)$ means that if i learns that θ holds, then i is almost certain that φ holds.

The analogous "all I approximately know" operator, $O_i^{(\mathscr{L},\delta,\varepsilon)}$, takes two parameters, δ and ε. As with the approximate belief operator B_i^δ, the δ tells us how close to 1 agent i's beliefs have to be. The ε gives us a lower bound on how likely each formula in $\mathscr{L}$ consistent with what is believed must be. Again, we also consider a conditional version of the operator. Define $O_i^{\mathscr{L},\nu,\varepsilon}\varphi$ (read "all agent i approximately knows with respect to $\mathscr{L}$ is φ") to be an abbreviation for

$$B_i^\delta(\varphi) \wedge (\wedge_{\psi \in \mathscr{L}}(\Diamond(\varphi \wedge \psi) \Rightarrow \langle B_i \rangle^\varepsilon(\psi))).$$

and define $O_i^{\mathscr{L},\delta,\varepsilon}(\varphi \mid \theta)$ (read "if agent i were to find out that θ holds, then all agent i approximately knows with respect to $\mathscr{L}$ is φ") to be an abbreviation for

$$B_i^\delta(\varphi \mid \theta) \wedge (\wedge_{\psi \in \mathscr{L}}(\Diamond(\varphi \wedge \psi \wedge \theta) \Rightarrow \langle B_i \rangle^\varepsilon(\psi \mid \theta))).$$

Finally, let $O_i^{\mathscr{L},\nu,\varepsilon}(\varphi_1, \ldots, \varphi_\ell)$ be an abbreviation for

$$O_i^{\mathscr{L},\delta,\varepsilon}(\varphi_1) \wedge O_i^{\mathscr{L},\delta,\varepsilon}(\varphi_2 \mid \neg\varphi_1) \wedge \ldots \wedge O_i^{\mathscr{L},\delta,\varepsilon}(\varphi_\ell \mid \neg\varphi_1 \wedge \ldots \wedge \neg\varphi_{\ell-1}).$$

To relate this definition to the definition in LPS structures, let $B_i^\delta(\varphi_1, \ldots, \varphi_\ell)$ be an abbreviation for

$$B_i^\delta(\varphi_1) \wedge B_i^\delta(\varphi_2 \mid \neg\varphi_1) \wedge \ldots \wedge B_i^\delta(\varphi_\ell \mid \neg\varphi_1 \wedge \ldots \wedge \neg\varphi_{\ell-1}).$$

Note that

$$\begin{aligned}(M,\omega) \models O_i^{\mathscr{L},\delta,\varepsilon}(\varphi_1, \ldots, \varphi_\ell) \text{ iff } \ & (M,\omega) \models B_i^\delta(\varphi_1, \ldots, \varphi_\ell) \text{ and,} \\ & \text{for all } \psi \in \mathscr{L}, \text{ if } (M,\omega) \models \Diamond(\varphi_1 \wedge \psi) \text{ then } \mu([\![\psi \mid \varphi_1]\!]_M) > \varepsilon \text{ and,} \\ & \text{for all } h \text{ with } 2 \leq h \leq \ell, \text{ if } (M,\omega) \models \Diamond(\neg\varphi_1 \wedge \ldots \wedge \neg\varphi_{h-1} \wedge \varphi_h \wedge \psi) \\ & \quad \text{then } (\mu([\![\psi]\!]_M \mid [\![\neg\varphi_1 \wedge \ldots \wedge \neg\varphi_{h-1}]\!]_M)_0 > \varepsilon.\end{aligned}$$

Thus, $O_i^{\mathscr{L},\delta,\varepsilon}(\varphi_1, \ldots, \varphi_\ell)$ really is the "approximate" analogue of $O_i^{\mathscr{L}}(\varphi_1, \ldots, \varphi_\ell)$. We now define the formulas $GRAT_i^{k,\delta,\varepsilon}$ in exactly the same way as $\mathscr{L}^0$-$GRAT^k$ except that we replace the LPS-based $O_i^{\mathscr{L}}$ operator with $O_i^{\mathscr{L},\delta,\varepsilon}$. In more detail, define $GRAT_i^{0,\delta,\varepsilon}$ to be *true*, and $GRAT_i^{k,\delta,\varepsilon}$ to be an abbreviation of

$$RAT_i \wedge O_i^{\mathscr{L}^0_{-i},\delta,\varepsilon}(GRAT_{-i}^{k,\delta,\varepsilon}, \ldots, GRAT_{-i}^{0,\delta,\varepsilon}).$$

That is, all agent i approximately knows is that all other agents are k-level rational, but if i were to find out that they are not, then all i approximately knows is that they are $(k-1)$-level rational and so on.

Theorem 5.1: *For all finite games Γ and all sufficiently small $\varepsilon > 0$, there exists some $\delta > 0$ such that the following are equivalent:*

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

(b) there exists a fully measurable structure M^k appropriate for Γ and a state ω^k in M^k such that $\mathbf{s}_i(\omega^k) = \sigma$ and $(M^k, \omega^k) \models GRAT_i^{k,\delta,\varepsilon}$.

In addition, if $\vec{\sigma} \in NWD^k$, then for all sufficiently small $\varepsilon > 0$, there exists some $\delta > 0$ and a fully measurable structure $\bar{M}^{k,\delta} = (\Omega^{k,\delta}, \mathbf{s}^{k,\delta}, \mathscr{F}^{k,\delta}, \mathscr{PR}_1^{k,\delta}, \ldots, \mathscr{PR}_n^{k,\delta})$ such that $\Omega^{k,\delta} = \{(k', i, \vec{\sigma}) : 0 \le k' \le k, 1 \le i \le n, \vec{\sigma} \in NWD^{k'}\}$, $\mathbf{s}^{k,\delta}(k', i, \vec{\sigma}) = \vec{\sigma}$, $\mathscr{F}^{k,\delta} = 2^{\Omega^k}$, and for all states $(k', i, \vec{\sigma}) \in \Omega^{k,\delta}$, we have $(\bar{M}^{k,\delta}, (k', i, \vec{\sigma})) \models GRAT_{-i}^{k',\delta,\varepsilon}$.[6]

6 Discussion

We have used the "all I know" operator introduced in our earlier paper to provide an epistemic characterization of IA that deals with Samuelson's conceptual concerns. We actually provided two characterizations, one in LPS structures and one in probability structures. The former uses a generalized belief operator, while the latter uses a generalized approximate belief operator. These operators may be of independent interest. For example, a logic with a generalized belief operator may be an appropriate logic in which to describe belief revision [**?**, **?**] and iterated belief revision [**?**], since it allows us to describe how beliefs would be revised. It clearly has deep connections with counterfactual reasoning as well. For example, in a logic of counterfactuals, a formula such as $B_i(\varphi_1, \varphi_2)$ can be viewed as an abbreviation of $B_i(\varphi_1) \wedge (\neg B_i \varphi_1 \rightarrow_i B_i \varphi_2)$, where $\rightarrow_i$ is a counterfactual operator (see [**?**] for a discussion of and semantics for this standard operator); the formula $B_i(\varphi_1, \ldots, \varphi_k)$ can be expressed using counterfactuals in a similar way. It would of interest to axiomatize the logic of generalized belief.

The more quantitative operator B_i^δ may also be of independent interest. Interestingly, in *cognitive hierarchy theory* (CHT) [**?**], there are assumed to be different types of players: roughly speaking, *level-k* players are assumed to be k-level rational, and players assign probabilities to a player being of level-k. Whereas in our characterization of IA, level-k players are assigned the highest probability, followed by level-$k-1$, and so on, in CHT it is the other way around. In any case, having an operator like B_i^δ may allow a more realistic characterization of players beliefs than a purely qualitative generalized belief operator.

Finally, while we have focused here only on IA, in other work [**?**], we have also used the notion of all I know to characterize Pearce's notion of *extensive-form rationalizability* [18], a well-studied solution concept in extensive-form games that also involves iterated deletion. That characterization too used a variant of the $PLAYCON_i$ formula, and thus does not address Samuelson's concerns. Although we have not yet checked details, it seems that we should also be able to get a characterization of extensive-form rationalizability using the techniques of this paper. All this suggests that thinking in terms of an "all I know" operator and generalized belief may provide further insights into solution concepts.

Acknowledgements

The first author is supported in part by NSF grants IIS-178108 and IIS-1703846, a grant from the Open Philanthropy Foundation, and ARO grant W911NF-17-1-0592. The second author is supported in part by NSF grant IIS-1703846. We thank Andres Perea and Burkhard Schipper for their helpful comments.

[6] How small ε has to be depends only on the game. Although the choice of δ depends on the choice of ε, ε plays no role in the construction of $\bar{M}^{k,\delta}$ (which is why we did not write $\bar{M}^{k,\delta,\varepsilon}$).

References

[1] P. Battigalli (1997): *On rationalizability in extensive games*. Journal of Economic Theory 74, pp. 40–61.

[2] P. Battigalli & M. Siniscalchi (2002): *Strong beliefs and forward-induction reasoning*. Journal of Economic Theory 106, pp. 356–391.

[3] B. D. Bernheim (1984): *Rationalizable strategic behavior*. Econometrica 52(4), pp. 1007–1028.

[4] L. E. Blume, A. Brandenburger & E. Dekel (1991): *Lexicographic probabilities and choice under uncertainty*. Econometrica 59(1), pp. 61–79, doi:10.2307/2938240.

[5] L. E. Blume, A. Brandenburger & E. Dekel (1991): *Lexicographic probabilities and equilibrium refinements*. Econometrica 59(1), pp. 81–98, doi:10.2307/2938241.

[6] A. Brandenburger & E. Dekel (1987): *Rationalizability and correlated equilibria*. Econometrica 55, pp. 1391–1402.

[7] A. Brandenburger, A. Friedenberg & J. Keisler (2008): *Admissibility in games*. Econometrica 76(2), pp. 307–352, doi:10.1111/j.1468-0262.2008.00835.x.

[8] E. Catonini & N. de Vito (2018): *Cautious belief and iterated admissibility*. Unpublished manuscript.

[9] S. Galanis & P. Barelli (2013): *Admissibility and event-rationality*. Games and Economic Behavior 77(1), pp. 21–40.

[10] J. Y. Halpern (2010): *Lexicographic probability, conditional probability, and nonstandard probability*. Games and Economic Behavior 68(1), pp. 155–179, doi:10.1016/j.geb.2009.03.013.

[11] J. Y. Halpern & G. Lakemeyer (2001): *Multi-agent only knowing*. Journal of Logic and Computation 11(1), pp. 41–70, doi:10.1093/logcom/11.1.41.

[12] J. Y. Halpern & R. Passjh (2009): *A logical characterization of iterated admissibility*. In: Theoretical Aspects of Rationality and Knowledge: Proc. Twelfth Conference (TARK 2009), pp. 146–155, doi:10.1145/1562814.1562836.

[13] H. J. Keisler & B.S. Lee (2015): *Common assumption of rationality*. Unpublished manuscript.

[14] B. S. Lee (2016): *Admissibility and assumption*. Journal of Economic Theory 163, pp. 42–72, doi:10.1016/j.jet.2016.01.006.

[15] H. J. Levesque (1990): *All I know: a study in autoepistemic logic*. Artificial Intelligence 42(3), pp. 263–309, doi:10.1016/0004-3702(90)90056-6.

[16] K.C. Lo (1999): *Nash equilibrium without mutual knowledge of rationality*. Economic Theory 14(3), pp. 621–633, doi:10.1007/s001990050344.

[17] M. J. Osborne & A. Rubinstein (1994): *A Course in Game Theory*. MIT Press, Cambridge, MA.

[18] D. G. Pearce (1984): *Rationalizable strategic behavior and the problem of perfection*. Econometrica 52(4), pp. 1029–1050, doi:10.2307/1911197.

[19] A. Perea (2012): *Epistemic Game Theory*. Cambridge University Press, Cambridge, U.K., doi:10.1017/CBO9780511844072.

[20] L. Samuelson (1992): *Dominated strategies and common knowledge*. Games and Economic Behavior 4, pp. 284–313, doi:10.1016/0899-8256(92)90020-S.

[21] M. Shimoji (2004): *On the equivalence of weak dominance and sequential best response*. Games and Economic Behavior 48, pp. 385–402.

[22] M. Shimoji & J. Watson (1998): *Conditional dominance, rationalizability, and game forms*. Journal of Economic Theory 83, pp. 161–195.

[23] D. Stahl (1995): *Lexicgraphic rationalizability and iterated admissibility*. Economic Letters 47, pp. 155–159, doi:10.1016/0165-1765(94)00530-F.

[24] T. C.-C. Tan & S. R. da C. Werlang (1988): *The Bayesian foundation of solution concepts of games*. Journal of Economic Theory 45(45), pp. 370–391.

[25] C. Yang (2016): *Weak assumption and iterative admissibility*. Journal of Economic Theory 158, pp. 87–101, doi:10.1016/j.jet.2015.03.009.