Epistemic Game Theory

Adam Brandenburger

June 22, 2008

John von Neumann and Oskar Morgenstern: *Theory of Games and Economic Behavior*, 1944

“[W]e wish to find the mathematically complete principles which define ‘rational behavior’ for the participants” (op.cit., p.31)

The search has continued ever since

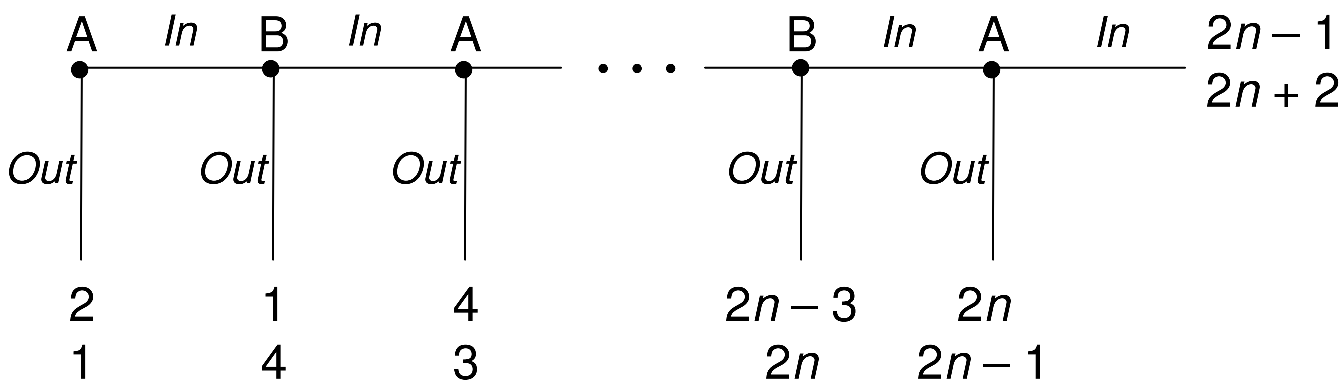
Prisoner's Dilemma

		B	
		"Cooperate"	*"Defect"*
A	*"Cooperate"*	3 3	4 0
	"Defect"	0 4	1 1

A = Ann

B = Bob

Centipede



A = Ann

B = Bob

Introduced by Rosenthal (1981)

The puzzle:

Is it better to secure a piece of the “pie” now?

Or, is it better to let the game continue for a while?

Nash Equilibrium

John Nash: “Non-Cooperative Games,” doctoral dissertation, 1950

“What would be a ‘rational’ prediction of the behavior to be expected of rationally playing the game in question?”

Nash Equilibrium in Centipede

In any Nash equilibrium of Centipede, Ann must choose *Out* immediately

Proof:

Let (s^a, s^b) be a pair of strategies

Let the path induced by (s^a, s^b) be *In* until the kth node and then *Out*

Suppose Bob moves at the kth node

Then s^a must tell Ann to play *In* at the $(k - 1)$th node

Let r^a be the strategy for Ann that plays *In* until the $(k - 1)$th node and then *Out*

Then r^a gives Ann a strictly higher payoff than s^a against s^b

(The argument extends to mixed strategies)

Nash Equilibrium: Theory vs. Evidence

The Nash-equilibrium prediction is not borne out in experiments
(e.g. Camerer 2003)

But Nash equilibrium is not a consequence of rational behavior on the part of the players!

A condition for Nash equilibrium:

> Each player is rational and each player assigns probability 1 to the actual strategies chosen by the other players
>
> (See Aumann and Brandenburger 1995 for this and also an "epistemic" treatment of mixed equilibria)

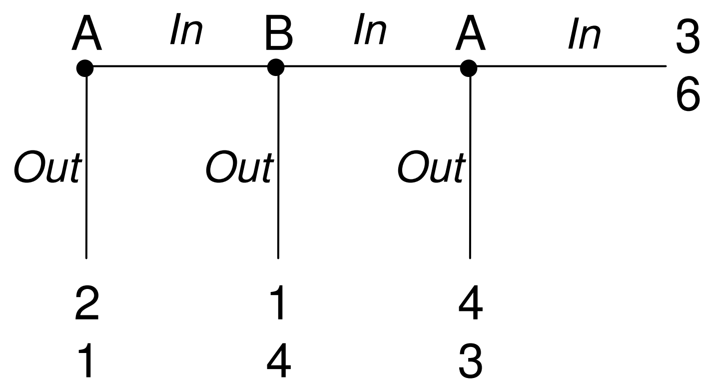
Rationality

What, then, is the implication of rationality in a game?

Or, what is the implication of supposing each player is rational, each player thinks the other players are rational, each player thinks the other players think …?

This is a central question of the epistemic program in game theory

Example: Three-Legged Centipede



How should Ann play?

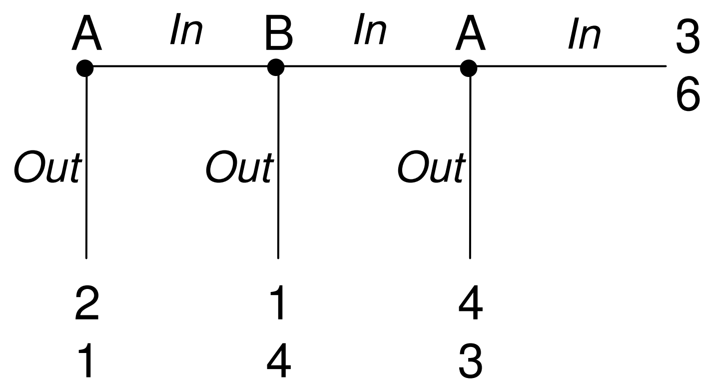
This depends on what she thinks Bob will do

This depends on what he thinks she will do (at the last node)

⟶ The Backward-Induction Algorithm

⟶ Ann plays *Out* at the first node

Three-Legged Centipede contd.



But what if Ann nevertheless plays *In* at the first node?

Bob might conclude that Ann is not playing rationally—and that she might play *In* at the last node

So, Bob might play *In*

So, it seems it can be rational for Ann to play *In* at the first node!

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

Epistemic Game Theory

Classically, a game was fully described as a game matrix (e.g. the Prisoner's Dilemma) or a game tree (e.g. Centipede)

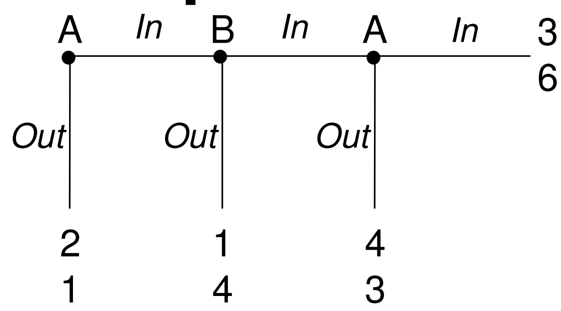
The epistemic approach adds a description of what the players believe about the strategies chosen, about what they believe other players believe about the strategies chosen, …

Formally, this is done via "epistemic type structures"

[Harsanyi (1967-8) introduced types to talk about uncertainty over the payoffs

The epistemic program adapts the technique to talk about uncertainty over the strategies chosen—or about uncertainty over both the payoffs and the strategies]

Example of Epistemic Analysis



Using conditional probability systems (Renyi 1955)

Type t^a

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

Type u^a

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

Type t^b

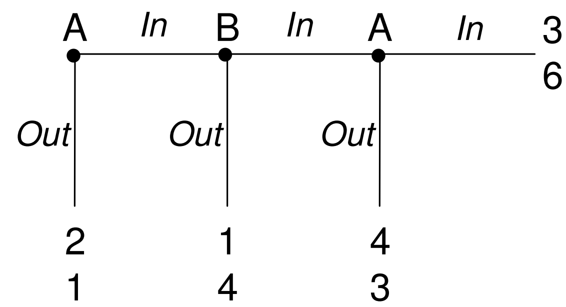
T^a \ S^a	*Out*	*In-Out*	*In-In*
u^a	1 [0]	0 [0]	0 [0]
t^a	0 [0]	0 [0]	0 [1]

Type u^b

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

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

Strong Belief



Bob doesn’t have to give up on Ann’s rationality if his node is reached (Battigalli and Siniscalchi 2002)

In this case, he will play *Out*

If Ann anticipates this, she will play *Out* at the first node

Theorem (Battigalli and Friedenberg, in preparation):

> Fix a CPS-based type structure for *n*-legged Centipede, and a state at which there is rationality and common “strong belief” of rationality. Then Ann plays *Out*.

An Impossibility Result

What is the prediction if

the players are rational

each strongly believes that the others are rational

and so on?

Consider:

A1: Ann is rational	B1: Bob is rational
A2: Ann is rational and strongly believes B1	B2: Bob is rational and strongly believes A1
A3: Ann is rational, strongly believes B1, and strongly believes B2	B3: Bob is rational, strongly believes A1, and strongly believes A2
• • •	• • •

Impossibility Result contd.

A1: Ann is rational	B1: Bob is rational
A2: Ann is rational and strongly believes B1	B2: Bob is rational and strongly believes A1
A3: Ann is rational, strongly believes B1, and strongly believes B2	B3: Bob is rational, strongly believes A1, and strongly believes A2
• • •	• • •

Definition (Battigalli and Siniscalchi 2002):

> A player strongly believes an event *S* if the player believes *S*, given any event *K* the player knows, provided only that *K* does not contradict *S*

Now suppose each of the events A1, A2, A3, … obtains

Let *S* be the event [B1 and B2 and B3 and …] and let *K* be not-*S*

K is consistent with B1, so Ann believes B1 given *K* (by A2)
K is consistent with B2, so Ann believes B2 given *K* (by A3)

• • •

So, Ann must believe *S* given *K*, which is impossible!

(Brandenburger, Friedenberg, and Keisler 2008)

Elements of the Formal Treatment: "Invariance"

What is the significance of the event not-[B1 and B2 and B3 and …]?

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

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

If we build epistemic game theory using a decision theory that satisfies admissibility (on the matrix), we can hope to get invariance at the game-theoretic level too

Invariance contd.

BFK (2008) build in admissibility (on the matrix) by using lexicographic probability systems (Blume, Brandenburger, and Dekel 1991) with full support

There is a canonical map from full-support LPS's to CPS's where the family of conditioning events consists of the open sets (BFK 2006)

The formal treatment of the impossibility result involves the event

K = not-[cl(B1) and cl(B2) and cl(B3) and …]

which is open

Elements of the Formal Treatment: “Completeness”

We need the idea that the events B1, B2, B3, … keep shrinking

This will be true if, for each *m*, Bob has types that strongly believe A(*m*-1) but not A*m*

The formal condition is that there are points in B*m* that are not in cl(B(*m*+1))

The impossibility result is couched in a “complete” structure (Brandenburger 2003)—one that contains, in a sense, all possible types

Discussion

Start with the "economic" data of a game:

Strategy sets ("Opportunity sets")

Payoff functions ("Tastes")

Can we derive implications about what the players will do and what they will believe?

The usual approach is to ask the question under the condition of Nash equilibrium

But we can ask the question under the more basic condition that each player is rational, strongly believes the others are rational, …

Formally, we pose the question in a complete structure—so no a priori restrictions are placed on what the players believe

Discussion contd.

But this is impossible

The conclusion appears to be that we must drop completeness—we must rule out a priori certain beliefs on the part of the players

Alternatively put, we must specify in the description of a game:

Strategy sets ("Opportunity sets")

Payoff functions ("Tastes")

Spaces of possible beliefs, beliefs about beliefs, … ("Context")

The traditional "economic" data are not enough

This is the thesis of the epistemic program in game theory