Epistemic Game Theory: Complete Information*

Adam Brandenburger†

First Version: October 2005
Current Version: May 2007

The epistemic program can be viewed as a methodical construction of game theory from its most basic elements–rationality and irrationality, belief and knowledge about such matters, beliefs about beliefs, knowledge about knowledge, and so on. To date, the epistemic field has been mainly focused on game matrices and trees–i.e., on the non-cooperative branch of game theory. It has been used to provide foundations for existing non-cooperative solution concepts, and also to uncover new solution concepts. The broader goal of the program is to provide a method of analyzing different sets of assumptions about games in a precise and uniform manner.

1 Epistemic Analysis

Under the epistemic approach, the traditional description of a game is augmented by a mathematical framework for talking about the rationality or irrationality of the players, their beliefs and knowledge, and related ideas.

The first step is to add sets of **types** for each of the players. The apparatus of types goes back to Harsanyi [40, 1967-68], who introduced it as a way to talk formally about the players' beliefs about the payoffs in a game, their beliefs about other players' beliefs about the payoffs, and so on. (See EPISTEMIC GAME THEORY: INCOMPLETE INFORMATION.) But the technique is equally useful to talk about uncertainty about the actual play of the game–i.e., about the players' beliefs about the strategies chosen in the game, their beliefs about other players' beliefs about the strategies, and so on. This survey will focus on this second source of uncertainty. It is also possible to treat both kinds of uncertainty together, using the same technique.

*To appear in *The New Palgrave Dictionary of Economics*, 2nd edition, edited by Steven Durlauf and Larry Blume, Macmillan, 2007. This survey is based on Brandenburger [27, 2007]. I am grateful to Springer for permission to use this material. I owe a great deal to joint work and many conversations with Robert Aumann, Eddie Dekel, Amanda Friedenberg, Jerry Keisler, and Harborne Stuart. My thanks to Konrad Grabiszewski for important input, John Nachbar for very helpful editorial advice, and Michael James for valuable assistance. The Stern School of Business provided financial support. egtc-05-13-07

†Address: Stern School of Business, New York University, New York, NY 10012, adam.brandenburger@stern.nyu.edu, www.stern.nyu.edu/∼abranden

We give a definition of a type structure as commonly used in the epistemic literature, and an example of its use.

Fix an n-player finite strategic-form game $\langle S^1, \ldots S^n, \pi^1, \ldots, \pi^n \rangle$. Some notation: Given sets $X^1, \ldots, X^n$, let $X = \times_{i=1}^n X^i$ and $X^{-i} = \times_{j \neq i} X^j$. Also, given a finite set Ω, write $\mathcal{M}(\Omega)$ for set of all probability measures on Ω.

Definition 1.1 *An $(S^1, \ldots, S^n)$-**based (finite) type structure** is a structure*

$$\langle S^1, \ldots, S^n; T^1, \ldots, T^n; \lambda^1, \ldots, \lambda^n \rangle,$$

*where each T^i is a finite set, and each $\lambda^i : T^i \to \mathcal{M}(S^{-i} \times T^{-i})$. Members of T^i are called **types** for player i. Members of $S \times T$ are called **states (of the world)**.*[1]

A particular state $(s^1, t^1, \ldots, s^n, t^n)$ describes the strategy chosen by each player, and also each player's type. Moreover, a type t^i for player i induces, via a natural induction, an entire hierarchy of beliefs–about the strategies chosen by the players $j \neq i$, about the beliefs of the players $j \neq i$, and so on. (See EPISTEMIC GAME THEORY: BELIEFS AND TYPES.)

Example 1.1 (A coordination game) *Consider the coordination game in Figure 1.1 (where Ann chooses the row and Bob the column), and the associated type structure in Figure 1.2.*[2]

	L	R
U	2, 2	0, 0
D	0, 0	1, 1

Figure 1.1

$\lambda^a(t^a)$

T^b \ S^b	L	R
u^b	0	½
t^b	0	½

$\lambda^a(u^a)$

T^b \ S^b	L	R
u^b	½	0
t^b	0	½

$\lambda^b(t^b)$

T^a \ S^a	U	D
u^a	0	½
t^a	0	½

$\lambda^b(u^b)$

T^a \ S^a	U	D
u^a	½	0
t^a	0	½

Figure 1.2

[1] For some purposes (see, e.g., Sections 4 and 6 below), it is important to consider infinite type structures. Topological assumptions are then made on the type spaces T_i.

[2] Similar to an example in Aumann and Brandenburger [9, 1995, pp.1166-1167].

There are two types t^a, u^a for Ann, and two types t^b, u^b for Bob. The measure associated with each type is as shown. Fix the state $(D,\ t^a, R,\ t^b)$. At this state, Ann plays D and Bob plays R. Ann is 'correct' about Bob's strategy. (Her type t^a assigns probability 1 to Bob's playing R.) Likewise, Bob is correct about Ann's strategy. Ann, though, thinks it possible Bob is wrong about her strategy. (Her type assigns probability 1/2 to type u^b for Bob, which assigns probability 1/2 to Ann's playing U, not D.) Again, likewise with Bob.

What about the rationality or irrationality of the players? At state $(D,\ t^a, R,\ t^b)$, Ann is rational. Her strategy maximizes her expected payoff, given her first-order belief (which assigns probability 1 to R). Likewise, Bob is rational. Ann, though, thinks it possible Bob is irrational. (She assigns probability 1/2 to (R, u^b). With type u^b, Bob gets a higher expected payoff from L than R.) The situation with Bob is again symmetric.

Summing up, the example is just a description of a game situation, not a prediction. A type structure is a descriptive tool. Note, too, that the example includes both rationality and irrationality, and also allows for incorrect as well as correct beliefs (e.g., Ann thinks it possible Bob is irrational, though in fact he isn't). These are typical features of the epistemic approach.

Two comments on type structures: First, we can ask whether Definition 1.1 above is to be taken as primitive or derived. Arguably, hierarchies of beliefs are the primitive, and types are simply a convenient tool for the analyst. See EPISTEMIC GAME THEORY: BELIEFS AND TYPES for further discussion.

Second, note that Definition 1.1 applies to finite games. These will be the focus of this survey. There is nothing yet approaching a developed literature on epistemic analysis of infinite games.

2 Early Results

A major use of type structures is to identify conditions on the players' rationality, beliefs, etc. that yield various solution concepts.

A very basic solution concept is iterated dominance. This involves deleting from the matrix all strongly dominated strategies, then deleting all strategies that become strongly dominated in the resulting submatrix, and so on until no further deletion is possible.[3] Call the remaining strategies the *iteratively undominated* (IU) strategies. There is a basic equivalence: A strategy is not strongly dominated if and only if there is a probability measure on the product of the other players' strategy sets under which it is optimal. Using this, IU can also be defined as follows: Delete from the matrix any strategy that isn't optimal under some measure on the product of the other players' strategy sets. Consider the resulting submatrix and delete strategies that don't pass this test on the submatrix, and so on.

The second definition suggests what a formal epistemic treatment of IU should look like. A rational player will choose a strategy which is optimal under some measure. This is the first round

[3]It is easy to check that in finite games (as considered in this survey), the residual set will always be nonempty.

of deletion. A player who is rational and believes the other players are rational will choose a strategy which is optimal under a measure that assigns probability 1 to the strategies remaining after the first round of deletion. This gives the second round of deletion. And so on.

Type structures allow a formal treatment of this idea. First the formal definition of rationality. This is a property of strategy-type pairs. Say (s^i, t^i) is **rational** if s^i maximizes player i's expected payoff under the marginal on S^{-i} of the measure $\lambda^i(t^i)$.

Say type t^i of player i **believes** an event $E \subseteq S^{-i} \times T^{-i}$ if $\lambda^i(t^i)(E) = 1$, and write

$$B^i(E) = \{t^i \in T^i : t^i \text{ believes } E\}.$$

Now, for each player i, let R_1^i be the set of all rational pairs (s^i, t^i), and for $m > 0$ define R_m^i inductively by

$$R_{m+1}^i = R_m^i \cap [S^i \times B^i(R_m^{-i})].$$

Definition 2.1 *If $(s^1, t^1, \ldots, s^n, t^n) \in R_{m+1}$, say there is* ***rationality and*** m***th-order belief of rationality (R***m***BR)*** *at this state. If $(s^1, t^1, \ldots, s^n, t^n) \in \bigcap_{m=1}^{\infty} R_m$, say there is* ***rationality and common belief of rationality (RCBR)*** *at this state.*

These definitions yield an epistemic characterization of IU: *Fix a type structure and a state $(s^1, t^1, \ldots, s^n, t^n)$ at which there is RCBR. Then the strategy profile $(s^1, \ldots, s^n)$ is IU. Conversely, fix an IU profile $(s^1, \ldots, s^n)$. There is a type structure and a state $(s^1, t^1, \ldots, s^n, t^n)$ at which there is RCBR.* Results like this can be found in the early literature–see, among others, Brandenburger-Dekel [28, 1987] and Tan-Werlang [59, 1988].

An important stimulus to the early literature was the pair of papers by Bernheim [16, 1984] and Pearce [48, 1984], which introduced the solution concept of *rationalizability.* This differs from IU by requiring on each round that a player's probability measure on the product of the other players' (remaining) strategy sets be a product measure–i.e., be independent. Thus the set of rationalizable strategy profiles is contained in the IU set. It is well known that there are games (with three or more players) in which inclusion is strict.

The argument for the independence assumption is that in non-cooperative game theory it is supposed that players do not coordinate their strategy choices. Interestingly though, correlation is consistent with the non-cooperative approach. This view is put forward in Aumann [4, 1987]. (Aumann [3, 1974] introduced the study of correlation into non-cooperative theory.) Consider an analogy to coin tossing. A correlated assessment over coin tosses is possible, if there is uncertainty over the coin's parameter or 'bias.' (The assessment is usually required to be conditionally i.i.d., given the parameter.) Likewise, in a game, Charlie might have a correlated assessment over Ann's and Bob's strategy choices, because, say, he thinks Ann and Bob have observed similar signals before the game (but is uncertain what the signal was).

The same epistemic tools used to understand IU can be used to characterize other solution concepts on the matrix. Aumann-Brandenburger [9, 1995, Preliminary Observation] point out that pure-strategy Nash equilibrium is characterized by the simple condition that each player is rational and assigns probability 1 to the actual strategies chosen by the other players. (Thus, in Example 1.1 above, these conditions hold at the state (D, t^a, R, t^b), and (D, R) is indeed a Nash equilibrium.) As far as mixed strategies are concerned, in the epistemic approach to games these don't play the central role that they do under equilibrium analysis. Built into the set-up of Section 1 is that each player makes a definite choice of (pure) strategy.[4] It is the other players who are uncertain about this choice. Harsanyi [41, 1973] originally proposed this shift in thinking about randomization. Aumann-Brandenburger [9, 1995] give an epistemic treatment of mixed-strategy Nash equilibrium along these lines.

Aumann [4, 1987] asks a question about an outside observer of a game. He provides conditions under which the observer's assessment of the strategies chosen will be the distribution of a correlated equilibrium (as defined in his [3, 1974] paper). The distinctive condition in [4, 1987] is the so-called Common Prior Assumption, which says that the probability assessment associated with each player's type is the same as the observer's assessment, except for being conditioned on what the type in question knows. A number of papers have investigated foundations for this assumption–see, among others, Morris [45, 1994], Samet [53, 1998], [54, 1998], Bonanno-Nehring [24, 1999], Feinberg [33, 2000], Halpern [39, 2002], and also the exchange between Gul [36, 1998] and Aumann [8, 1998].

3 Next Steps: The Tree

An important next step in the epistemic program was extending the analysis to game trees. A big motivation for this was to understand the logical foundation of backward induction (BI). At first sight, BI is one of the easiest ideas in game theory. If Ann, the last player to move, is rational, she will make the BI choice. If Bob, the second-to-last player to move, is rational and thinks Ann is rational, he will make the choice that is maximal given that Ann makes the BI choice–i.e., he too will make the BI choice. And so on back in the tree, until the BI path is identified (Aumann [5, 1995]).

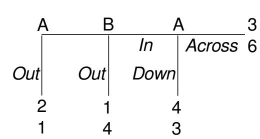

Figure 3.1

For example, Figure 3.1 is three-legged Centipede (Rosenthal [51, 1981]). (The top payoffs are Ann's, and the bottom payoffs are Bob's.) BI says Ann plays *Out* at her first node. But what if she doesn't? How will Bob react? Perhaps Bob will conclude that Ann is an irrational player, who plays *Across*. That is, Bob might play *In*, hoping to get a payoff of 6 (better than 4 from *Out*).

[4]If a player does have the option of making a randomized choice, this can be added to the (pure) strategy set. Indeed, in a finite game, a finite number of such choices can be added.

Perhaps, anticipating this, Ann will in fact play *Down*, hoping to get 4 (better than 2 from playing *Out*).

Many papers have examined this conceptual puzzle with BI–see, among others, Binmore [19, 1987], Bicchieri [17, 1988], [18, 1989], Basu [11, 1990], Bonanno [23, 1991], and Reny [49, 1992].

A key step in resolving the puzzle is extending the epistemic tools of Section 1, to be able to talk formally about rationality, beliefs etc. in the tree.

Example 3.1 (Three-Legged Centipede) *Figure 3.2 is a type structure for Three-Legged Centipede.*

$\lambda^a(t^a)$

T^b \ S^b	*Out*	*In*
u^b	0	0
t^b	0	1

$\lambda^a(u^a)$

T^b \ S^b	*Out*	*In*
u^b	1 [0]	0 [0]
t^b	0 [0]	0 [1]

$\lambda^b(t^b)$

T^a \ S^a	*Out*	*Down*	*Across*
u^a	1 [0]	0 [0]	0 [0]
t^a	0 [0]	0 [0]	0 [1]

$\lambda^b(u^b)$

T^a \ S^a	*Out*	*Down*	*Across*
u^a	1 [0]	0 [0]	0 [0]
t^a	0 [0]	0 [1]	0 [0]

Figure 3.2

There are two types t^a, u^a for Ann. Type t^a for Ann has the measure shown in the top-left matrix. It assigns probability 1 to (In, t^b) for Bob. Type u^a has two associated measures–shown in the top-right matrix. The first measure (the numbers without parentheses) assigns probability 1 to (Out, u^b) for Bob. In this case, we also specify a second measure for Ann, because we want to specify what Ann thinks at her second node, too. Reaching this node is assigned positive probability (in fact, probability 1) under Ann's type t^a, but probability 0 under her type u^a. So, for type u^a, there isn't a well-defined conditional probability measure at Ann's second node. This is why we (separately) specify a second measure for Ann's type u^a: it is the measure in square brackets. If type u^a, Ann assigns probability 1 to (In, t^b) at her second node.

There are also two types t^b, u^b for Bob. Both types initially assign probability 1 to Ann's playing Out. For both of Bob's types, there isn't a well-defined conditional probability measure at his node. At his node, Bob's type t^b assigns probability 1 to $\{(Across,\ t^a)\}$, while his type u^b assigns probability 1 to $\{(Down,\ t^a)\}$.

This is a simple illustration of the concept of a *conditional probability system* (*CPS*), due to Rényi [50, 1955]. A CPS specifies a family of conditioning events E and a measure p_E for each such event, together with certain restrictions on these measures. The interpretation is that p_E is what the player believes, after observing E. Even if $p_\Omega(E) = 0$ (where Ω is the entire space), the measure p_E is still specified. That is, even if E is 'unexpected,' the player has a measure if E nevertheless happens. This is why CPS's are well-suited to epistemic analysis of game trees–where we need to be able to describe how players react to the unexpected.

Myerson [47, 1991, Ch.1] provided a preference-based axiomatization of a class of CPS's. Battigalli and Siniscalchi [13, 1999], [14, 2002] further developed both the pure theory and the game-theoretic application of CPS's (see below).

Suppose the true state in Figure 3.2 is (*Down*, t^a, *In*, t^b). In particular, Ann plays *Down*, expecting Bob to play *In*. Bob plays *In*, expecting (at his node) Ann to play *Across*. Ann expects a payoff of 4 (and gets this). Bob expects a payoff of 6 (but gets only 3). In everyday language, we can says that Ann successfully bluffs Bob.[5]

But what about epistemic conditions? Are the players rational in this situation? Does each think the other is rational? Etc.

To answer, we need a definition of rationality with CPS's. Fix a strategy-type pair (s^i, t^i), where t^i is associated with a CPS. Call this pair **rational** (**in the tree**) if the following holds: Fix any information set H for i allowed by s^i, and look at the measure on the other players' strategies, given H. (This means given the event that the other players' strategies allow H.) Require that s^i maximizes i's expected payoff under this measure, among all strategies r^i of i that allow H.

With this definition, the rational strategy-type pairs in Figure 3.2 are (*Down*, t^a), (*Out*, u^a), (*In*, t^b), and (*Out*, u^b).

Next, what does Ann think about Bob's rationality? To answer, we need a CPS-analog to belief (as defined in Section 2). Ben Porath [15, 1997] proposed the following:[6] Say player i **initially believes** event E if, under i's CPS, E gets probability 1 at the root of the tree. (Formally, the conditioning event consists of all strategy profiles of the other players.) Battigalli-Siniscalchi [14, 2002] strengthened this definition to: Say player i **strongly believes** event E if, under i's CPS, E gets probability 1 at every information set at which E is possible. Under initial belief, E also gets probability 1 at any information set H that gets positive probability under i's initial measure (i.e., i's measure given the root). This is just standard conditioning on non-null events. But under strong belief, this conclusion holds for any information set H which has a nonempty intersection with E–even if H is null under i's initial measure. This is why strong belief is stronger than initial belief.

Let us apply these definitions to Figure 3.2. Does Ann initially believe that Bob is rational? Yes. Both of Ann's types initially believe Bob is rational. Type t^a initially assigns probability 1 to the rational pair (*In*, t^b). Type u^a initially assigns probability 1 to the rational pair (*Out*, u^b). In fact, both types strongly believe Bob is rational. Since, under type t^a, Ann's second node gets positive probability (in fact, probability 1) under her initial measure, we need only check this for type u^a. But at Ann's second node, type u^a assigns probability 1 to the rational pair (*In*, t^b).

Turning to Bob, both of his types initially believe that Ann is rational. Type u^b even strongly believes Ann is rational; but type t^b doesn't. This is because, at Bob's node, type t^b assigns positive probability (in fact, probability 1) to the irrational pair (*Across*, t^a).

[5] At the state (*Down*, t^a, *In*, t^b), the bluff works. By contrast, at the state (*Down*, t^a, *Out*, u^b), Ann attempts the bluff and it fails.

[6] We have taken the liberty of changing terminology, consistency with "strong belief" below.

Staying with initial belief (we come back to strong belief below), we can parallel Definition 2.1 and define inductively **rationality and mth-order initial belief of rationality** (**RmIBR**) at a state of a type structure, and **rationality and common initial belief of rationality** (**RCIBR**). (See Ben Porath [15, 1997].) In Figure 3.2, since all four types initially believe the other player is rational, a simple induction gives that at the state (*Down*, t^a, *In*, t^b) for instance, RCIBR holds.

In words, Ann plays across at her first node, believing (initially) that Bob will play *In*, so she can get a payoff of 4. Why would Bob play *In*? Because he initially believes that Ann plays *Out*. But in the probability-0 event that Ann plays across at her first node, Bob then assigns probability 1 to Ann's playing across at her second node–i.e., to Ann's being irrational. He therefore (rationally) plays *In*. All this is consistent with RCIBR.

4 Conditions for Backward Induction

Interestingly, this is exactly the line of reasoning which, as we said, was the original stimulus for investigating the foundations of BI. So, there is no difficulty with it–we've just seen a formal set-up in which it holds. The resolution of the BI puzzle is simply to accept that the BI path may not result.

But one can also argue that RCIBR is not the right condition: it is too weak. In the above example, Bob realizes that he might be 'surprised' in the play of the game–that's why he has a CPS, not just an ordinary probability measure. If he realizes he might be surprised, should he abandon his (initial) belief that Ann is rational when he is surprised? Bob's type t^b does so. This is the step taken by Battigalli-Siniscalchi [14, 2002] with their concept of strong belief. The argument says that we want t^b to strongly believe, not just initially believe, that Ann is rational. Type t^b will strongly believe Ann is rational if we move the probability-1 weight (in square brackets) on (*Across*, t^a) to (*Down*, t^a). But now (*In*, t^b) isn't rational for Bob, so Ann doesn't (even initially) believe Bob is rational. It looks as if the example unravels.

We can again parallel Definition 2.1 and define inductively **rationality and mth-order strong belief of rationality** (**RmSBR**), and **rationality and common strong belief of rationality** (**RCSBR**). (See Battigalli-Siniscalchi [14, 2002].) The question is then: Does RCSBR yield BI?

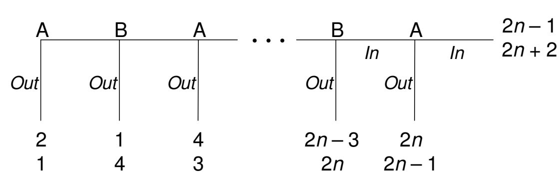

Figure 4.1

The answer is yes: *Fix a CPS-based type structure for n-legged Centipede (Figure 4.1), and a state at which there is RCSBR. Then Ann plays Out.* The result follows from Friedenberg [34, 2002], who shows that in a PI game (satisfying certain payoff restrictions), RCSBR yields a Nash-equilibrium outcome. In Centipede, there is a unique Nash path and it coincides with the BI path. Of course, this isn't true in general.

Example 4.1 (A second coordination game) *Consider the coordination game in Figure 4.2 and the associated CPS-based type structure in Figure 4.3.*

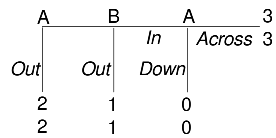

Figure 4.2

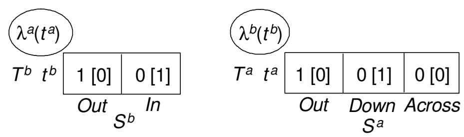

Figure 4.3

The rational strategy-type pairs are (Out, t^a) and (Out, t^b) for Ann and Bob respectively. Ann's type t^a strongly believes $\{$(Out, t^b)$\}$, and Bob's type t^b strongly believes $\{$(Out, t^a)$\}$. By induction, RCSBR holds at the state (Out, t^a,Out, t^b).

Here, the BI path need not be played under RCSBR. The key is to see that both (*Down*, t^a) and (*Across*, t^a) are irrational for Ann, since she (strongly) believes Bob plays *Out*. So at his node, Bob can't believe Ann is rational. If he considers it sufficiently more likely Ann will play *Down* rather than *Across*, he will rationally play *Out* (as happens). In short, if Ann doesn't play *Out*, she is irrational and so 'all bets are off' as to what she will do. She could play *Down*.

This situation may be surprising, at least at first blush, but there does not appear to be anything conceptually wrong with it. Indeed, it points to an interesting way in which the players in a game can literally be trapped by their beliefs–which here prevent them from getting their mutually preferred $(3, 3)$ outcome.

But one can also argue differently: If Ann forgoes the payoff of 2 she can get by playing *Out* at the first node, then surely she must be playing *Across* to get 3. Playing *Down* to get 0 makes little sense since this is lower than the payoff she gave up at the first node.[7] But if Bob considers *Across* (sufficiently) more likely than *Down*, he will play *In*. Presumably then, Ann will indeed play *Across*, and the BI path results.

There is no contradiction with the previous analysis because in Figure 3.5, Ann is irrational once she doesn't play *Out*, so we can't say Ann should then rationally play *Across* not *Down*. To make *Across* rational for Ann, we have to add more types to the structure–specifically, we would want to

[7] This is forward-induction reasoning à la Kohlberg-Mertens [42, 1986, Section 2.3], introduced in the context of non-PI games. Interestingly, epistemic analysis makes clear that the issue already arises in PI games (such as Figure 3.4).

add a second type of Ann that assigns (initial) probability 1 to Bob's playing *In* not *Out*. This key insight is due to Stalnaker [58, 1998] and Battigalli-Siniscalchi [14, 2002].

Battigalli-Siniscalchi formulate a general result of this kind. They consider a **complete** CPS-based type structure, which contains, in a certain sense, every possible type of each player,[8] and prove: *Fix a complete CPS-based type structure. If there is RCSBR at the state* $(s^1, t^1, \ldots, s^n, t^n)$*, then the strategy profile* $(s^1, \ldots, s^n)$ *is extensive-form rationalizable. Conversely, if the profile* $(s^1, \ldots, s^n)$ *is extensive-form rationalizable, then there is a state* $(s^1, t^1, \ldots, s^n, t^n)$ *at which there is RCBR.*

The extensive-form rationalizability strategies (Pearce [48, 1984]) yield the BI outcome in a PI game,[9] so the Battigalli-Siniscalchi analysis gives epistemic conditions for BI.

There are other routes to getting BI in PI games. Asheim [1, 2001] develops an epistemic analysis using the properness concept (Myerson [46, 1978]). Go back to Example 4.1. The properness idea says that Bob's type t^b should view (*Across*, t^a) as infinitely more likely than (*Down*, t^a) since *Across* is the less costly 'mistake' for Ann, given her type t^a. Unlike the completeness route taken above, the irrationality of both *Down* and *Across* (given Ann's type t^a) is accepted. But the relative ranking of these 'mistakes' must be in the right order. With this ranking, Bob is irrational to play *Out* rather than *In*. Ann presumably will play *Across*, and we get BI again. Asheim [1, 2001] formulates a general such result.

Another strand of the literature on BI employs knowledge models rather than belief models. As pointed out in Example 1.1, players' beliefs don't have to be correct in any sense. For example, a type might even assign probability 1 to a strategy-type pair for another player different from the actual one. Knowledge as usually formalized is different, in that if a player knows an event E, then E indeed happens.

Aumann [5, 1995] formulates a knowledge-based epistemic model for PI trees. In his set-up, the condition of common knowledge of rationality implies that the players choose their BI strategies. Stalnaker [57, 1996] finds that non-BI outcomes are possible, under a different formulation of the same condition. The explanation lies in differences in how counterfactuals are treated. These play an important role in a knowledge-based analysis, when we talk about what a player thinks at an information set that cannot be reached given what he knows. Halpern [38, 2001] provides a synthesis in which these differences can be understood. See also the exchange between Binmore [20, 1996] and Aumann [6, 1996], and the analyses by Samet [52, 1996], Balkenborg and Winter [10, 1997], and Halpern [37, 1999].

Aumann [7, 1998] provides knowledge-based epistemic conditions under which Ann plays *Out* in Centipede. The conditions are weaker than in his [5, 1995] paper, and the conclusion weaker (about outcomes not strategies). There is an obvious parallel between this result and the belief-based result on Centipede we stated in Section 4 (also about outcomes). More generally, there may be an analogy between counterfactuals in knowledge models and extended probabilities in belief models.

[8] A complete type structure will be uncountably infinite.

[9] Under an assumption ruling out certain payoff ties (Battigalli [12, 1997]).

But, for one thing, completeness is crucial to the belief-based approach, as we have seen, and an analogous concept does not appear to be present in the knowledge-based approach. As yet, there does not appear to be any formal treatment of the relationship between the two approaches.

5 Next Steps: Weak Dominance

Extending the epistemic analysis of games from the matrix to the tree has been the focus of much recent work in the literature. Another area has been extending the analysis on the matrix from strong dominance (described in Section 2) to weak dominance.

Weak dominance (admissibility) says that a player considers as possible (even if unlikely) any of the strategies for the other players. In the game context, we are naturally led to consider iterated admissibility (IA)–the weak-dominance analog to IU. This is an old concept in game theory, going back at least to Gale [35, 1953]. Like BI, it is a powerful solution concept, delivering sharp answers in many games–Bertrand, auctions, voting games, and others.[10]

But also like BI, there is a conceptual puzzle. Suppose Ann conforms to the admissibility requirement, so that she considers possible any of Bob's strategies. Suppose Bob also conforms to the requirement, and this leads him not to play a strategy, say L. If Ann thinks Bob adheres to the requirement (as he does), then she can rule out Bob's playing L. But this conflicts with the requirement that she not rule anything out. (See Samuelson [55, 1992].)

Can a sound argument be made for IA? To investigate this, the epistemic tools of Section 1 have to be extended again.

Example 5.1 (Bertrand) *Figure 5.1 is a Bertrand pricing game, where each firm chooses a price in* $\{0, 1, 2, 3\}$.[11] *(The left payoff is to A, the right payoff to B. Each firm has capacity of two units and zero cost. Two units are demanded. If the firms charge the same price, they each sell one unit.) Figure 5.2 is an associated type structure (with one type for each player).*

		B			
		3	2	1	0
A	3	3, 3	0, 4	0, 2	0, 0
	2	4, 0	2, 2	0, 2	0, 0
	1	2, 0	2, 0	1, 1	0, 0
	0	0, 0	0, 0	0, 0	0, 0

Figure 5.1

[10] Mertens [44, 1989, p.582] and Marx-Swinkels [43, 1997, p.224-225] list various games involving weak dominance.
[11] Ken Corts kindly provided this example.

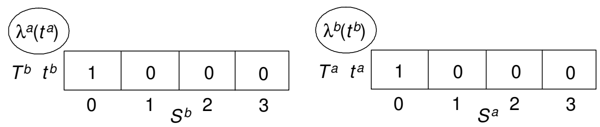

Figure 5.2

The rational strategy-type pairs are $R_1^a = \{0,1,2,3\} \times \{t^a\}$ *and* $R_1^b = \{0,1,2,3\} \times \{t^b\}$. *Since both types assign positive probability only to a rational strategy-type pair of the other player, we get* $R_m^a = R_1^a$ *and* $R_m^b = R_1^b$ *for all* m. *In particular, there is RCBR at the state* $(3, t^a, 3, t^b)$.

But a price of 3 is inadmissible (as is a price of 0). The IA set is just $\{(1,1)\}$, where each firm charges the lowest price above cost. (This is a plausible scenario: While pricing at cost is inadmissible, competition forces price down to the first price above cost.)

A tool to incorporate admissibility is *lexicographic probability systems* (*LPS's*), introduced and axiomatized by Blume, Brandenburger, and Dekel [21, 1991], [22, 1991]. An LPS specifies a sequence of probability measures. The interpretation is that the first measure is the player's primary hypothesis about the true state. But the player recognizes that his primary hypothesis might be mistaken, and so also forms a secondary hypothesis. This is his second measure. Then his tertiary hypothesis, and so on. The primary states can be thought of as infinitely more likely than the secondary states, which are infinitely more likely than the tertiary states, etc. Stahl [56, 1995], Stalnaker [58, 1998], Asheim [1, 2001], Brandenburger-Friedenberg-Keisler [29, 2006], and Asheim and Perea [2, 2005], among other papers, use LPS's.

Example 5.2 (Bertrand contd.) *Figure 5.3 is a type structure for Bertrand (Figure 5.1) that now specifies LPS's.*

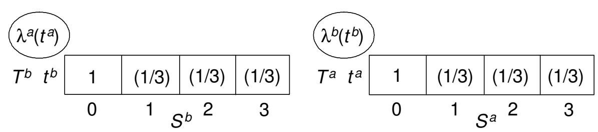

Figure 5.3

Each player has a primary hypothesis which assigns probability 1 to the other player's charging a price of 0. But each player also has a secondary hypothesis that assigns equal probability to each of the three remaining choices of the other player. This measure is shown in parentheses. Note that every state (i.e., strategy-type pair) gets positive probability under some measure. But states can also be ruled out, in the sense that they can be give infinitely less weight than other states.

What about epistemic conditions? Are the players rational in this situation? Does each think the other is rational? Etc.

To answer, we need a definition of rationality with LPS's. Fix strategy-type pairs (s^i, t^i) and (r^i, t^i) of player i, where t^i is now associated with an LPS. Calculate the tuple of expected payoffs to i from s^i, using first the primary measure associated with t^i, then the secondary measure associated with t^i, etc. Calculate the corresponding tuple for r^i. If the first tuple lexicographically exceeds the second, then s^i is preferred to r^i.[12] A strategy-type pair (s^i, t^i) is **rational (in the lexicographic sense)** if s^i is maximal under this ranking.

So $(3, t^a)$ and $(3, t^b)$ are irrational. All choices give each player an expected payoff of 0 under the primary measure. But a price of 2 gives each player an expected payoff of 2 under the secondary measure, as opposed to an expected payoff of 1 from a price of 3. Conceptually, we want $(3, t^a)$ and $(3, t^b)$ to be irrational (because a price of 3 is inadmissible).

What does each player think about the other's rationality? For this, we again need an LPS-based definition. An early candidate in the literature was: Say player i **believes** event E **at the 1st level** if E gets primary probability 1 under i's LPS (Börgers [25, 1994], Brandenburger [26, 1992]). A stronger concept is: Say i **assumes** E if all states not in E are infinitely less likely than all states in E, under i's LPS (Brandenburger-Friedenberg-Keisler [29, 2006]). In other words, a player who assumes E recognizes E may not happen, but is prepared to 'count on' E versus not-E.

In Figure 5.3, type t^a doesn't 1st-level believe (so certainly doesn't assume) the other player is rational. Likewise with t^b. Again, this is right conceptually.

6 Conditions for Iterated Admissibility

Once again we can parallel Definition 2.1 and define inductively **rationality and mth-order 1st-level belief of rationality (Rm1BR)** at a state of a type structure, and **rationality and common 1st-level belief of rationality (RC1BR)**. Likewise, one can define **rationality and mth-order assumption of rationality (RmAR)**, and **rationality and common assumption of rationality (RCAR)**. What do these conditions yield?

In fact, just as we saw in Sections 3 and 4 that neither RCIBR not RCSBR yields BI, so neither RC1BR nor RCAR yields IA. RC1BR is characterized by the $S^\infty W$ concept (Dekel and Fudenberg [31, 1990]), i.e., the set of strategies that remain after one round of deletion of inadmissible strategies followed by iterated deletion of strongly dominated strategies. RCAR is characterized by the self-admissible set concept (Brandenburger-Friedenberg-Keisler [29, 2006]). Self-admissible sets may be viewed as the weak-dominance analog to Pearce ([48, 1984]) best-response sets.

But while the IA set is one self-admissible set is a game, there may well be others. To select the IA set, a completeness assumption is needed, similar to Section 4: *Fix a complete LPS-based type structure. If there is RmAR at the state $(s^1, t^1, \ldots, s^n, t^n)$, then the strategy profile $(s^1, \ldots, s^n)$ survives $(m + 1)$ rounds of iterated admissibility. Conversely, if the profile $(s^1, \ldots, s^n)$ survives $(m + 1)$ rounds of iterated admissibility, then there is a state $(s^1, t^1, \ldots, s^n, t^n)$ at which there is RmAR* ([29, 2006]).

This result is stated for RmAR and not RCAR. See the next section for the reason. Of course, for a given game, there is an m such that IA stabilizes after m rounds.

[12] If $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, then x lexicographically exceeds y if $y_j > x_j$ implies $x_k > y_k$ for some $k < j$.

IA yields the BI outcome in a PI game,[13] so, understanding IA gives, in particular, another analysis of BI.

Related analyses of IA include Stahl [56, 1995] and Ewerhart [32, 2002]. Stahl uses LPS's and directly assumes that Ann considers one of Bob's strategies infinitely less likely than another if the first is eliminated on an earlier round of IA than the second. Ewerhart gives an analysis of IA couched in terms of provability (from mathematical logic).

7 Strategic vs. Extensive Analysis

Kohlberg-Mertens [42, 1986, Section 2.4] argued that a 'fully rational' analysis of games should be invariant–i.e., should depend only on the fully reduced strategic form of a game.[14] In this, they appealed to early results in game theory (Dalkey [30, 1953], Thompson [60, 1952]) which established that two trees sharing the same reduced strategic form differ from each other by a (finite) sequence of elementary transformations of the tree, each of which can be argued to be 'strategically inessential.' Kohlberg-Mertens added a fourth transformation involving convex combinations, to get to the fully reduced strategic form.

In decision theory, invariance is implied by (and implies) admissibility. (Kohlberg-Mertens [42, 1986, Section 2.7] gave the essential idea. See Brandenburger [27] for the decision-theory argument.) If we build up our game analysis using a decision theory that satisfies admissibility, we can hope to get invariance at this level too. LPS-based decision theory satisfies admissibility. Indeed, IA, and also the $S^{\infty}W$ and self-admissible set concepts, are invariant in the Kohlberg-Mertens sense. The extensive-form rationalizability concept (Section 4) is not.

There does appear to be a price paid for invariance, however. The extensive-form conditions of RCSBR and (CPS-based) completeness are consistent (in any tree). That is, for any tree, we can build a complete type structure and find a state at which RCSBR holds. But Brandenburger-Friedenberg-Keisler [29, 2006] shows the strategic-form conditions of RCAR and (LPS-based) completeness are inconsistent (in any matrix satisfying a non-triviality condition).

A possible interpretation is that rationality, even as a theoretical concept, appears to be inherently limited. There are purely theoretical limits to the Kohlberg-Mertens notion of a 'fully rational' analysis of games.

The epistemic program has uncovered a number of impossibility results (see EPISTEMIC GAME THEORY: BELIEFS AND TYPES for some others). We don't see this as a deficiency of the program, but rather as a sign it has reached a certain depth and maturity. Also, central to the program is the analysis of scenarios (we have seen several in this survey) that are 'a long way from' these theoretical limits. Under the epistemic approach to game theory there is not one right set of assumptions to make about a game.

[13] Again ruling out certain payoff ties (Marx-Swinkels [43, 1997]).

[14] The strategic form after elimination of any (pure) strategies that are duplicates or convex combinations of other strategies.

References

[1] Asheim, G., "Proper Rationalizability in Lexicographic Beliefs," *International Journal of Game Theory*, 30, 2001, 453-478.

[2] Asheim, G., and A. Perea, "Sequential and Quasi-Perfect Rationalizability in Extensive Games," *Games and Economic Behavior*, 53, 2005, 15-42.

[3] Aumann, R., "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 1974, 67-96.

[4] Aumann, R., "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1987, 1-18.

[5] Aumann, R., "Backward Induction and Common Knowledge of Rationality," *Games and Economic Behavior*, 8, 1995, 6-19.

[6] Aumann, R., "Reply to Binmore," *Games and Economic Behavior*, 17, 1996, 138-146.

[7] Aumann, R., "On the Centipede Game," *Games and Economic Behavior*, 23, 1998, 97-105.

[8] Aumann, R., "Common Priors: A Reply to Gul," *Econometrica*, 66, 1998, 929-938.

[9] Aumann, R., and A. Brandenburger, "Epistemic Conditions for Nash Equilibrium," *Econometrica*, 63, 1995, 1161-1180.

[10] Balkenborg, D., and E. Winter, "A Necessary and Sufficient Epistemic Condition for Playing Backward Induction," *Journal of Mathematical Economics*, 27, 1997, 325-345.

[11] Basu, K., "On the Existence of a Rationality Definition for Extensive Games," *International Journal of Game Theory*, 19, 1990, 33-44.

[12] Battigalli, P., "On Rationalizability in Extensive Games," *Journal of Economic Theory*, 74, 1997, 40-61.

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

[14] Battigalli, P., and M. Siniscalchi, "Strong Belief and Forward-Induction Reasoning," *Journal of Economic Theory*, 106, 2002, 356-391.

[15] Ben Porath, E., "Rationality, Nash Equilibrium, and Backward Induction in Perfect Information Games," *Review of Economic Studies*, 64, 1997, 23-46.

[16] Bernheim, D., "Rationalizable Strategic Behavior," *Econometrica*, 52, 1984, 1007-1028.

[17] Bicchieri, C., "Strategic Behavior and Counterfactuals," *Synthese*, 76, 1988, 135-169.

[18] Bicchieri, C., "Self-Refuting Theories of Strategic Interaction: A Paradox of Common Knowledge," *Erkenntnis*, 30, 1989, 69-85.

[19] Binmore, K., "Modelling Rational Players I," *Economics and Philosophy*, 3, 1987, 179-214.

[20] Binmore, K., "A Note on Backward Induction," *Games and Economic Behavior*, 17, 1996, 135-137.

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

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

[23] Bonanno, G., "The Logic of Rational Play in Games of Perfect Information," *Economics and Philosophy*, 7, 1991, 37-65.

[24] Bonanno, G., and K. Nehring, "How to Make Sense of the Common Prior Assumption under Incomplete Information," *International Journal of Game Theory*, 28, 1999, 409-434.

[25] Börgers, T., "Weak Dominance and Approximate Common Knowledge," *Journal of Economic Theory*, 64, 1994, 265-276.

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

[27] Brandenburger, A., "The Power of Paradox: Some Recent Results in Interactive Epistemology," *International Journal of Game Theory*, 35, 2007, 465-492.

[28] Brandenburger, A., and E. Dekel, "Rationalizability and Correlated Equilibria," *Econometrica*, 55, 1987, 1391-1402.

[29] Brandenburger, A., A. Friedenberg, and H.J. Keisler, "Admissibility in Games," 2006. Available at www.stern.nyu.edu/∼abranden.

[30] Dalkey, N., "Equivalence of Information Patterns and Essentially Determinate Games," in H. Kuhn and A. Tucker (eds.), *Contributions to the Theory of Games*, Vol. 2, Princeton University Press, 1953, 217-244.

[31] Dekel, E., and D. Fudenberg, "Rational Behavior with Payoff Uncertainty," *Journal of Economic Theory*, 52, 1990, 243-267.

[32] Ewerhart, C., "Ex-ante Justifiable Behavior, Common Knowledge, and Iterated Admissibility," 2002. Available at http://mail.wiwi.uni-bonn.de/users/cewerhart.

[33] Feinberg, Y., "Characterizing Common Priors in Terms of Posteriors," *Journal of Economic Theory*, 91, 2000, 127-179.

[34] Friedenberg, A., "When Common Belief is Correct Belief," 2002. Available at www.olin.wustl.edu/faculty/friedenberg.

[35] Gale, D., "A Theory of n-Person Games with Perfect Information," *Proceedings of the National Academy of Sciences*, 39, 1953, 496–501.

[36] Gul, F., "A Comment on Aumann's Bayesian View," Econometrica, 66, 1998, 923-927.

[37] Halpern, J., "Hypothetical Knowledge and Counterfactual Reasoning," *International Journal of Game Theory*, 28, 1999, 315-330.

[38] Halpern, J., "Substantive Rationality and Backward Induction," *Games and Economic Behavior*, 37, 2001, 425-435.

[39] Halpern, J., "Characterizing the Common Prior Assumption," *Journal of Economic Theory*, 106, 2002, 316-355.

[40] Harsanyi, J., "Games with Incomplete Information Played by 'Bayesian' Players, I-III," *Management Science*, 14, 1967-68, 159-182, 320-334, 486-502.

[41] Harsanyi, J., "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed Strategy Equilibrium Points," *International Journal of Game Theory*, 2, 1973, 1-23.

[42] Kohlberg, E., and J.-F. Mertens, "On the Strategic Stability of Equilibria," *Econometrica*, 54, 1986, 1003-1037.

[43] Marx, L., and J. Swinkels, "Order Independence for Iterated Weak Dominance," *Games and Economic Behavior*, 18, 1997, 219-245.

[44] Mertens, J-F., "Stable Equilibria–A Reformulation," *Mathematics of Operations Research*, 14, 1989, 575-625.

[45] Morris, S., "Trade with Heterogeneous Prior Beliefs and Asymmetric Information," *Econometrica*, 62, 1994, 1327-1347.

[46] Myerson, R., "Refinements of the Nash Equilibrium Concept," , *International Journal of Game Theory*, 1, 1978, 73-80.

[47] Myerson, R., *Game Theory*, Harvard University Press, 1991.

[48] Pearce, D., "Rational Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1984, 1029-1050.

[49] Reny, P., "Rationality in Extensive Form Games," *Journal of Economic Perspectives*, 6, 1992, 103-118.

[50] Rényi, A., "On a New Axiomatic Theory of Probability," *Acta Mathematica Academiae Scientiarum Hungaricae*, 6, 1955, 285-335.

[51] Rosenthal, R., "Games of Perfect Information, Predatory Pricing and the Chain-Store Paradox," *Journal of Economic Theory*, 25, 1981, 92-100.

[52] Samet, D., "Hypothetical Knowledge and Games with Perfect Information," *Games and Economic Behavior*, 17, 1996, 230-251.

[53] Samet, D., "Common Priors and the Separation of Convex Sets," *Games and Economic Behavior*, 24, 1998, 172-174.

[54] Samet, D., "Iterated Expectations and Common Priors," *Games and Economic Behavior*, 24, 1998, 131-141.

[55] Samuelson, L., "Dominated Strategies and Common Knowledge," *Games and Economic Behavior*, 4, 1992, 284-313.

[56] Stahl, D., "Lexicographic Rationalizability and Iterated Admissibility," *Economic Letters*, 47, 1995, 155-159.

[57] Stalnaker, R., "Knowledge, Belief and Counterfactual Reasoning in Games," *Economics and Philosophy*, 12, 1996, 133-163.

[58] Stalnaker, R., "Belief Revision in Games: Forward and Backward Induction," *Mathematical Social Sciences*, 36, 1998, 31-56.

[59] Tan, T., and S. Werlang, "The Bayesian Foundations of Solution Concepts of Games," *Journal of Economic Theory*, 45, 1988, 370-391.

[60] Thompson, F., "Equivalence of Games in Extensive Form," Research Memorandum RM-759, The RAND Corporation, 1952.