Rationality

Lecture 14

Eric Pacuit

Center for Logic and Philosophy of Science
Tilburg University
ai.stanford.edu/∼epacuit
e.j.pacuit@uvt.nl

December 1, 2010

		Bob	
		D	C
Ann	D	4,4	1,3
	C	3,1	2,2

What should/will Ann (Bob) do?

		Bob	
		D	C
Ann	D	4,4	1,3
	C	3,1	2,2

Assurance Game

What should/will Ann (Bob) do?

Bob

Ann	D	C
D	3,3	1,4
C	4,1	2,2

Prisoner's Dilemma

Bob

Ann	D	C
D	4,4	1,3
C	3,1	2,2

Assurance Game

What should/will Ann (Bob) do?

Nozick: Symbolic Utility

"Yet the symbolic value of an act is not determined solely by *that* act.

Nozick: Symbolic Utility

"Yet the symbolic value of an act is not determined solely by *that* act. The act's meaning can depend upon what other acts are available with what payoffs and what acts also are available to the other party or parties.

Nozick: Symbolic Utility

"Yet the symbolic value of an act is not determined solely by *that* act. The act's meaning can depend upon what other acts are available with what payoffs and what acts also are available to the other party or parties. What the act symbolizes is something it symbolizes when done in *that* particular situation, in preference to *those* particular alternatives.

Nozick: Symbolic Utility

"Yet the symbolic value of an act is not determined solely by *that* act. The act's meaning can depend upon what other acts are available with what payoffs and what acts also are available to the other party or parties. What the act symbolizes is something it symbolizes when done in *that* particular situation, in preference to *those* particular alternatives. If an act symbolizes "being a cooperative person," it will have that meaning not simply because it has the two possible payoffs it does

Nozick: Symbolic Utility

"Yet the symbolic value of an act is not determined solely by *that* act. The act's meaning can depend upon what other acts are available with what payoffs and what acts also are available to the other party or parties. What the act symbolizes is something it symbolizes when done in *that* particular situation, in preference to *those* particular alternatives. If an act symbolizes "being a cooperative person," it will have that meaning not simply because it has the two possible payoffs it does but also because it occupies a particular position within the two-person matrix — that is, being a dominated action that (when joined with the other person's dominated action) yield a higher payoff to each than does the combination of dominated actions. " (pg. 55)

R. Nozick. *The Nature of Rationality*. Princeton University Press, 1993.

		Bob	
		D	C
Ann	D	3,3	1,4
	C	4,1	2,2

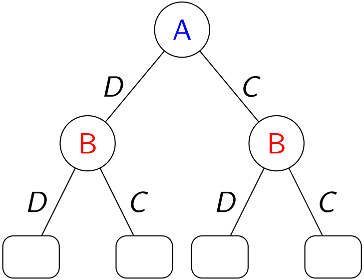
Prisoner's Dilemma

What should/will Ann (Bob) do?

Bob

Ann		D	C
	D	3,3	1,4
	C	4,1	2,2

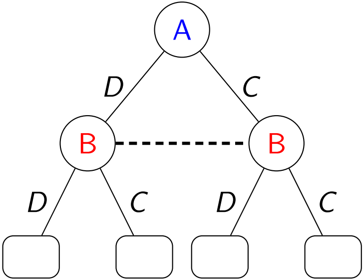
Prisoner's Dilemma



What should/will Ann (Bob) do?

		Bob	
		D	C
Ann	D	3,3	1,4
	C	4,1	2,2

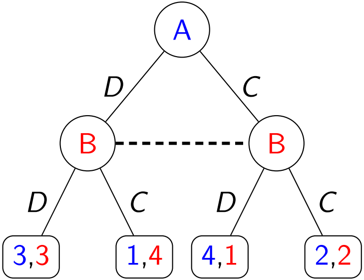
Prisoner's Dilemma



What should/will Ann (Bob) do?

		Bob	
		D	C
Ann	D	3,3	1,4
	C	4,1	2,2

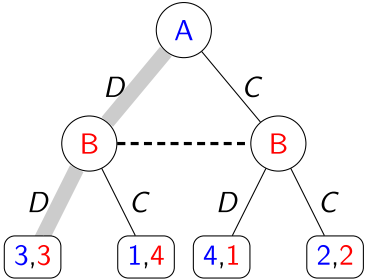
Prisoner's Dilemma



What should/will Ann (Bob) do?

		Bob	
		D	C
Ann	D	3,3	1,4
	C	4,1	2,2

Prisoner's Dilemma

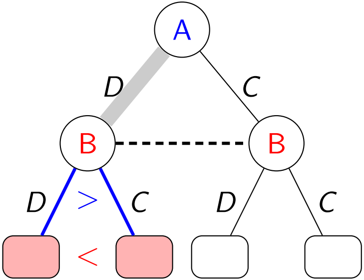


What should/will Ann (Bob) do?

Bob

Ann	D	C
D	3,3	1,4
C	4,1	2,2

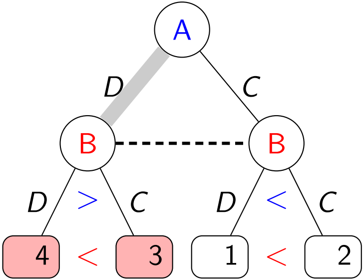
Prisoner's Dilemma



What should/will Ann (Bob) do?

		Bob	
		D	C
Ann	D	3,3	1,4
	C	4,1	2,2

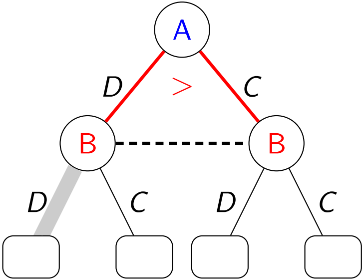
Prisoner's Dilemma



What should/will Ann (Bob) do?

Ann \ Bob	D	C
D	3,3	1,4
C	4,1	2,2

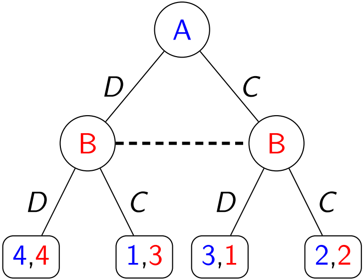
Prisoner's Dilemma



What should/will Ann (Bob) do?

		Bob	
		D	C
Ann	D	3,3	1,4
	C	4,1	2,2

Prisoner's Dilemma



What should/will Ann (Bob) do?

“Game theorists think it just plain wrong to claim that the Prisoners’ Dilemma embodies the essence of the problem of human cooperation.

"Game theorists think it just plain wrong to claim that the Prisoners' Dilemma embodies the essence of the problem of human cooperation. On the contrary, it represents a situation in which the dice are as loaded against the emergence of cooperation as they could possibly be. If the great game of life played by the hum species were the Prisoner's Dilemma, we wouldn't have evolved as social animals!

"Game theorists think it just plain wrong to claim that the Prisoners' Dilemma embodies the essence of the problem of human cooperation. On the contrary, it represents a situation in which the dice are as loaded against the emergence of cooperation as they could possibly be. If the great game of life played by the hum species were the Prisoner's Dilemma, we wouldn't have evolved as social animals! .... No paradox of rationality exists. Rational players don't cooperate in the Prisoners' Dilemma, because the conditions necessary for rational cooperation are absent in this game." (pg. 63)

K. Binmore. *Natural Justice*. Oxford University Press, 2005.

Hi-Low

		Bob	
		D	C
Ann	D	3,3	0,0
	C	0,0	1,1

What should/will Ann (Bob) do?

N. Bardsley, J. Mehta, C. Starmer and R. Sugden. *The Nature of Salience Revisited: Cognitive Hierarch Theory* versus *Team Reasoning*. Economic Journal.

N. Bardsley, J. Mehta, C. Starmer and R. Sugden. *The Nature of Salience Revisited: Cognitive Hierarch Theory* versus *Team Reasoning*. Economic Journal.

$$\{water, beer, sherry, whisky, wine\}$$

N. Bardsley, J. Mehta, C. Starmer and R. Sugden. *The Nature of Salience Revisited: Cognitive Hierarch Theory* versus *Team Reasoning*. Economic Journal.

$$\{water, beer, sherry, whisky, wine\}$$

Task 1: pick an option

N. Bardsley, J. Mehta, C. Starmer and R. Sugden. *The Nature of Salience Revisited: Cognitive Hierarch Theory* versus *Team Reasoning*. Economic Journal.

$$\{\textbf{water}, \textit{beer}, \textit{sherry}, \textit{whisky}, \textit{wine}\}$$

Task 1: pick an option

N. Bardsley, J. Mehta, C. Starmer and R. Sugden. *The Nature of Salience Revisited: Cognitive Hierarch Theory* versus *Team Reasoning*. Economic Journal.

$$\{\textbf{water}, \textit{beer}, \textit{sherry}, \textit{whisky}, \textit{wine}\}$$

Task 1: pick an option

Task 2: guess what your opponent picked

N. Bardsley, J. Mehta, C. Starmer and R. Sugden. *The Nature of Salience Revisited: Cognitive Hierarch Theory* versus *Team Reasoning*. Economic Journal.

$$\{\textbf{water}, \textit{beer}, \textit{sherry}, \textit{whisky}, \textit{wine}\}$$

Task 1: pick an option
Task 2: guess what your opponent picked
Task 3: try to coordinate with your (unknown) partner

N. Bardsley, J. Mehta, C. Starmer and R. Sugden. *The Nature of Salience Revisited: Cognitive Hierarch Theory* versus *Team Reasoning*. Economic Journal.

$$\{\textbf{water}, beer, sherry, whisky, wine\}$$

Task 1: pick an option

Task 2: guess what your opponent picked

Task 3: try to coordinate with your (unknown) partner

	pick	guess	coordinate
water	20	15	38
beer	13	26	11
sherry	4	1	0
whisky	6	6	5
wine	10	4	2

Footballer Example

A and B are players in the same football team. A has the ball, but an opposing player is converging on him. He can pass the ball to B, who has a chance to shoot. There are two directions in which A can move the ball, *left* and *right*, and correspondingly, two directions in which B can run to intercept the pass. If both choose *left* there is a 10% chance that a goal will be scored. If they both choose *right*, there is a 11% change. Otherwise, the chance is zero. There is no time for communication; the two players must act simultaneously.

What should they do?

R. Sugden. *The Logic of Team Reasoning*. Philosophical Explorations (6)3, pgs. 165 - 181 (2003).

Column

		l	r
Row	l		
	r		

Column

		l	r
Row	l	(10,10)	(0,0)
	r	(0,0)	(11,11)

		Column	
		l	r
Row	l	(10,10)	(0,0)
	r	(0,0)	(11,11)

Row: What should I do?

		Column	
		l	r
Row	l	(10,10)	(0,0)
	r	(0,0)	(11,11)

Row: What should I do? (r if the probability of Column choosing r is $> \frac{10}{21}$ and l if the probability of Column choosing l is $> \frac{11}{21}$)

Column

		l	r
Row	l	(10,10)	(0,0)
	r	(0,0)	(11,11)

Row: What should *we* do?

Column

		l	r
Row	l	(10,10)	(0,0)
	r	(0,0)	(11,11)

Team Reasoning: escape from the infinite regress? why should this "mode of reasoning" be endorsed?

"The basic intellectual premise, or working hypothesis, for rational players in this game seems to be the premise that some rule must be used if success is to exceed coincidence, and that the best rule to be found, whatever its rationalization, is consequently a rational rule." (Thomas Schelling)

Rationality in Interaction

What does it mean to be rational when the outcome of one's action depends upon the actions of other people and everyone is trying to guess what the others will do?

Rationality in Interaction

What does it mean to be rational when the outcome of one's action depends upon the actions of other people and everyone is trying to guess what the others will do?

In social interaction, rationality has to be enriched with further assumptions about individuals' **mutual knowledge and beliefs**, *but these assumptions are not without consequence.*

C. Bicchieri. *Rationality and Game Theory*. Chapter 10 in [HR].

Example: Common Knowledge

Suppose there are two friends Ann and Bob are on a bus separated by a crowd.

Example: Common Knowledge

Suppose there are two friends Ann and Bob are on a bus separated by a crowd. Before the bus comes to the next stop a mutual friend from outside the bus yells "get off at the next stop to get a drink?".

Example: Common Knowledge

Suppose there are two friends Ann and Bob are on a bus separated by a crowd. Before the bus comes to the next stop a mutual friend from outside the bus yells “get off at the next stop to get a drink?”.

Say Ann is standing near the front door and Bob near the back door.

Example: Common Knowledge

Suppose there are two friends Ann and Bob are on a bus separated by a crowd. Before the bus comes to the next stop a mutual friend from outside the bus yells "get off at the next stop to get a drink?".

Say Ann is standing near the front door and Bob near the back door. When the bus comes to a stop, will they get off?

Example: Common Knowledge

Suppose there are two friends Ann and Bob are on a bus separated by a crowd. Before the bus comes to the next stop a mutual friend from outside the bus yells "get off at the next stop to get a drink?".

Say Ann is standing near the front door and Bob near the back door. When the bus comes to a stop, will they get off?

D. Lewis. *Convention*. 1969.

M. Chwe. *Rational Ritual*. 2001.

"*Common Knowledge*" is informally described as what any fool would know, given a certain situation: It encompasses what is relevant, agreed upon, established by precedent, assumed, being attended to, salient, or in the conversational record.

"*Common Knowledge*" is informally described as what any fool would know, given a certain situation: It encompasses what is relevant, agreed upon, established by precedent, assumed, being attended to, salient, or in the conversational record.

It is not Common Knowledge who "(formally) defined" Common Knowledge!

The first formal definition of common knowledge?

M. Friedell. *On the Structure of Shared Awareness*. Behavioral Science (1969).

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

The first formal definition of common knowledge?

M. Friedell. *On the Structure of Shared Awareness*. Behavioral Science (1969).

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

The first rigorous analysis of common knowledge

D. Lewis. *Convention, A Philosophical Study*. 1969.

The first formal definition of common knowledge?

M. Friedell. *On the Structure of Shared Awareness*. Behavioral Science (1969).

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

The first rigorous analysis of common knowledge

D. Lewis. *Convention, A Philosophical Study*. 1969.

Fixed-point definition: $\gamma :=$ *i* and *j* know that (φ and γ)

G. Harman. *Review of* Linguistic Behavior. Language (1977).

J. Barwise. *Three views of Common Knowledge*. TARK (1987).

The first formal definition of common knowledge?

M. Friedell. *On the Structure of Shared Awareness*. Behavioral Science (1969).

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

The first rigorous analysis of common knowledge

D. Lewis. *Convention, A Philosophical Study*. 1969.

Fixed-point definition: $\gamma :=$ i and j know that (φ and γ)

G. Harman. *Review of* Linguistic Behavior. Language (1977).

J. Barwise. *Three views of Common Knowledge*. TARK (1987).

Shared situation: There is a *shared situation* s such that (1) s entails φ, (2) s entails everyone knows φ, plus other conditions

H. Clark and C. Marshall. *Definite Reference and Mutual Knowledge*. 1981.

M. Gilbert. *On Social Facts*. Princeton University Press (1989).

P. Vanderschraaf and G. Sillari. *"Common Knowledge", The Stanford Encyclopedia of Philosophy (2009)*.
`http://plato.stanford.edu/entries/common-knowledge/`.

Suppose you are told "Ann and Bob are going together,"' and respond "sure, that's common knowledge." What you mean is not only that everyone knows this, but also that the announcement is pointless, occasions no surprise, reveals nothing new; in effect, that the situation after the announcement does not differ from that before. ...the event "Ann and Bob are going together" — call it E — is common knowledge if and only if some event — call it F — happened that entails E and also entails all players' knowing F (like all players met Ann and Bob at an intimate party). *(Aumann, pg. 271, footnote 8)*

Key Assumptions

CK1 The structure of the game, including players' strategy sets and payoff functions, is common knowledge among the players.

CK2 The players are rational (i.e., they are expected utility maximizers) and this is common knowledge.

Common Knowledge of Rationality: Iterated Removal of Strictly Dominated Strategies

Bob

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

Common Knowledge of Rationality: Iterated Removal of Strictly Dominated Strategies

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

There is no prior such that R is rational for Bob.

Common Knowledge of Rationality: Iterated Removal of Strictly Dominated Strategies

Bob

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

If Ann knows this, then she does not consider R a option for Bob

Common Knowledge of Rationality: Iterated Removal of Strictly Dominated Strategies

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

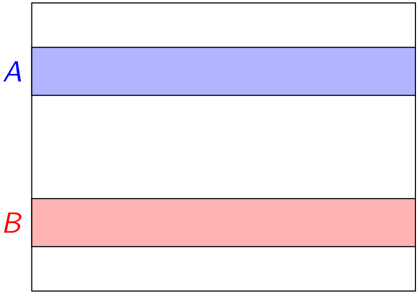
So, U is the only rational choice.

Common knowledge of rationality (players will not choose strictly dominated actions) leads to a process of iterated removal of strictly dominated strategies.

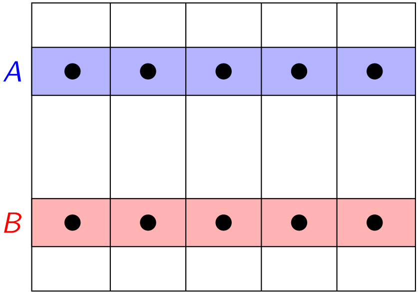
Common knowledge of rationality (players will not choose strictly dominated actions) leads to a process of iterated removal of strictly dominated strategies.

What about *weak dominance*?

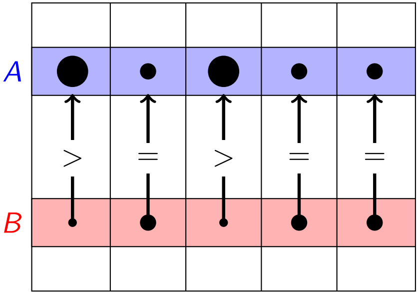
Weak Dominance



Weak Dominance



Weak Dominance



Iterated Admissibility

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

Iterated Admissibility

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

Suppose rationality incorporates *weak dominance* (i.e., *admissibility* or *cautiousness*).

Iterated Admissibility

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

Suppose rationality incorporates *weak dominance* (i.e., *admissibility* or *cautiousness*).

1. Both Row and Column should use a *full-support* probability measure
2. But if Row thinks that Column is **rational** then should she not assign probability 1 to L?

Iterated Admissibility

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

Suppose rationality incorporates *weak dominance* (i.e., *admissibility* or *cautiousness*).

1. Both Row and Column should use a *full-support* probability measure
2. But if Row thinks that Column is **rational** then should she not assign probability 1 to L?

The condition that the players incorporate admissibility *into their rationality calculations seems to conflict with the condition that the players think the other players are rational (there is a tension between admissibility and strategic reasoning)*

Iterated Removal of Weakly Dominated Strategies

Ann \ Bob	L	R
T	1,1	1,0
B	1,0	0,1

Iterated Removal of Weakly Dominated Strategies

Ann \ Bob	L	R
T	1,1	1,0
B	1,0	0,1

T weakly dominates B

Iterated Removal of Weakly Dominated Strategies

Bob

		L	R
Ann	T	1,1	1,0
	B	1,0	0,1

Then L strictly dominates R.

Iterated Removal of Weakly Dominated Strategies

Bob

Ann		L	R
	T	1,1	1,0
	B	1,0	0,1

The IA set

Iterated Removal of Weakly Dominated Strategies

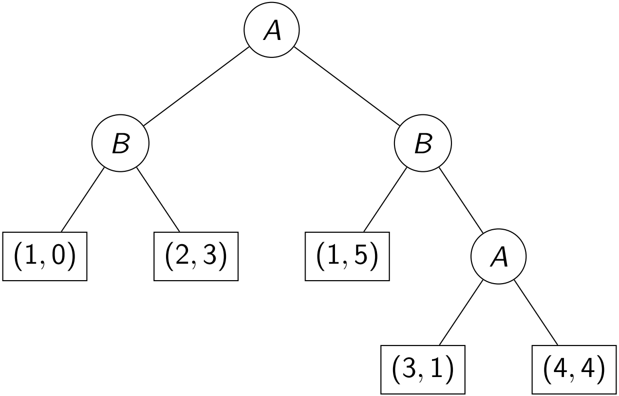
Bob

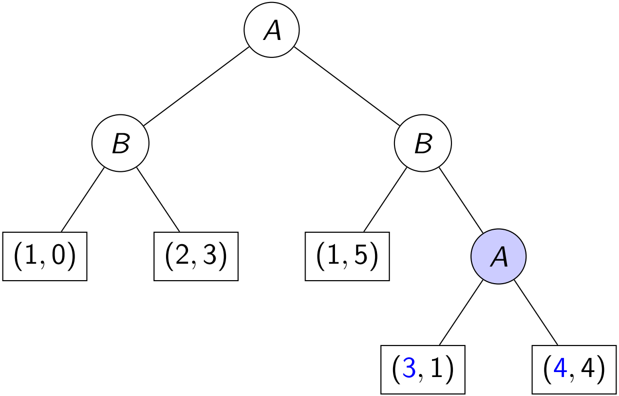
Ann		L	R
	T	1,1	1,0
	B	1,0	0,1

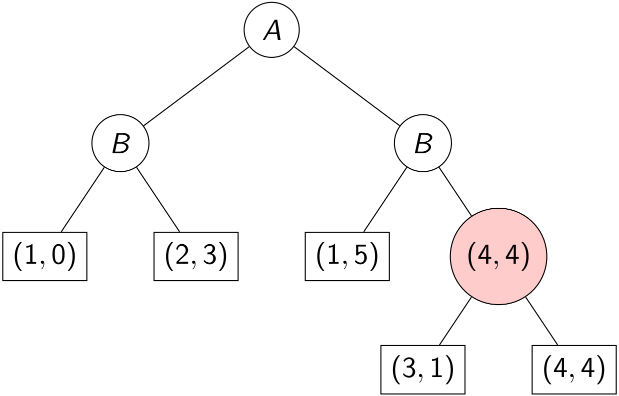
But, now what is the reason for not playing B?

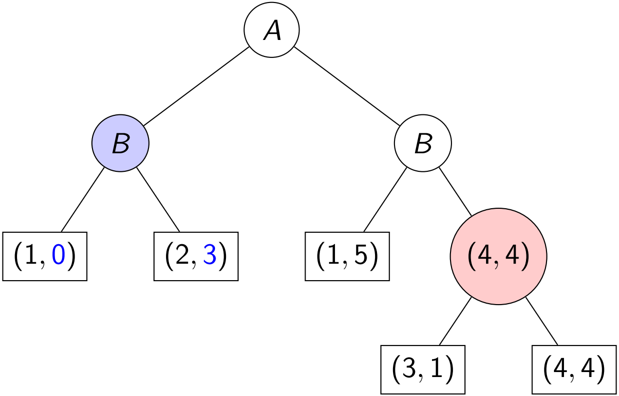
Backwards Induction

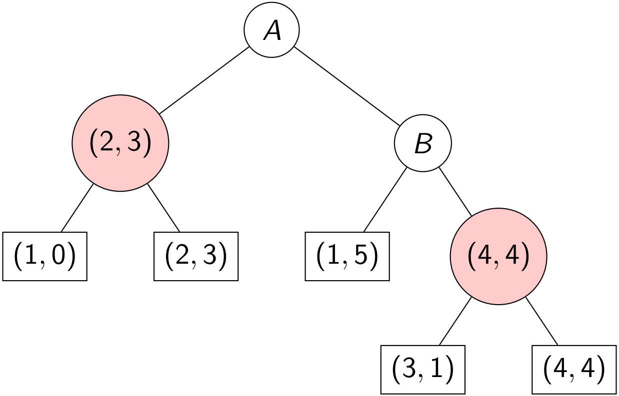
Invented by Zermelo, Backwards Induction is an iterative algorithm for “solving” and extensive game.

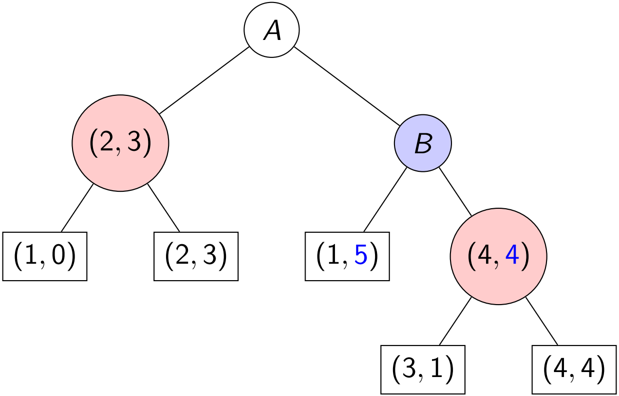

A
B
B
(1, 0)
(2, 3)
(1, 5)
A
(3, 1)
(4, 4)

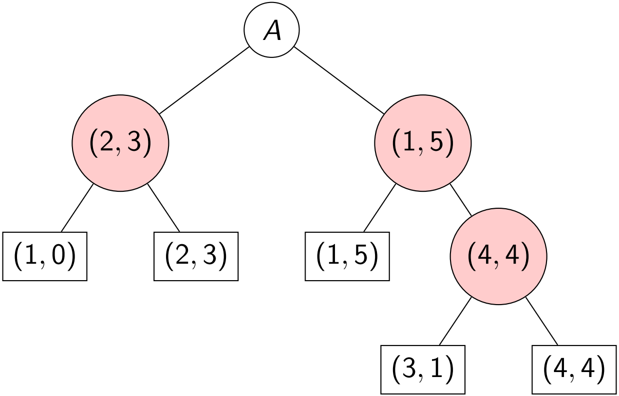

A
B
B
(1, 0)
(2, 3)
(1, 5)
A
(3, 1)
(4, 4)

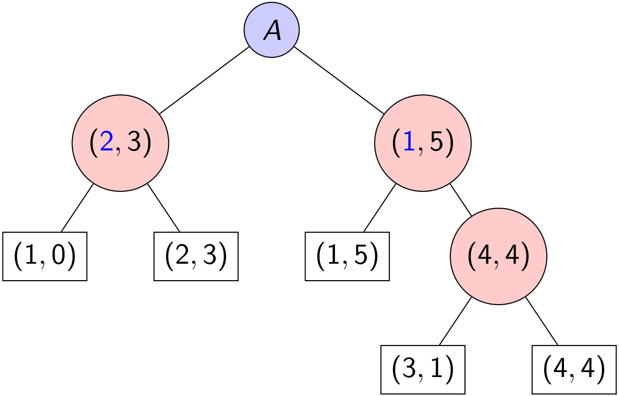

A
B
B
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)

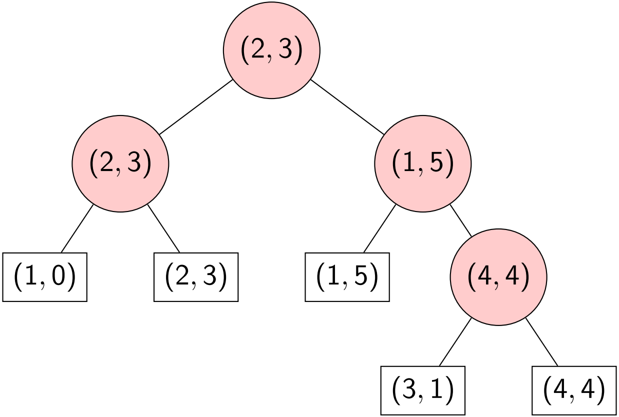

A
B
B
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)

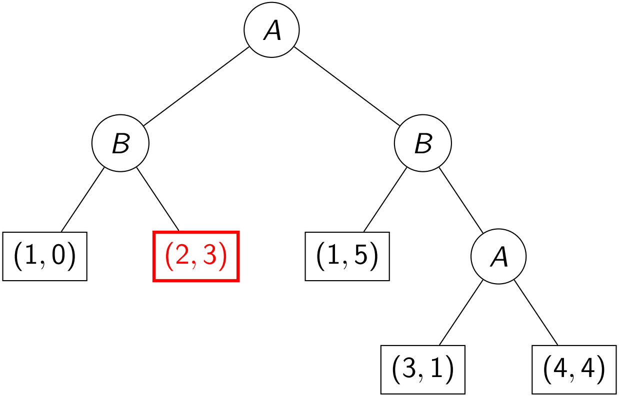

A
(2, 3)
B
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)

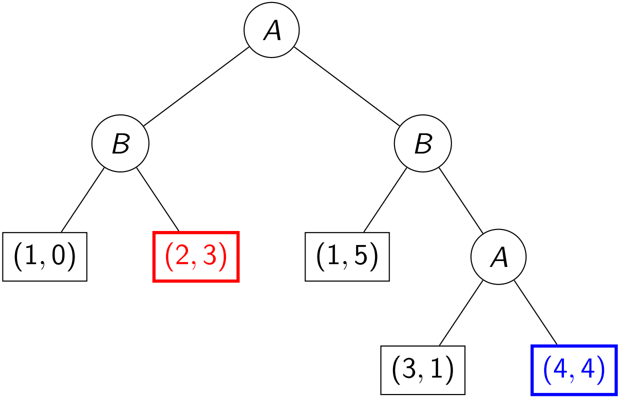

A
(2, 3)
B
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)


A
$(2,3)$
$(1,5)$
$(1,0)$
$(2,3)$
$(1,5)$
$(4,4)$
$(3,1)$
$(4,4)$

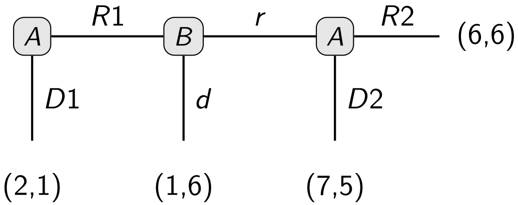

A
(2, 3)
(1, 5)
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)


(2, 3)
(2, 3)
(1, 5)
(1, 0)
(2, 3)
(1, 5)
(4, 4)
(3, 1)
(4, 4)

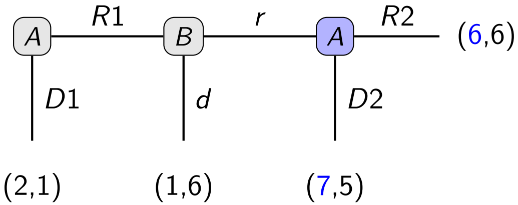

A
B
B
(1, 0)
(2, 3)
(1, 5)
A
(3, 1)
(4, 4)


A
B
B
(1, 0)
(2, 3)
(1, 5)
A
(3, 1)
(4, 4)

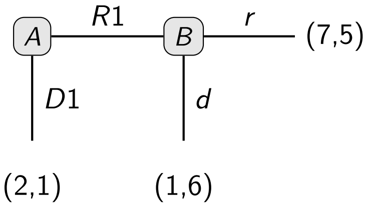
BI Puzzle



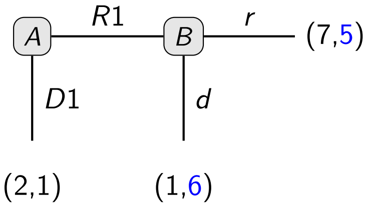
BI Puzzle



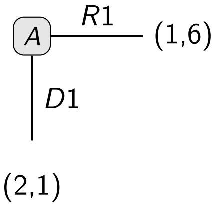
BI Puzzle



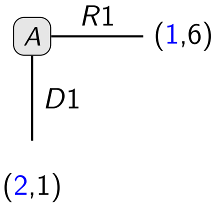
BI Puzzle



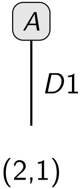
BI Puzzle



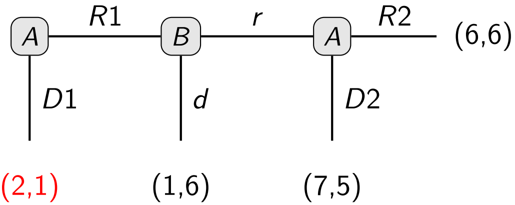
BI Puzzle



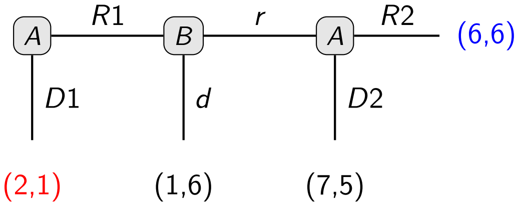
BI Puzzle



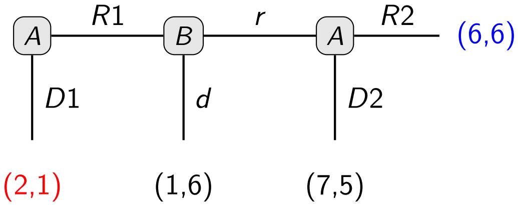
BI Puzzle



But what if...



But what if...



- Are the players *irrational*?
- What argument leads to the BI solution?

Repeated Prisoner's Dilemma

	C	D
C	3,3	0,4
D	4,0	1,1

Repeated Prisoner's Dilemma

	C	D
C	3,3	0,4
D	4,0	1,1

Repeated Prisoner's Dilemma

	C	D
C	3,3	0,4
D	4,0	1,1

	C	D
C	3,3	0,4
D	4,0	1,1

	C	D
C	3,3	0,4
D	4,0	1,1

	C	D
C	3,3	0,4
D	4,0	1,1

$\cdots$

Repeated Prisoner's Dilemma

	C	D
C	3,3	0,4
D	4,0	1,1

	C	D
C	3,3	0,4
D	4,0	1,1

	C	D
C	3,3	0,4
D	4,0	1,1

	C	D
C	3,3	0,4
D	4,0	1,1

Repeated Prisoner's Dilemma

	C	D
C	3,3	0,4
D	4,0	1,1

	C	D
C	3,3	0,4
D	4,0	1,1

	C	D
C	3,3	0,4
D	4,0	1,1

	C	D
C	3,3	0,4
D	4,0	1,1

Repeated Prisoner's Dilemma

	C	D
C	3,3	0,4
D	4,0	1,1

	C	D
C	3,3	0,4
D	4,0	1,1

	C	D
C	3,3	0,4
D	4,0	1,1

	C	D
C	3,3	0,4
D	4,0	1,1

What about "tit-for-tat"?

Repeated Prisoner's Dilemma

	C	D
C	3,3	0,4
D	4,0	1,1

	C	D
C	3,3	0,4
D	4,0	1,1

	C	D
C	3,3	0,4
D	4,0	1,1

	C	D
C	3,3	0,4
D	4,0	1,1

What about "tit-for-tat"?

Is anything missing in these models?

Formally, a game is described by its strategy sets and payoff functions.

Formally, a game is described by its strategy sets and payoff functions. But in real life, may other parameters are relevant; there is a lot more going on. Situations that substantively are vastly different may nevertheless correspond to precisely the same strategic game.

Formally, a game is described by its strategy sets and payoff functions. But in real life, may other parameters are relevant; there is a lot more going on. Situations that substantively are vastly different may nevertheless correspond to precisely the same strategic game. For example, in a parliamentary democracy with three parties, the winning coalitions are the same whether the parties hold a third of the seats, or, say, 49%, 39%, and 12% respectively.

Formally, a game is described by its strategy sets and payoff functions. But in real life, may other parameters are relevant; there is a lot more going on. Situations that substantively are vastly different may nevertheless correspond to precisely the same strategic game. For example, in a parliamentary democracy with three parties, the winning coalitions are the same whether the parties hold a third of the seats, or, say, 49%, 39%, and 12% respectively. But the political situations are quite different.

Formally, a game is described by its strategy sets and payoff functions. But in real life, may other parameters are relevant; there is a lot more going on. Situations that substantively are vastly different may nevertheless correspond to precisely the same strategic game. For example, in a parliamentary democracy with three parties, the winning coalitions are the same whether the parties hold a third of the seats, or, say, 49%, 39%, and 12% respectively. But the political situations are quite different. The difference lies in the attitudes of the players, in their expectations about each other, in custom, and in history, though the rules of the game do not distinguish between the two situations.

R. Aumann and J. H. Dreze. *Rational Expectation in Games*. American Economic Review (2008).

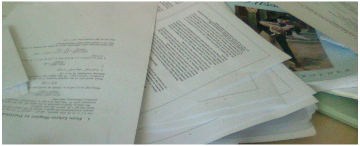
Two questions

- What should the players *do* in a game-theoretic situation and what should they expect? (Assuming everyone is **rational** and recognize each other's rationality)

- What are the assumptions about rationality and the players' knowledge/beliefs underlying the various solution concepts? *Why* would the agents' follow a particular solution concept?

Writing a paper together

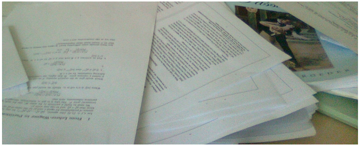
	C	D
C		
D		



Writing a paper together

Problem of Cooperation.

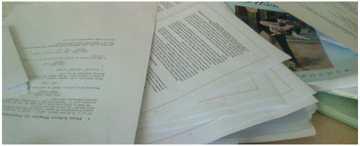
	C	D
C	3,3	0,4
D	4,0	1,1



Writing a paper together

Problem of Coordination.

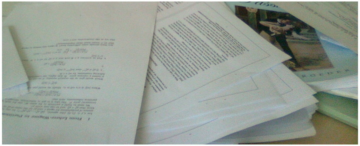
	C	D
C	3,3	0,0
D	0,0	1,1



Writing a paper together

	C	D
C		
D		

Intuitively, we solve these problem by working together.
This is the question of collective agency.



Individual vs collective agency

Different contexts of agency

Different contexts of agency

- Individual decision making and individual action against nature.
  - Ex: Gambling.

Different contexts of agency

- Individual decision making and individual action against nature.

- Individual decision making in interaction.
  - Ex: Playing chess.

Different contexts of agency

- Individual decision making and individual action against nature.
- Individual decision making in interaction.
- Collective decision making.
  - Ex: Carrying the piano.

Different contexts of agency

- Individual decision making and individual action against nature.
- Individual decision making in interaction.
- Collective decision making.

Next: Social Choice Theory and Group Preferences