Foundations of Game Theory

Adam Brandenburger

Version 12/31/07

Sources:

"The Power of Paradox: Some Recent Developments in Interactive Epistemology," *International Journal of Game Theory*, 35, 2007, 465-492

"Epistemic Game Theory: An Overview" and "Epistemic Game Theory: Complete Information," to appear in *The New Palgrave Dictionary of Economics*, Palgrave Macmillan, 2008

Available at www.stern.nyu.edu/~abranden

Introduction

The starting point for most of game theory is a “solution concept”—such as Nash equilibrium or one of its many variants, backward induction, or iterated dominance of various kinds

These are usually thought of as the embodiment of “rational behavior” in some way and used to analyze game situations

The starting point for most game theory is more of an endpoint of work in foundations

Here, the primitives are more basic:

- The very idea of rational—or irrational—behavior needs to be formalized
- So does what each player might know or believe about the game—including about the rationality or irrationality of other players
- Foundational work shows that even what each player knows or believes about what other players know or believe, and so on, can matter

Historical Context

The foundations ("epistemic") program can be seen as a response to the equilibrium refinements program of the 1980s

In the refinements program, the starting point was Nash equilibrium, and various modifications of equilibrium were proposed and interpreted as reflecting one or another underlying notion of rationality (plus belief in rationality, etc.)

"In this way, we may eventually reach an axiomatisation, and an interpretation in terms of rationality, without imposing any explicit preconception about what rationality exactly means, except for some general a priori requirement[s]"
--Mertens (1989, p.583)

The epistemic program is different!

The Formalism

We augment the traditional description of a game 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

Harsanyi (1967-8) introduced the types concept 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

The technique is equally useful to talk about uncertainty about the 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

It is also possible to treat both kinds of uncertainty together, using the same technique

Type Structures

Fix an *n*-player finite strategic-form game

$$<S^1, \ldots, S^n, \pi^1, \ldots, \pi^n>$$

An (S^1, …, S^n)**-based** (**finite**) **type structure** is a structure

$$<S^1, \ldots, S^n; T^1, \ldots, T^n; \lambda^1, \ldots, \lambda^n>$$

where each T^i is a finite set, and each $\lambda^i : T^i \rightarrow M(S^{-i} \times T^{-i})$

Members of T^i are called **types** for player *i*

Members of $S \times T$ are called **states** (**of the world**)

From Types to Hierarchies

A 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 $j \neq i$, etc.

Also, what about going from hierarchies to types?

See:

Armbruster and Böge (1979)

Böge and Eisele (1979)

Mertens and Zamir (1985)

Brandenburger and Dekel (1993)

Heifetz (1993)

and others

Example of a Type Structure

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

At the state (D, t^a, R, t^b):

i. Ann and Bob are each ‘correct’ about the other’s strategy

ii. Ann thinks it possible Bob is wrong about her strategy

iii. Ann and Bob are each rational

iv. Ann thinks it possible Bob is irrational

Similar to an example in Aumann and Brandenburger (1995)

Rationality and Common Belief of Rationality

A strategy-type pair (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 for player i **believes** an event $E \subseteq S^{-i} \times T^{-i}$ if $\lambda^i(t^i)(E) = 1$

Write

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

For each player i, let R^i_1 be the set of all rational pairs (s^i, t^i) and for $m > 0$ define R^i_m inductively by

$$R^i_{m+1} = R^i_m \cap [S^i \times B^i(R^{-i}_m)]$$

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 $(s^1, t^1, \ldots, s^n, t^n) \in \cap_{m=1}^{\infty} R_m$, say there is **rationality and common belief of rationality** (**RCBR**) at this state

Early Results: Characterization of RCBR

Fix a type structure and a state (s^1*,* t^1*, …,* s^n*,* t^n) *at which there is RCBR. Then the strategy profile* (s^1*, …,* s^n) *is* ***iteratively undominated****. Conversely, fix an iteratively undominated profile* (s^1*, …,* s^n)*. There is a type structure and a state* (s^1*,* t^1*, …,* s^n*,* t^n) *at which there is RCBR.*

The key: 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

See:

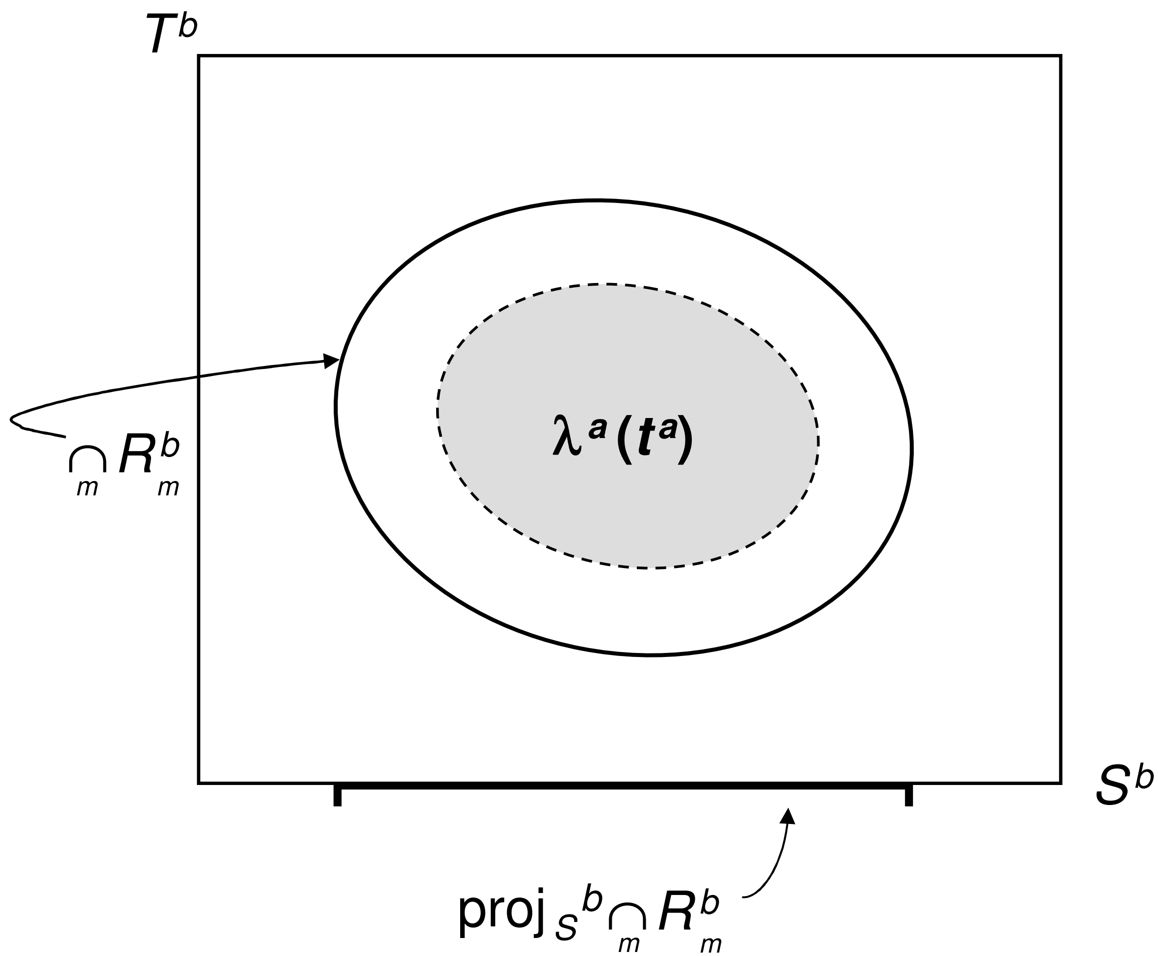
Brandenburger and Dekel (1987)

Tan and Werlang (1988)

and others

Proof of Forward Direction

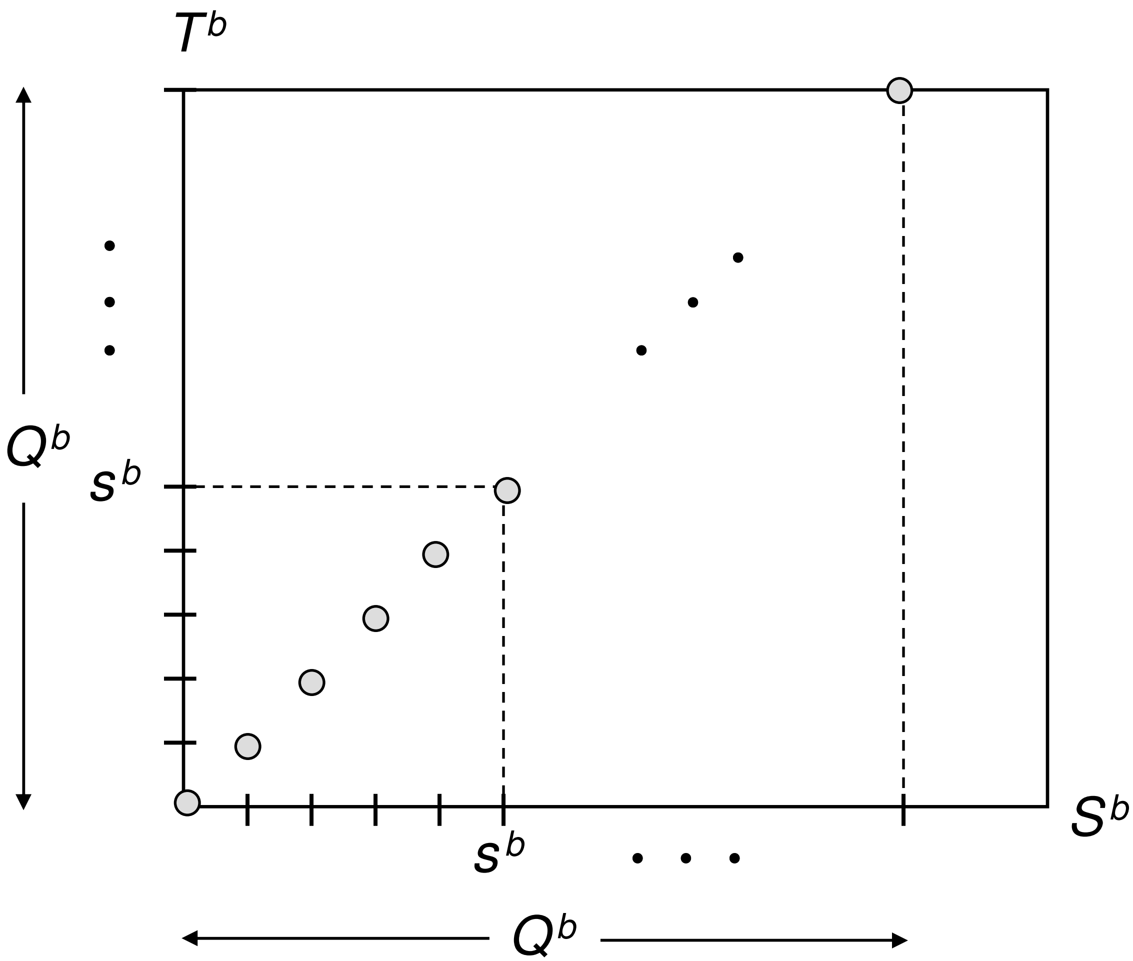
Consider $(s^a, t^a) \in \bigcap_m R^a_m$



We show that $\text{proj}_{S^a} \bigcap_m R^a_m \times \text{proj}_{S^b} \bigcap_m R^b_m$ is a best-response set (Pearce 1984)

(Note the use of conjunction!)

Proof of Converse Direction



Let $Q^a \times Q^b$ be a best-response set

For a given $t^a = s^a$, put the weights along the diagonal

Early Results contd.: Rationalizability

Call a strategy "good" if there is a product measure on the product of the other players' strategy sets under which it is optimal

The **rationalizable** strategies are those that survive iterated elimination of bad strategies (Bernheim 1984 and Pearce 1984)

But is the independence assumption implied by the assumption of non-cooperative play? Or, is some form of conditional independence more appropriate?

See Aumann (1974, 1987) for "extrinsic correlation"—i.e., conditioning on external signals

See Brandenburger and Friedenberg (2004) for "intrinsic correlation"—i.e., conditioning on the hierarchies of beliefs

Early Results contd.: Nash and Correlated Equilibrium

Nash equilibrium:

Pure equilibrium is characterized by the condition that each player is rational and assigns probability 1 to the actual strategies chosen by the other players

From an epistemic perspective, Nash equilibrium is a special case

Mixed equilibrium:

Under the epistemic approach, each player makes a definite choice of (pure) strategy, and it is the other players who are uncertain about this choice (cf. Harsanyi 1973)

Aumann and Brandenburger (1995) give an epistemic characterization of mixed equilibrium along these lines

Correlated equilibrium:

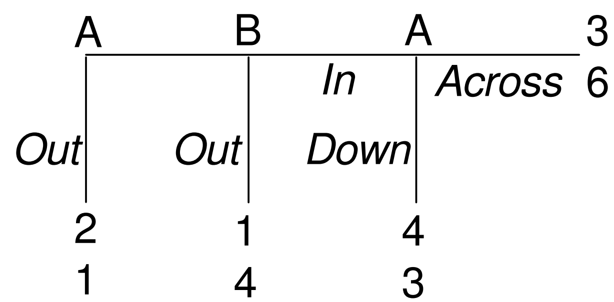
See Aumann (1987)

Next Steps:
The Tree

How to do epistemic analysis on the tree?

A big motivation is to understand the logical foundation of backward induction

Centipede (Rosenthal 1981):



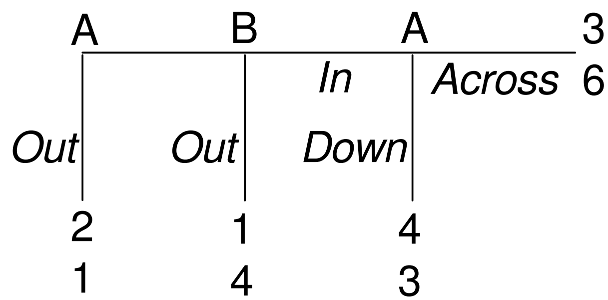
It seems that if Ann is rational, Bob thinks Ann is rational , …, then Ann plays *Out*

But what if she doesn't?

Perhaps she shouldn't!

See Binmore (1987), Bicchieri (1988, 1989), Basu (1990), Bonanno (1991), Reny (1992), and others

Type Structures for Trees



$\lambda^a(t^a)$

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]

Conditional Probability Systems

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

So, 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

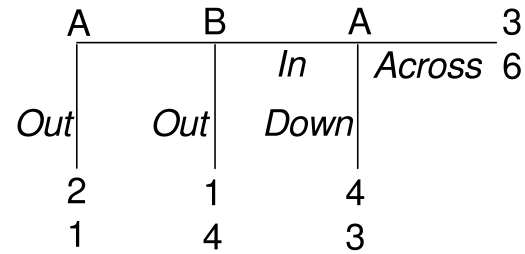
See:

Rényi (1955)

Myerson (1991)

Battigalli and Siniscalchi (1999, 2002)

Initial Belief



$\lambda^a(t^a)$

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]

$\lambda^b(t^b)$

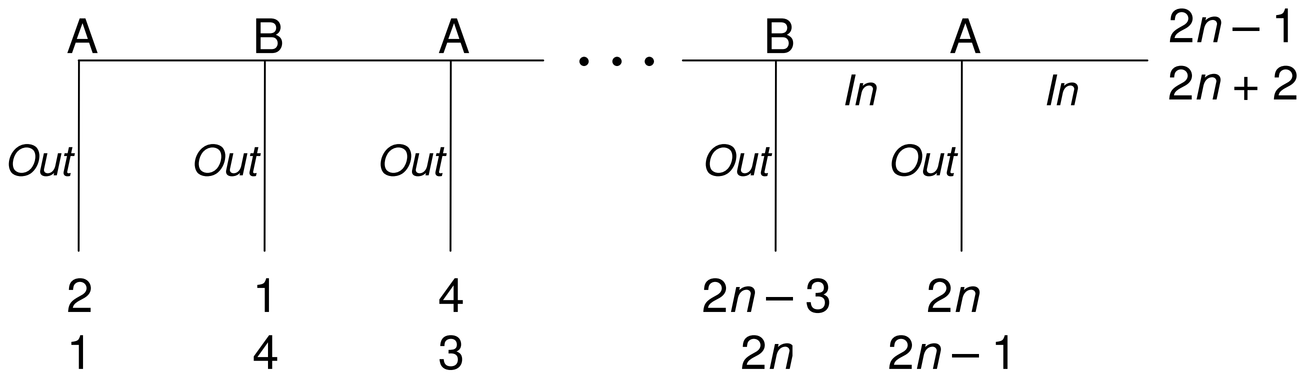
T^a \ S^a	*Out*	*Down*	*Across*
u^a	1 [0]	0 [0]	0 [0]
t^a	0 [0]	0 [0]	0 [1]

$\lambda^b(u^b)$

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

At the state (*Down*, t^a, *In*, t^b), there is **rationality** (**in the tree**) **and common initial belief of rationality** (Ben Porath 1997)

Strong Belief

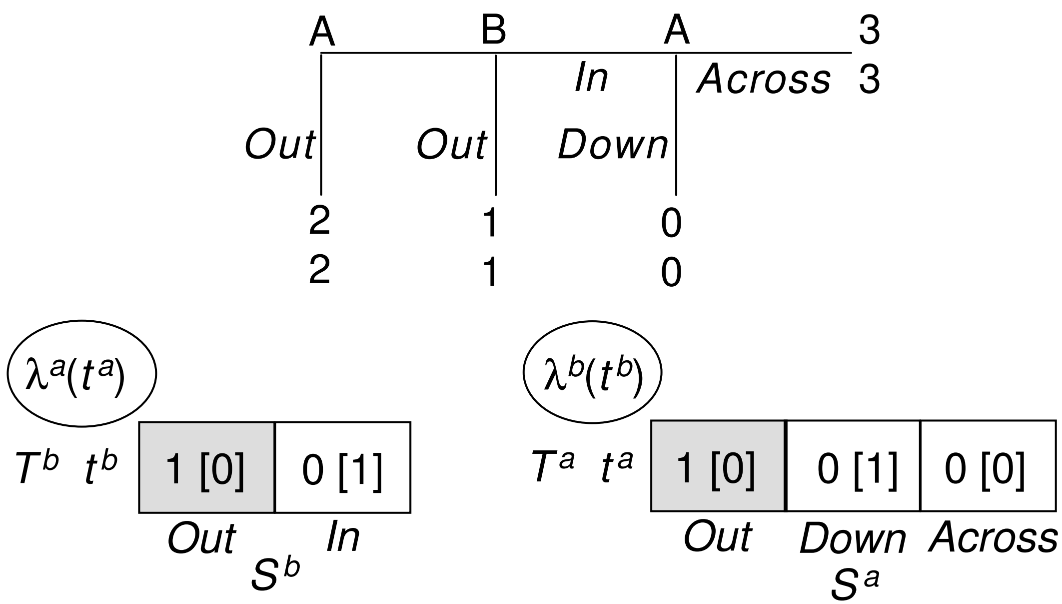


*Fix a CPS-based type structure for n-legged Centipede and a state at which there is **rationality** (**in the tree**) **and common strong belief of rationality** (Battigalli and Siniscalchi 2002)*

Then Ann plays Out

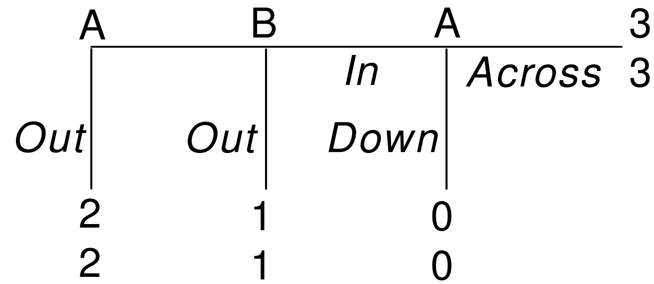
See Friedenberg (2002)

Backward Induction?



At the state (*Out*, t^a, *Out*, t^b), there is **rationality** (**in the tree**) **and common strong belief of rationality**

Adding Types



$\lambda^a(t^a)$

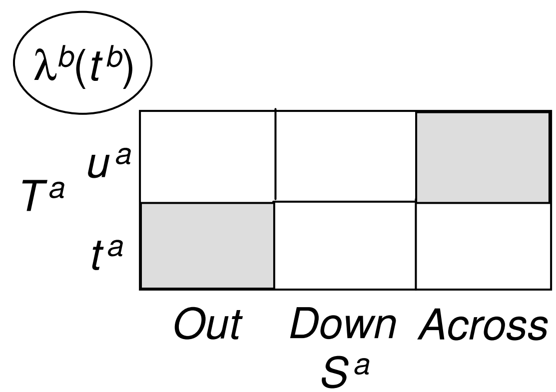
T^b	Out	In
t^b	1 [0]	0 [1]

S^b

$\lambda^a(u^a)$

T^b	Out	In
t^b	0	1

S^b



Now, type t^a for Ann doesn't (strongly) believe Bob is rational

Conditions for Backward Induction

Fix a complete CPS-based type structure. If there is rationality and common strong belief of rationality 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.*

See Battigalli and Siniscalchi (2002)

Extensive-form rationalizability (Pearce 1984, Battigalli 1997) is an iterated-dominance concept on the tree (despite the name, it does not make an independence assumption)

It yields the backward-induction outcome in perfect-information trees under a no-ties condition (Battigalli 1997)

* A **complete** type structure is two-way surjective (Brandenburger 2003)

Is Completeness Possible?

Consider the configuration of beliefs:

> Ann believes that Bob believes that Ann believes that what Bob believes (about Ann's type) is wrong

Ask:

> Does Ann believe that what Bob believes is wrong?
>
> Does Ann not believe that what Bob believes is wrong?

This shows that the above configuration is impossible

Can it be expressed in the (formal) language the players use?

Whether or not completeness is possible depends on the language used by the players

See Brandenburger and Keisler (2006)

Other Routes to Backward Induction

Asheim (2001) is an epistemic analysis using the properness concept (Myerson 1978)

Aumann (1995) formulates a knowledge-based epistemic model for PI trees, in which common knowledge of rationality implies that the players choose their backward-induction strategies

Stalnaker (1996) says that common knowledge of rationality does not imply backward induction

See Halpern (2001) for a synthesis (the two papers treat counterfactuals differently)

What is the relationship between the belief-based and knowledge-based approaches?

Belief vs. Knowledge

Philosophically, the belief-based approach takes the view that

only observables are knowable

unobservables are subject to belief, not knowledge

in particular, other players' strategies are unobservables, and only moves are observables

Next Steps:
Weak Dominance

What is the epistemic analysis of admissibility (weak dominance)—including iterated admissibility?

Note:

> A strategy is admissible if and only if there is a full-support probability measure on the product of the other players' strategy sets under which it is optimal

A puzzle (Samuelson 1992):

> Suppose Ann conforms to the admissibility requirement, so that, presumably, she should put positive weight on all of Bob's strategies. Suppose Bob also conforms to the requirement, and this leads him not to play *L*, say (which is inadmissible). If Ann thinks Bob adheres to the requirement (as he does), should she put zero weight on *L*?

Invariance

Kohlberg-Mertens (1986) argued that a “fully rational” analysis of games should be invariant to “strategically inessential” transformations of the tree (Dalkey 1953, Thompson 1952)

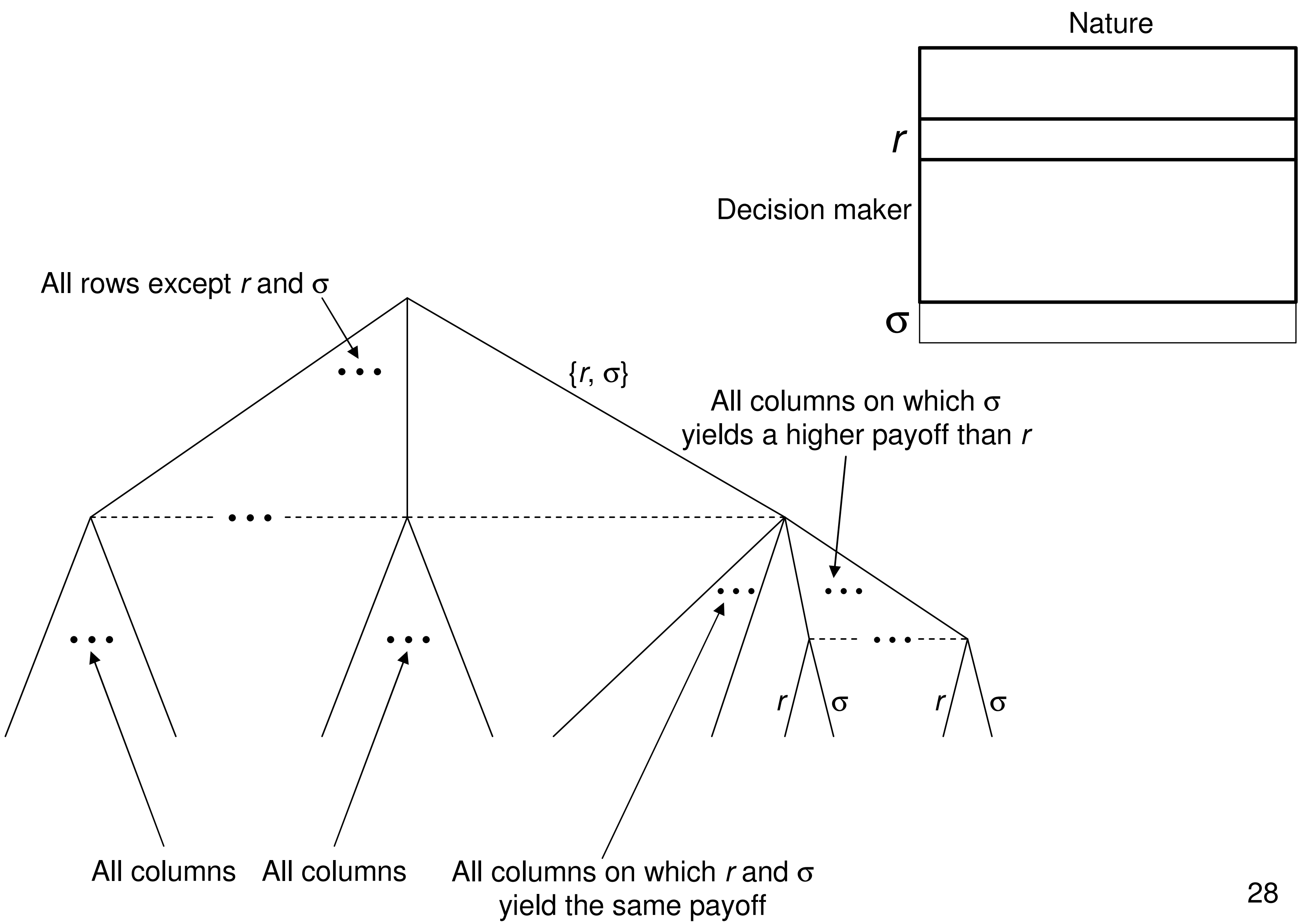
In decision theory, admissibility ensures invariance—in fact, is equivalent to it: A strategy in a decision matrix is admissible if and only if it is rational in every decision tree that reduces to this matrix (Kohlberg-Mertens 1986)

If we build up our game analysis using a decision theory which satisfies admissibility, we can hope to get invariance at this level too

(But note that an admissible solution concept need not be invariant! An example is (strategic-form) perfect equilibrium*)

* Amanda Friedenberg kindly pointed this out

Invariance Implies Admissibility



Type Structures for Admissibility

A \ B	3	2	1	0
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

$\lambda^a(t^a)$

T^b	S^b: 0	1	2	3
t^b	(1/3)	(1/3)	1	(1/3)

$\lambda^b(t^b)$

T^a	S^a: 0	1	2	3
t^a	(1/3)	(1/3)	1	(1/3)

Lexicographic Probability Systems

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, etc.

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.

See Blume, Brandenburger, and Dekel (1991)

Inclusion vs. Exclusion

LPS's solve the inclusion-exclusion problem associated with admissibility in games

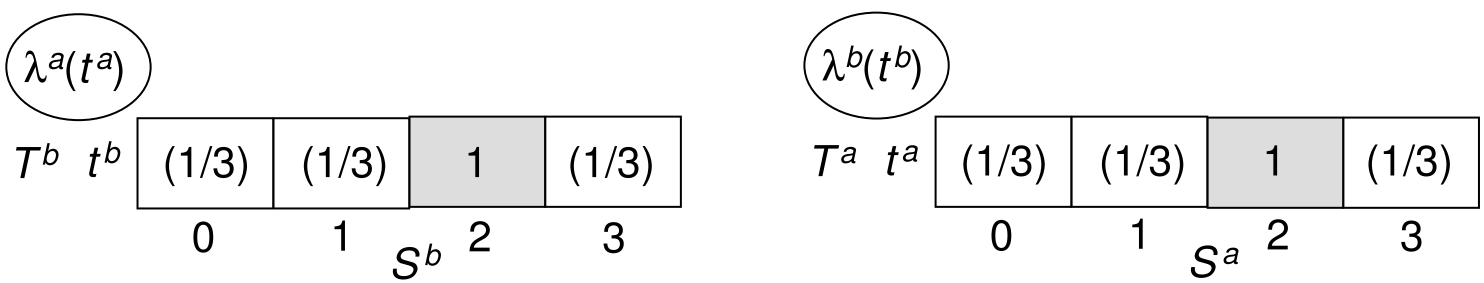
Strategies that get infinitesimal weight can be viewed as

included (because they don't get zero weight)

excluded (because they get only infinitesimal weight)

Assumption

A \ B	3	2	1	0
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



At the state (2**,** t^a, 2, t^b), there is (**lexicographic**) **rationality and common assumption of rationality** (Brandenburger, Friedenberg, and Keisler 2006)

Characterizations

Fix an LPS-based type structure and a state $(s^1, t^1, \ldots, s^n, t^n)$ *at which there is (lexicographic) rationality and common assumption of rationality. Then the strategy profile* $(s^1, \ldots, s^n)$ *belongs to a **self-admissible set**. Conversely, fix a profile* $(s^1, \ldots, s^n)$ *that belongs to a self-admissible set. There is an LPS-based type structure and a state* $(s^1, t^1, \ldots, s^n, t^n)$ *at which there is (lexicographic) rationality and common assumption of rationality.*

*Fix a **complete** LPS-based type structure. If there is (lexicographic) rationality and mth-order assumption of rationality 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 lexicographic) rationality and mth-order assumption of rationality.*

See Brandenburger, Friedenberg, and Keisler (2006)

Iterated admissibility yields the backward-induction outcome in perfect-information trees under a no-ties condition (Battigalli 1997)

Relationship between CPS's and LPS's

CPS's are suited to extensive-form analysis

LPS's are suited to strategic-form analysis

Nevertheless, one can establish some formal relationships

See:

Blume, Brandenburger, and Dekel (1989)

Hammond (1994)

Halpern (2003)

Battigalli (2004)

Asheim and Søvik (2005)

Brandenburger, Friedenberg, and Keisler (2006b)

An Impossibility Result

Admissibility

Asks a player to take all states into consideration

Rationality and common assumption of rationality

Asks a player to assume rationality and mth-order assumption of rationality for all m

Completeness

Asks a player to consider all possible types implied by the model

All three together are not possible (under some technical conditions)

See Brandenburger, Friedenberg, and Keisler (2006)

There appears to be an inherent limit on the ability of players to reason about all possibilities in a game

It is as if every time we think we finally get a hold on what rational behaviour means, we find ourselves having grasped only a shadow. Maybe this means there is excessive ύβρις *in this endeavour: that rationality is something belonging to the gods themselves, and that should not be stolen from them. Maybe it is the tree of knowledge itself, that we should not touch?*

--Mertens (1989, p.583)

Conclusion

The equilibrium refinements program is “top down”

Epistemic game theory is “bottom up”

Nash equilibrium plays a much smaller role in epistemic game theory

Under the epistemic approach, there is no one right set of conditions to impose on a game

(In particular, the inconsistency of certain criteria is not fatal)

The goal is to be able to analyze many different sets of conditions about games in a precise and uniform manner

References

All references are in the source papers except for

Asheim, G., and Y. Søvik, “Preference-Based Belief Operators,” *Mathematical Social Sciences*, 50, 2005, 61-82

Battigalli, P., “Interactive Epistemology in the Extensive and Normal Form,” unpublished, 2004

Blume, L., A. Brandenburger, and E. Dekel, “An Overview of Lexicographic Choice Under Uncertainty,” *Annals of Operations Research*, 19, 1989, 231-246

Brandenburger, A., ““On the Existence of a ‘Complete’ Possibility Structure,” in M. Basili, N. Dimitri, and I. Gilboa (eds.), *Cognitive Processes and Economic Behavior*, Routledge, 2003, 30-34

Brandenburger, A., and A., Friedenberg, “Intrinsic Correlation in Games,” 2004, forthcoming in *Journal of Economic Theory*; at www.stern.nyu.edu/~abranden

Brandenburger, A., A., Friedenberg, and H.J. Keisler, “Notes on the Relationship Between Strong Belief and Assumption,” 2006b, at www.stern.nyu.edu/~abranden

References contd.

Halpern, J., "Lexicographic Probability, Conditional Probability, and Nonstandard Probability," 2003, at www.cs.cornell.edu/home/halpern

Hammond, P., "Elementary Non-Archimedean Representations of Probability for Decision Theory and Games," in P. Humphreys (ed.), *Patrick Suppes: Scientific Philosopher, Vol. I: Probability and Probabilistic Causality*, Kluwer, 1994, 25-59