Foundations of Game Theory

Adam Brandenburger

08/28/10

Introduction

The starting point for most of non-cooperative game theory is a solution concept such as

– Nash equilibrium or variants
– backward induction
– iterated dominance in various forms . . .

These are usually thought of as the embodiment of "rational behavior" in some way

This starting point is more of an endpoint in foundations

In foundations, the primitives are more basic and we need to formalize the meaning of

– **rationality** and **irrationality**
– the players' **knowledge** or **beliefs** about the game (including about other players' knowledge or beliefs, . . .)

and investigate the implications of assumptions involving these concepts

From Equilibrium to Epistemics

The **epistemic** approach to foundations 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 (together with 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].*

–J.-F. Mertens*

The epistemic program is different!

* "Stable Equilibria–A Reformulation," *Mathematics of Operations Research*, 14, 1989, 575-625

Extending 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 notions

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 payoff functions, their beliefs about other players' beliefs about the payoff functions, . . .

The technique is equally useful to talk about the players' beliefs about the strategies chosen, their beliefs about other players' beliefs about the strategies chosen, . . .

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

$$\langle S^1, \ldots, S^n; \pi^1, \ldots, \pi^n \rangle$$

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)**

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, a hierarchy of beliefs–about the strategies chosen by the players $j \neq i$, about the beliefs held by the players $j \neq i$, etc.

What about going from hierarchies of beliefs to types?

Specifically, given a hierarchy of beliefs, is there a type structure and a type in that structure that induces this hierarchy?

Is there one type structure that suffices?

We won't address these important "foundations of foundations" questions here

Example of a Type Structure

Bob

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

Type t^a:

	L	R
t^b	0	1/2
v^b	1/2	0

Type t^b:

	U	D
t^a	1/4	0
v^a	0	3/4

Type v^a:

	L	R
t^b	0	0
v^b	1	0

Type v^b:

	U	D
t^a	0	0
v^a	0	1

At the true state–(U, t^a, R, t^b), say–we can calculate the players' hierarchies of beliefs over:

(i) strategies

(ii) rationality and irrationality

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_1^i be the set of all rational pairs (s^i, t^i) and for $m > 0$ define R_m^i inductively by

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

If $(s^1, t^1, \ldots, s^n, t^n) \in R_{m+1}$, say there is **rationality and mth-order belief of rationality (RmBR)** 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

Zeroth Theorem of Epistemic Game Theory

Theorem

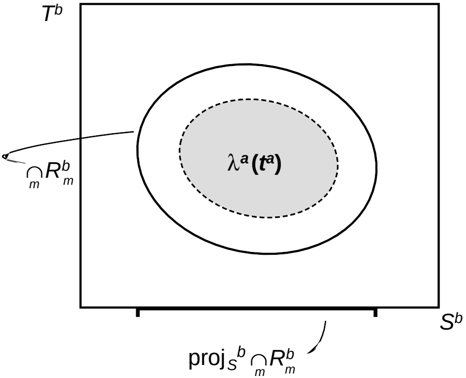
Fix a type structure and a state $(s^1, t^1, \ldots, s^n, t^n)$ *at which there is RCBR. Then the strategy profile* $(s^1, \ldots, s^n)$ *is* **iteratively undominated**. *Conversely, fix an iteratively undominated profile* $(s^1, \ldots, s^n)$. *There is a type structure and a state* $(s^1, t^1, \ldots,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

Early results along these lines are in Brandenburger and Dekel (1987) and Tan and Werlang (1988)

Modern improvements can be found in Battigalli and Siniscalchi (2002) and Battigalli and Friedenberg (2009)

Zeroth Theorem: Proof of the Forward Direction

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

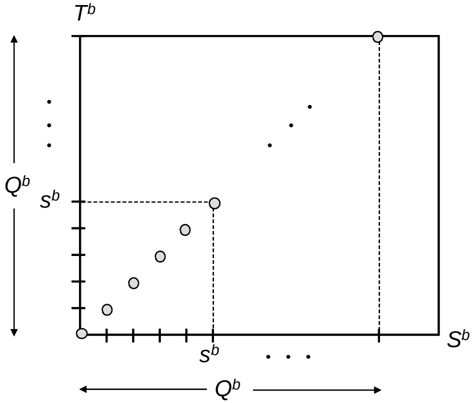


$\text{proj}_{S^a} \bigcap_m R^a_m \times \text{proj}_{S^b} \bigcap_m R^b_m$ is a best-response set (Pearce 1984)

Note the use of infinite conjunction

Zeroth Theorem: Proof of the Converse Direction

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



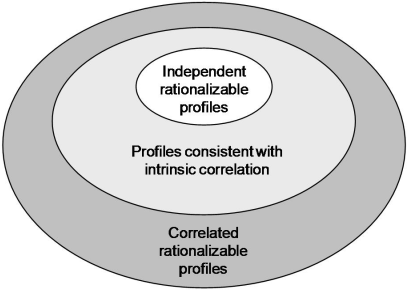
For a given $t^a = s^a$, put the weights along the diagonal

Rationalizability and Beyond

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 really implied by the assumption of non-cooperative play?

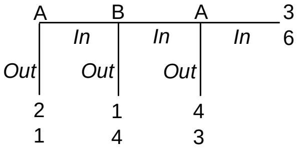


See "Intrinsic Correlation in Games," by A. Brandenburger and A. Friedenberg, *Journal of Economic Theory*, 141, 2008, 28-67

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)

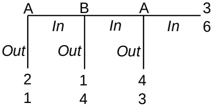
It seems that if Ann is rational, Bob is rational and thinks Ann is rational , . . . , then Ann plays *Out*

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



Type t^a:

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

S^b

Type u^a:

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

S^b

Type t^b:

T^a	*Out*	*In-Out*	*In-In*
u^a	1 [0]	0 [0]	0 [0]
t^a	0 [0]	0 [0]	0 [1]

S^a

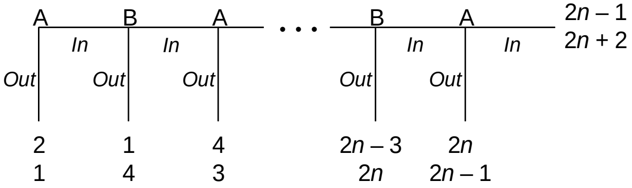
Type u^b:

T^a	*Out*	*In-Out*	*In-In*
u^a	1 [0]	0 [0]	0 [0]
t^a	0 [0]	0 [1]	0 [0]

S^a

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

Strong Belief

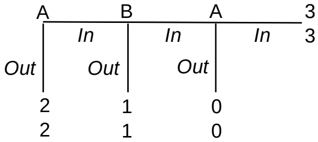


Theorem

Fix a type structure (with conditional probability systems) for n-legged Centipede and a state at which there is rationality (in the tree) and common **strong belief** *of rationality. Then, Ann plays Out.*

On strong belief, see Battigalli and Siniscalchi (2002)
For the result, see Battigalli and Friedenberg (2009)

Backward Induction?



Type t^a:

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

S^b

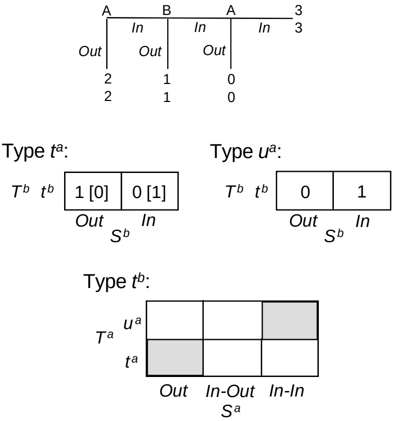
Type t^b:

T^a		Out	In-Out	In-In
	t^a	1 [0]	0 [1]	0 [0]

S^a

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

Adding Types



Now, type t^a for Ann does not (strongly) believe Bob is rational

Conditions for Backward Induction

Theorem

Fix a complete type structure (with conditional probability systems). 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 RCSBR.*

See Battigalli and Siniscalchi (2002)

Extensive-form rationalizability (Pearce 1984) 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)

'Large' and 'Small' Type Spaces

Type structures that, in one or another sense, contain all possible beliefs:

(i) terminal structures (Böge and Eisele 1979)
(ii) canonically-built (aka universal) structures (Mertens and Zamir 1985)
(iii) complete structures (Brandenburger 2003)

Some things–but not everything!–are known about the relationships among these three concepts

Small type structures are also important!

> *We think of a particular . . . structure as giving the "context" in which the game is played. In line with Savage's Small-Worlds idea in decision theory (Foundations of Statistics, 1954), who the players are in the given game can be seen as a shorthand for their experiences before the game. The players' possible characteristics–including their possible types–then reflect the prior history or context. Each different type structure reflects a different context for the game.**

* "Admissibility in Games," by A. Brandenburger, A. Friedenberg, and HJ. Keisler, *Econometrica*, 76, 2008, 307-352

The Knowledge-Based Approach

Aumann (1995) formulates a knowledge-based epistemic model for perfect-information 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 Artemov (2010) for the resolution

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

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** (avoidance of weakly dominated strategies) ensures invariance–indeed:

Fact

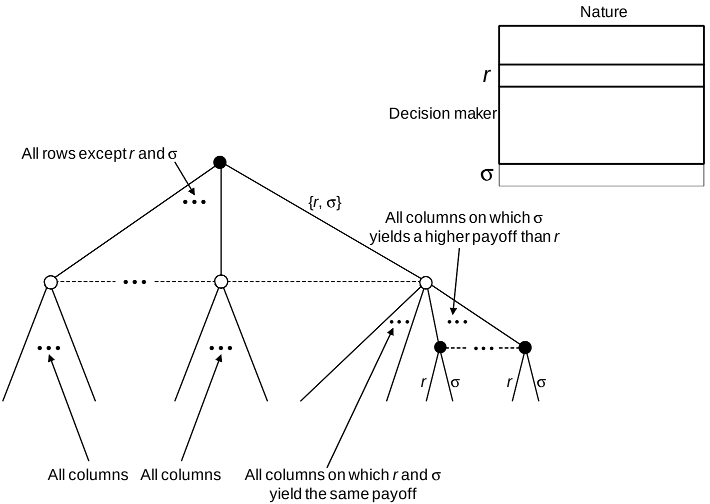
A strategy in a decision matrix is admissible if and only if it is rational in every decision tree that reduces to this matrix.

So, if we build up our game theory using a decision theory that satisfies admissibility, we can hope to get invariance at this level too

Note that an admissible solution concept need not be invariant–an example is (strategic-form) perfect equilibrium*

* Amanda Friedenberg pointed this out

Invariance Implies Admissibility



Epistemics of Admissibility

Fact

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*?*

Conditions for Iterated Admissibility

Theorem

Fix a complete type structure (with lexicographic probability systems). 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 (2008), which also has results for incomplete type structures

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

An Impossibility Result

Theorem

Fix a complete continuous type structure (with lexicographic probability systems). Then, under a non-triviality condition, there is no state at which there is (lexicographic) rationality and common assumption of rationality.

See Brandenburger, Friedenberg, and Keisler (2008)

An interpretation:

(i) admissibility asks a player to take all states into consideration

(ii) rationality and common assumption of rationality asks a player to reason to all levels

(iii) completeness asks a player to consider all possible types

In some sense, this is too much rationality to ask for!

Keisler (2009) and Lee (2009) provide positive results by omitting continuity–interpretation?

A Little Poetry

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 $\upsilon\beta\rho\iota\varsigma$ *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?*

–J.-F. 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