

# Introduction to Epistemic Reasoning in Interaction

Eric Pacuit

Center for Logic and Philosophy of Science  
Tilburg University  
`ai.stanford.edu/~epacuit`

December 9, 2009

---

We are interested in reasoning about rational agents interacting in *social* situations.

---

We are interested in reasoning about rational agents interacting in *social* situations.

- ▶ Philosophy (social philosophy, epistemology)
- ▶ Game Theory
- ▶ Social Choice Theory
- ▶ AI (multiagent systems)

We are interested in reasoning about **rational agents** interacting in *social* situations.

*What is a rational agent?*

- ▶ maximize expected utility (instrumentally rational)
- ▶ react to observations
- ▶ revise beliefs when learning a *surprising* piece of information
- ▶ understand higher-order information
- ▶ plans for the future
- ▶ asks questions
- ▶ ????

We are interested in **reasoning about** rational agents interacting in *social* situations.

There is a jungle of formal systems!

- ▶ logics of informational attitudes (knowledge, beliefs, certainty)
- ▶ logics of action & agency
- ▶ temporal logics/dynamic logics
- ▶ logics of motivational attitudes (preferences, intentions)

*(Not to mention various game-theoretic/social choice models and logical languages for reasoning about them)*

We are interested in **reasoning about** rational agents interacting in *social* situations.

There is a jungle of formal systems!

- ▶ How do we compare different logical systems studying the same phenomena?
- ▶ How *complex* is it to reason about rational agents?
- ▶ (How) should we *merge* the various logical systems?
- ▶ What do the logical frameworks contribute to the discussion on rational agency?

*and logical languages for reasoning about them)*

We are interested in reasoning about rational agents **interacting in *social situations***.

- ▶ playing a game (eg. a card game)
- ▶ having a conversation
- ▶ executing a *social procedure*
- ▶ making a group decision (eg., coordination problem)
- ▶ ....

## Example (1)

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



## Example (1)

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 (1)

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 (1)

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 (1)

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.

## Example (2)

$A$  and  $B$  are players in the same football team.  $A$  has the ball, but an opposing player is converging on him.

## Example (2)

$A$  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.

## Example (2)

$A$  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.

## Example (2)

$A$  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.



## Example (2)

$A$  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.

## Example (2)

$A$  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.

## Example (2)

$A$  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.

## Example (2)

$A$  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).

## Goals for today

1. An Introduction to (Formal) Tools for Analytic Philosophy
  - Strategic reasoning
  - Epistemic reasoning

*focus on intuitions not technical details*

2. Epistemic Reasoning in Games

## Just Enough Game Theory

*“Game theory is a bag of analytical tools designed to help us understand the phenomena that we observe when decision-makers interact.”*

Osborne and Rubinstein. *Introduction to Game Theory*. MIT Press .

## Just Enough Game Theory

*“Game theory is a bag of analytical tools designed to help us understand the phenomena that we observe when decision-makers interact.”*

Osborne and Rubinstein. *Introduction to Game Theory*. MIT Press .

A **game** is a description of strategic interaction that includes

- ▶ actions the players *can* take
- ▶ description of the players' interests (i.e., preferences),

## Just Enough Game Theory

*“Game theory is a bag of analytical tools designed to help us understand the phenomena that we observe when decision-makers interact.”*

Osborne and Rubinstein. *Introduction to Game Theory*. MIT Press .

A **game** is a description of strategic interaction that includes

- ▶ actions the players *can* take
- ▶ description of the players' interests (i.e., preferences),

*It does not specify the actions that the players do take.*



A **solution concept** is a systematic description of the outcomes that may emerge in a family of games.

This is the starting point for most of game theory and includes many variants: Nash equilibrium, backwards inductions, 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.

A **solution concept** is a systematic description of the outcomes that may emerge in a family of games.

This is the starting point for most of game theory and includes many variants: Nash equilibrium, backwards inductions, 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.

For this course, **solution concepts** are more of an *endpoint*.

## Some Formal Details

- ▶ Strategic Games
- ▶ Nash Equilibrium
- ▶ Extensive Games
- ▶ Backwards Induction

## Strategic Games

A **strategic games** is a tuple  $\langle N, \{A_i\}_{i \in N}, \{\succeq_i\}_{i \in N} \rangle$  where

- ▶  $N$  is a finite set of **players**

## Strategic Games

A **strategic games** is a tuple  $\langle N, \{A_i\}_{i \in N}, \{\succeq_i\}_{i \in N} \rangle$  where

- ▶  $N$  is a finite set of **players**
- ▶ for each  $i \in N$ ,  $A_i$  is a nonempty set of **actions**

## Strategic Games

A **strategic games** is a tuple  $\langle N, \{A_i\}_{i \in N}, \{\succsim_i\}_{i \in N} \rangle$  where

- ▶  $N$  is a finite set of **players**
- ▶ for each  $i \in N$ ,  $A_i$  is a nonempty set of **actions**
- ▶ for each  $i \in N$ ,  $\succsim_i$  is a **preference relation** on  $A = \prod_{i \in N} A_i$   
(Often  $\succsim_i$  are represented by **utility functions**  $u_i : A \rightarrow \mathbb{R}$ )

## Strategic Games: Comments on Preferences

- ▶ Preferences may be over a set of consequences  $C$ . Assume  $g : A \rightarrow C$  and  $\{\succeq_i^* \mid i \in N\}$  a set of preferences on  $C$ . Then for  $a, b \in A$ ,

$$a \succeq_i b \text{ iff } g(a) \succeq_i^* g(b)$$

- ▶ Consequences may be affected by exogenous random variable whose realization is not known before choosing actions. Let  $\Omega$  be a set of states, then define  $g : A \times \Omega \rightarrow C$ . Where  $g(a|\cdot)$  is interpreted as a *lottery*.
- ▶ Often  $\succeq_i$  are represented by **utility functions**  $u_i : A \rightarrow \mathbb{R}$

## Strategic Games: Example

		Column	
		r	l
Row	u		
	d		

- ▶  $N = \{Row, Column\}$
- ▶  $A_{Row} = \{u, d\}, A_{Column} = \{r, l\}$
- ▶  $(u, r) \succeq_{Row} (d, l) \succeq_{Row} (u, l) \sim_{Row} (d, r)$   
 $(u, r) \succeq_{Column} (d, l) \succeq_{Column} (u, l) \sim_{Column} (d, r)$



## Strategic Games: Example

		Column	
		r	l
Row	u	(2,2)	(0,0)
	d	(0,0)	(1,1)

- ▶  $N = \{Row, Column\}$
- ▶  $A_{Row} = \{u, d\}$ ,  $A_{Column} = \{r, l\}$
- ▶  $u_{Row} : A_{Row} \times A_{Column} \rightarrow \{0, 1, 2\}$ ,  
 $u_{Column} : A_{Row} \times A_{Column} \rightarrow \{0, 1, 2\}$  with  
 $u_{Row}(u, r) = u_{Column}(u, r) = 2$ ,  
 $u_{Row}(d, l) = u_{Column}(d, l) = 2$ ,  
and  $u_x(u, l) = u_x(d, r) = 0$  for  $x \in N$ .

## Nash Equilibrium

Let  $\langle N, \{A_i\}_{i \in N}, \{\succeq_i\}_{i \in N} \rangle$  be a strategic game

For  $a_{-i} \in A_{-i}$ , let

$$B_i(a_{-i}) = \{a_i \in A_i \mid (a_{-i}, a_i) \succeq_i (a_{-i}, a'_i) \forall a'_i \in A_i\}$$

$B_i$  is the **best-response** function.

## Nash Equilibrium

Let  $\langle N, \{A_i\}_{i \in N}, \{\succeq_i\}_{i \in N} \rangle$  be a strategic game

For  $a_{-i} \in A_{-i}$ , let

$$B_i(a_{-i}) = \{a_i \in A_i \mid (a_{-i}, a_i) \succeq_i (a_{-i}, a'_i) \forall a'_i \in A_i\}$$

$B_i$  is the **best-response** function.

$a^* \in A$  is a **Nash equilibrium** iff  $a_i^* \in B_i(a_{-i}^*)$  for all  $i \in N$ .

## Strategic Games Example: Bach or Stravinsky?

	$b_c$	$s_c$
$b_r$	2,1	0,0
$s_r$	0,0	1,2

## Strategic Games Example: Bach or Stravinsky?

	$b_c$	$s_c$
$b_r$	2,1	0,0
$s_r$	0,0	1,2

$$N = \{r, c\} \quad A_r = \{b_r, s_r\}, A_c = \{b_c, s_c\}$$

## Strategic Games Example: Bach or Stravinsky?

	$b_c$	$s_c$
$b_r$	2,1	0,0
$s_r$	0,0	1,2

$$N = \{r, c\} \quad A_r = \{b_r, s_r\}, A_c = \{b_c, s_c\}$$

$$B_r(b_c) = \{b_r\}$$

$$B_r(s_c) = \{s_r\}$$

## Strategic Games Example: Bach or Stravinsky?

	$b_c$	$s_c$
$b_r$	2,1	0,0
$s_r$	0,0	1,2

$$N = \{r, c\} \quad A_r = \{b_r, s_r\}, A_c = \{b_c, s_c\}$$

$$B_r(b_c) = \{b_r\}$$

$$B_r(s_c) = \{s_r\}$$

$$B_c(b_r) = \{b_c\}$$

$$B_c(s_r) = \{s_c\}$$

## Strategic Games Example: Bach or Stravinsky?

	$b_c$	$s_c$
$b_r$	2,1	0,0
$s_r$	0,0	1,2

$$N = \{r, c\} \quad A_r = \{b_r, s_r\}, A_c = \{b_c, s_c\}$$

$$B_r(b_c) = \{b_r\}$$

$$B_r(s_c) = \{s_r\}$$

$$B_c(b_r) = \{b_c\}$$

$$B_c(s_r) = \{s_c\}$$

$(b_r, b_c)$  is a Nash Equilibrium

$(s_r, s_c)$  is a Nash Equilibrium



## Another Example: Mozart or Mahler?

	<i>Mo</i>	<i>Ma</i>
<i>Mo</i>	2,2	0,0
<i>Ma</i>	0,0	1,1

## Another Example: Prisoner's Dilemma

	<i>D</i>	<i>C</i>
<i>D</i>	3,3	0,4
<i>C</i>	4,0	1,1

## Extensive Games

In strategic games, strategies are chosen *once and for all at the start of the game*

**Extensive games** are explicit descriptions of the **sequential structure** of the decision problem encountered by the players in a strategic situation.

## Extensive Games

- ▶ A set  $N$  of players
- ▶ A set  $H$  is a set of **sequences**, or **histories**, that is “closed” and contains all finite prefixes. I.e.,
  - The empty sequence is in  $H$
  - If  $(a_k)_{k=1,\dots,K} \in H$  and  $L < K$  then  $(a_k)_{k=1,\dots,L} \in H$
  - If an infinite sequence  $(a_k)_{k=1}^\infty$  satisfies  $(a_k)_{k=1,\dots,L} \in H$  for each  $L \geq 1$ , then  $(a_k)_{k=1}^\infty \in H$

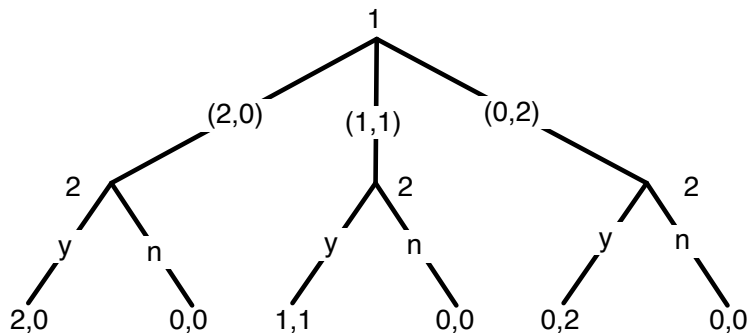
## Extensive Games

- ▶ Let  $Z \subseteq H$  be the set of **terminal** histories.
- ▶ A function  $P : H - Z \rightarrow N$
- ▶ For each  $i \in N$ , a relation  $\succeq_i$  on  $Z$ .
- ▶ For a nonterminal history  $h$ , let  $A(h) = \{a \mid (h, a) \in H\}$

# Extensive Games: Example

Suppose two people use the following procedure to share two desirable goods. One of them proposes an allocation, which the other accepts or rejects. In the event of rejection, neither person receives either of the objects. Each person cares only about the number of objects he obtains.

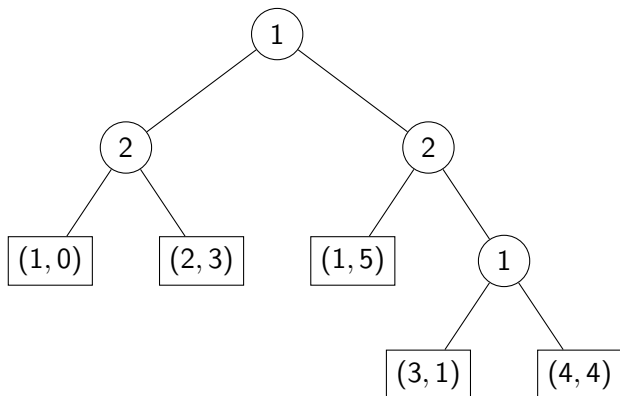
## Example

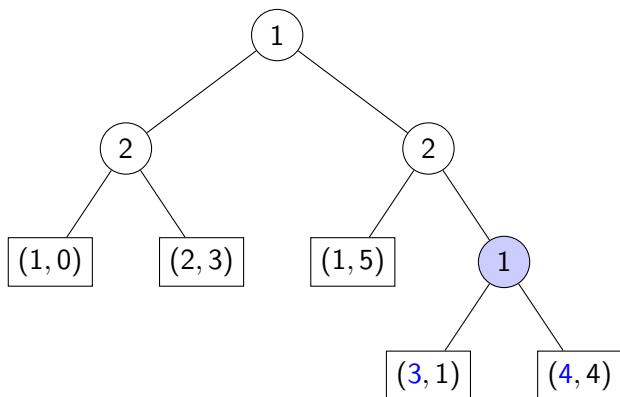


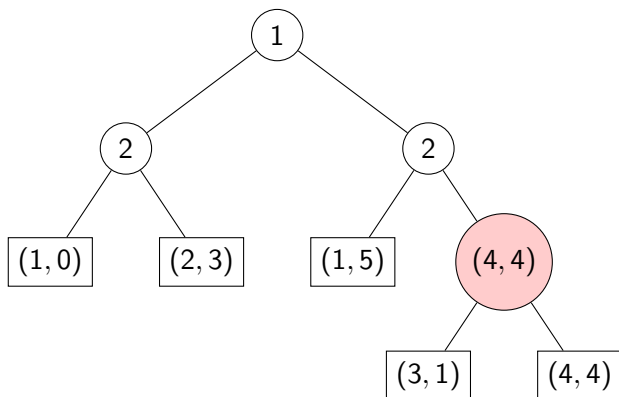
## Backwards Induction

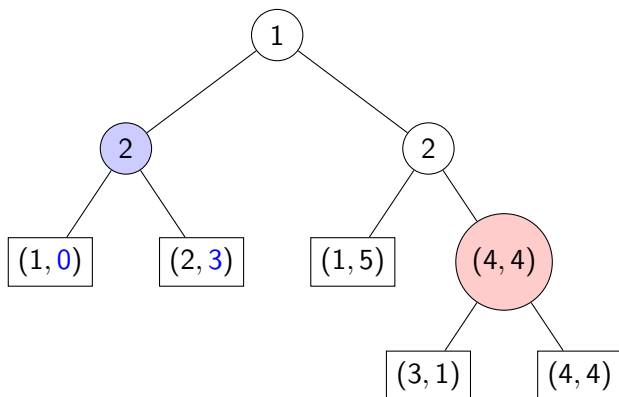
Invented by Zermelo, Backwards Induction is an iterative algorithm for “solving” an extensive game.

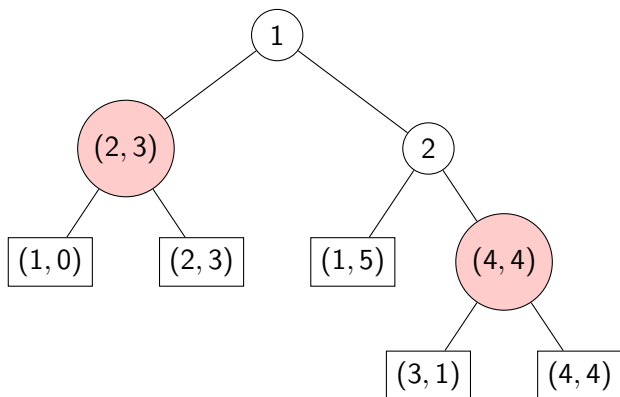


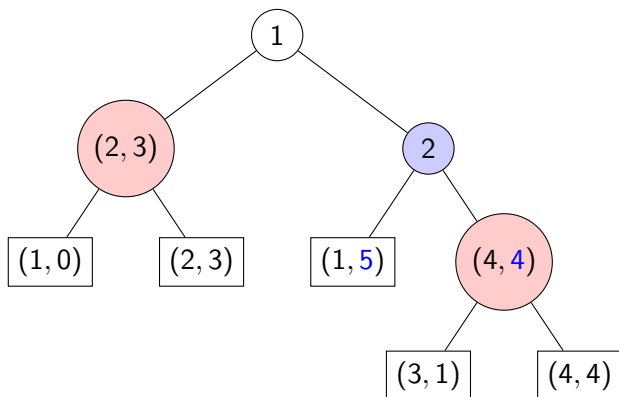


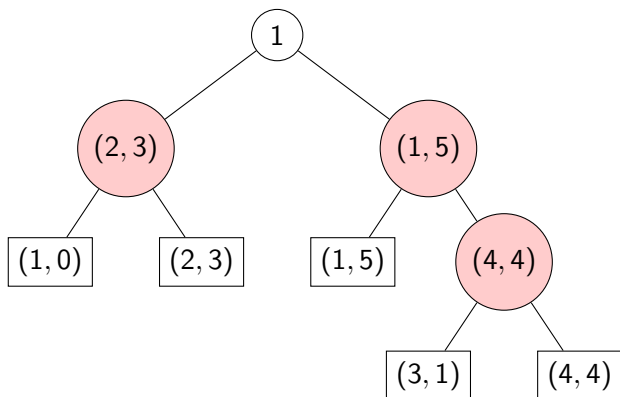


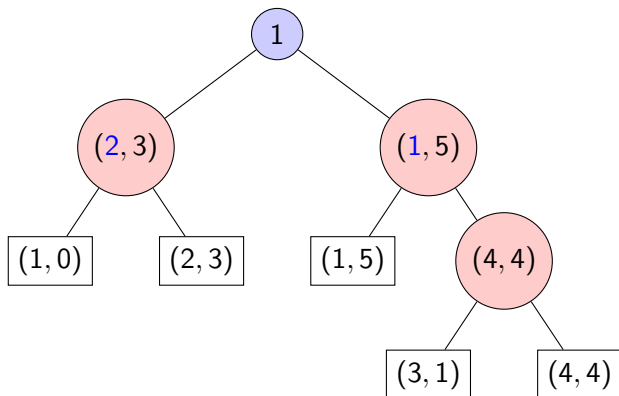




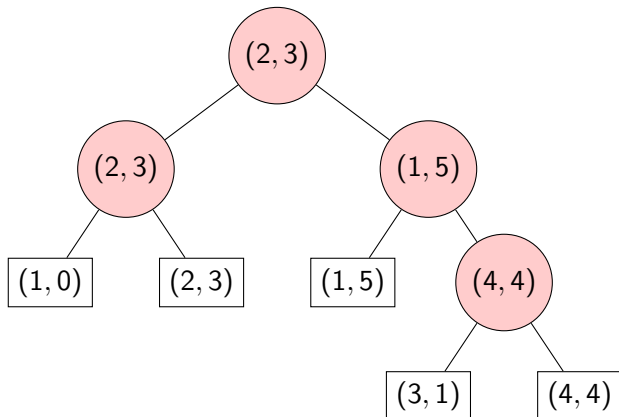


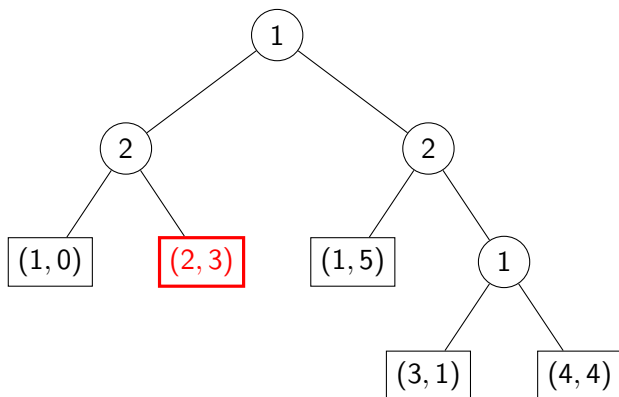


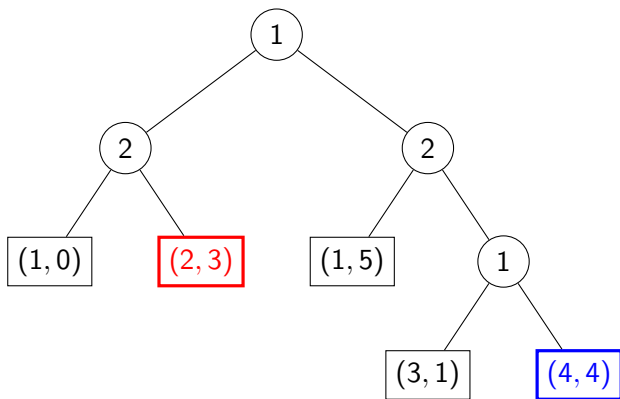












Is anything missing in these models?

*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, many 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, many 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, many 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, many 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?

To answer these questions, we need a (mathematical) framework to study each of the following issues:

- ▶ Rationality: “Ann is rational”
- ▶ Knowledge/Beliefs: “Bob believes (knows) Ann is rational”
- ▶ Higher-order Knowledge/Beliefs: “Ann knows that Bob knows that Ann is rational”, “it is common knowledge that all agents are rational”.

## Rationality

*...to understand the fundamental ideas of game theory, one should begin by studying decision theory. -R. Myerson (Game Theory)*

## Rationality

*...to understand the fundamental ideas of game theory, one should begin by studying decision theory.* -R. Myerson (*Game Theory*)

- ▶ Any rational decision-maker's behaviour should be describable by a utility function, which gives a quantitative characterization of his preference for outcomes and prizes and

## Rationality

*...to understand the fundamental ideas of game theory, one should begin by studying decision theory.* -R. Myerson (*Game Theory*)

- ▶ Any rational decision-maker's behaviour should be describable by a utility function, which gives a quantitative characterization of his preference for outcomes and prizes and
- ▶ a subjective probability distribution, which characterizes his beliefs about all relevant unknown factors.

## Rationality

*...to understand the fundamental ideas of game theory, one should begin by studying decision theory.* -R. Myerson (*Game Theory*)

- ▶ Any rational decision-maker's behaviour should be describable by a utility function, which gives a quantitative characterization of his preference for outcomes and prizes and
- ▶ a subjective probability distribution, which characterizes his beliefs about all relevant unknown factors.

We assume each agent **rational** in the sense that the agent **maximizes his expected utility given his current information**

(See Savage, Ramsey, Aumann & Anscombe, Jeffrey, etc.)

## Game Theory: Uncertainty

The players may be

- ▶ uncertain about the objective parameters of the environment
- ▶ imperfectly informed about events that happen in the game
- ▶ uncertain about action of the other players that are not deterministic
- ▶ uncertain about the reasoning of the other players

**Much more to discuss here!**



## Describing the Players Knowledge and Beliefs

Fix a set of possible states (**complete descriptions of a situation**). Two main approaches to describe beliefs (knowledge):

- ▶ Set-theoretical (Kripke Structures, Aumann Structures): For each state and each agent  $i$ , specify a set of states that  $i$  considers possible.
- ▶ Probabilistic (Bayesian Models, Harsanyi Type Spaces): For each state, define a (subjective) probability function over the set of states for each agent.

J. Hintikka. *Knowledge and Belief*. 1962, recently republished.

## Single-Agent Epistemic Logic: The Language

$\varphi$  is a formula of Epistemic Logic ( $\mathcal{L}$ ) if it is of the form

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi$$

## Single-Agent Epistemic Logic: The Language

$\varphi$  is a formula of Epistemic Logic ( $\mathcal{L}$ ) if it is of the form

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi$$

- ▶  $p \in \text{At}$  is an **atomic fact**.
  - “It is raining”
  - “The talk is at 2PM”
  - “The card on the table is a 7 of Hearts”

## Single-Agent Epistemic Logic: The Language

$\varphi$  is a formula of Epistemic Logic ( $\mathcal{L}$ ) if it is of the form

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi$$

- ▶  $p \in \text{At}$  is an **atomic fact**.
- ▶ The usual propositional language ( $\mathcal{L}_0$ )

## Single-Agent Epistemic Logic: The Language

$\varphi$  is a formula of Epistemic Logic ( $\mathcal{L}$ ) if it is of the form

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi$$

- ▶  $p \in \text{At}$  is an **atomic fact**.
- ▶ The usual propositional language ( $\mathcal{L}_0$ )
- ▶  $K\varphi$  is intended to mean “**The agent knows that  $\varphi$  is true**”.

## Single-Agent Epistemic Logic: The Language

$\varphi$  is a formula of Epistemic Logic ( $\mathcal{L}$ ) if it is of the form

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi$$

- ▶  $p \in \text{At}$  is an **atomic fact**.
- ▶ The usual propositional language ( $\mathcal{L}_0$ )
- ▶  $K\varphi$  is intended to mean “**The agent knows that  $\varphi$  is true**”.
- ▶ The usual definitions for  $\rightarrow, \vee, \leftrightarrow$  apply
- ▶ Define  $L\varphi$  as  $\neg K\neg\varphi$

## Single-Agent Epistemic Logic: The Language

$\varphi$  is a formula of Epistemic Logic ( $\mathcal{L}$ ) if it is of the form

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi$$

$K(p \rightarrow q)$ : “Ann knows that  $p$  implies  $q$ ”

$Kp \vee \neg Kp$ :

$Kp \vee K\neg p$ :

$L\varphi$ :

$KL\varphi$ :



## Single-Agent Epistemic Logic: The Language

$\varphi$  is a formula of Epistemic Logic ( $\mathcal{L}$ ) if it is of the form

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi$$

$K(p \rightarrow q)$ : “Ann knows that  $p$  implies  $q$ ”

$Kp \vee \neg Kp$ : “either Ann does or does not know  $p$ ”

$Kp \vee K\neg p$ : “Ann knows whether  $p$  is true”

$L\varphi$ :

$KL\varphi$ :

## Single-Agent Epistemic Logic: The Language

$\varphi$  is a formula of Epistemic Logic ( $\mathcal{L}$ ) if it is of the form

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi$$

$K(p \rightarrow q)$ : “Ann knows that  $p$  implies  $q$ ”

$Kp \vee \neg Kp$ : “either Ann does or does not know  $p$ ”

$Kp \vee K\neg p$ : “Ann knows whether  $p$  is true”

$L\varphi$ : “ $\varphi$  is an epistemic possibility”

$KL\varphi$ : “Ann knows that she thinks  $\varphi$  is possible”

## Single-Agent Epistemic Logic: Kripke Models

$$\mathcal{M} = \langle W, R, V \rangle$$

## Single-Agent Epistemic Logic: Kripke Models

$$\mathcal{M} = \langle W, R, V \rangle$$

- ▶  $W \neq \emptyset$  is the set of all relevant situations (states of affairs, possible worlds)

## Single-Agent Epistemic Logic: Kripke Models

$$\mathcal{M} = \langle W, R, V \rangle$$

- ▶  $W \neq \emptyset$  is the set of all relevant situations (states of affairs, possible worlds)
- ▶  $R \subseteq W \times W$  represents the information of the agent:  $wRv$  provided “ $w$  and  $v$  are epistemically indistinguishable”

## Single-Agent Epistemic Logic: Kripke Models

$$\mathcal{M} = \langle W, R, V \rangle$$

- ▶  $W \neq \emptyset$  is the set of all relevant situations (states of affairs, possible worlds)
- ▶  $R \subseteq W \times W$  represents the information of the agent:  $wRv$  provided “ $w$  and  $v$  are epistemically indistinguishable”
- ▶  $V : At \rightarrow \wp(W)$  is a **valuation function** assigning propositional variables to worlds

### Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

### Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

What are the relevant states?

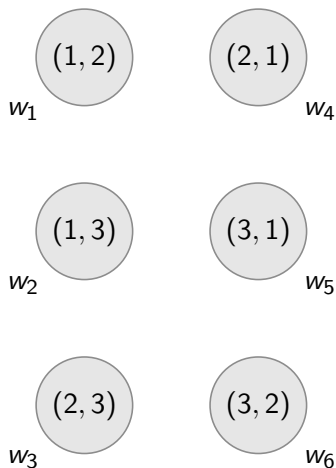


## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

What are the relevant states?

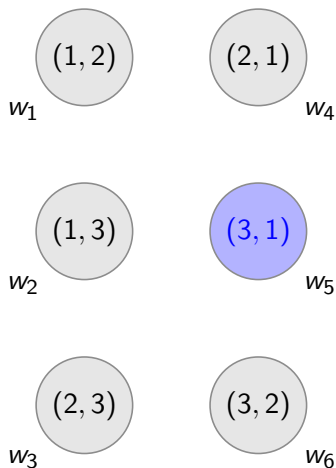


## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

Ann receives card 3 and card 1  
is put on the table

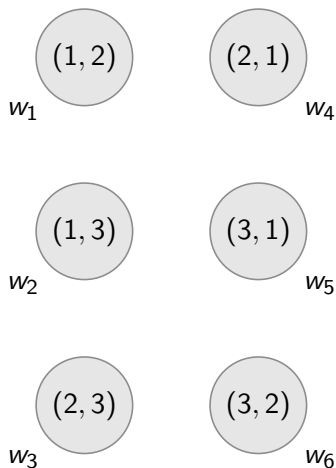


## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

What information does Ann  
have?

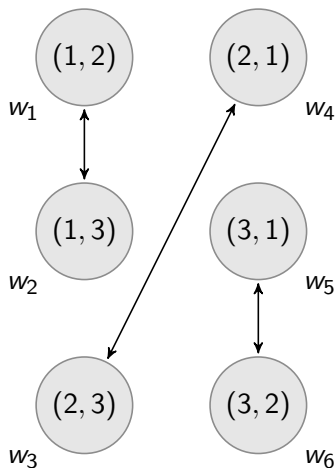


## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

What information does Ann  
have?

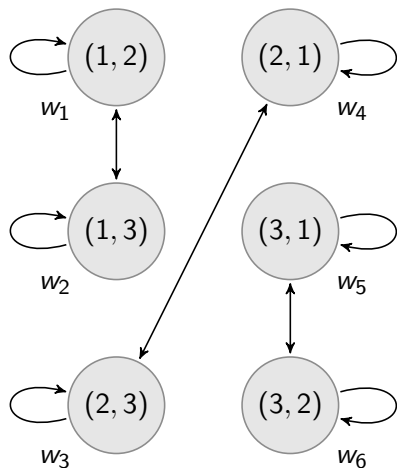


## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

What information does Ann  
have?



## Example

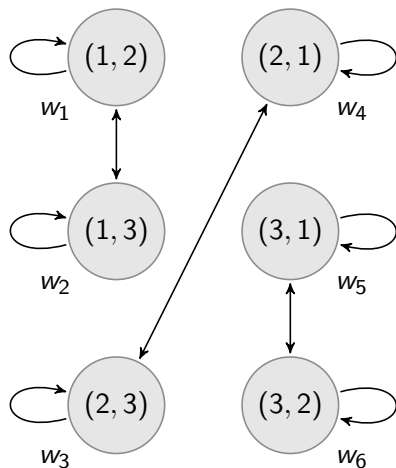
Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

Suppose  $H_i$  is intended to  
mean “Ann has card  $i$ ”

$T_i$  is intended to mean “card  $i$   
is on the table”

Eg.,  $V(H_1) = \{w_1, w_2\}$



## Example

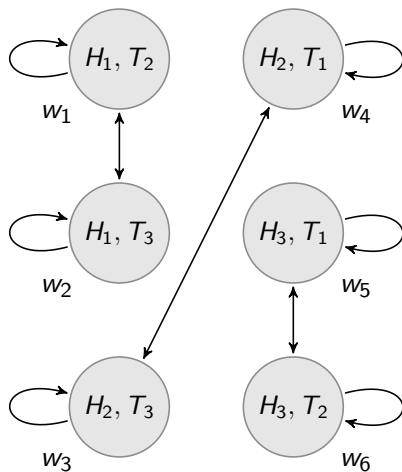
Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

Suppose  $H_i$  is intended to  
mean “Ann has card  $i$ ”

$T_i$  is intended to mean “card  $i$   
is on the table”

Eg.,  $V(H_1) = \{w_1, w_2\}$



## Single Agent Epistemic Logic: Truth in a Model

Given  $\varphi \in \mathcal{L}$ , a Kripke model  $\mathcal{M} = \langle W, R, V \rangle$  and  $w \in W$

$\mathcal{M}, w \models \varphi$  means “in  $\mathcal{M}$ , if the actual state is  $w$ , then  $\varphi$  is true”



## Single Agent Epistemic Logic: Truth in a Model

Given  $\varphi \in \mathcal{L}$ , a Kripke model  $\mathcal{M} = \langle W, R, V \rangle$  and  $w \in W$

$\mathcal{M}, w \models \varphi$  is defined as follows:

- ▶  $\mathcal{M}, w \models p$  iff  $w \in V(p)$  (with  $p \in \text{At}$ )
- ▶  $\mathcal{M}, w \models \neg\varphi$  if  $\mathcal{M}, w \not\models \varphi$
- ▶  $\mathcal{M}, w \models \varphi \wedge \psi$  if  $\mathcal{M}, w \models \varphi$  and  $\mathcal{M}, w \models \psi$
- ▶  $\mathcal{M}, w \models K\varphi$  if for each  $v \in W$ , if  $wRv$ , then  $\mathcal{M}, v \models \varphi$

## Single Agent Epistemic Logic: Truth in a Model

Given  $\varphi \in \mathcal{L}$ , a Kripke model  $\mathcal{M} = \langle W, R, V \rangle$  and  $w \in W$

$\mathcal{M}, w \models \varphi$  is defined as follows:

- ✓  $\mathcal{M}, w \models p$  iff  $w \in V(p)$  (with  $p \in \text{At}$ )
- ▶  $\mathcal{M}, w \models \neg\varphi$  if  $\mathcal{M}, w \not\models \varphi$
- ▶  $\mathcal{M}, w \models \varphi \wedge \psi$  if  $\mathcal{M}, w \models \varphi$  and  $\mathcal{M}, w \models \psi$
- ▶  $\mathcal{M}, w \models K\varphi$  if for each  $v \in W$ , if  $wRv$ , then  $\mathcal{M}, v \models \varphi$

## Single Agent Epistemic Logic: Truth in a Model

Given  $\varphi \in \mathcal{L}$ , a Kripke model  $\mathcal{M} = \langle W, R, V \rangle$  and  $w \in W$

$\mathcal{M}, w \models \varphi$  is defined as follows:

- ✓  $\mathcal{M}, w \models p$  iff  $w \in V(p)$  (with  $p \in \text{At}$ )
- ✓  $\mathcal{M}, w \models \neg\varphi$  if  $\mathcal{M}, w \not\models \varphi$
- ✓  $\mathcal{M}, w \models \varphi \wedge \psi$  if  $\mathcal{M}, w \models \varphi$  and  $\mathcal{M}, w \models \psi$
- ▶  $\mathcal{M}, w \models K\varphi$  if for each  $v \in W$ , if  $wRv$ , then  $\mathcal{M}, v \models \varphi$

## Single Agent Epistemic Logic: Truth in a Model

Given  $\varphi \in \mathcal{L}$ , a Kripke model  $\mathcal{M} = \langle W, R, V \rangle$  and  $w \in W$

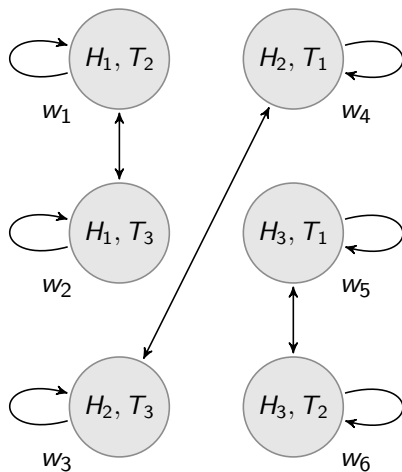
$\mathcal{M}, w \models \varphi$  is defined as follows:

- ✓  $\mathcal{M}, w \models p$  iff  $w \in V(p)$  (with  $p \in \text{At}$ )
- ✓  $\mathcal{M}, w \models \neg\varphi$  if  $\mathcal{M}, w \not\models \varphi$
- ✓  $\mathcal{M}, w \models \varphi \wedge \psi$  if  $\mathcal{M}, w \models \varphi$  and  $\mathcal{M}, w \models \psi$
- ✓  $\mathcal{M}, w \models K\varphi$  if for each  $v \in W$ , if  $wRv$ , then  $\mathcal{M}, v \models \varphi$

## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

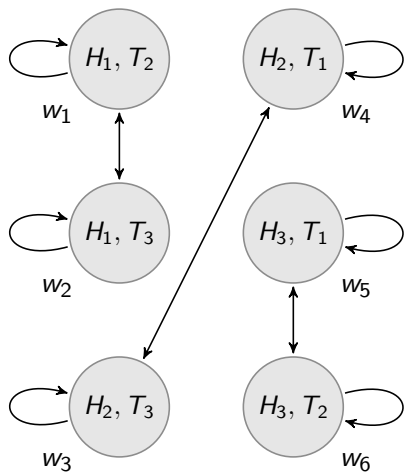


## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

Suppose that Ann receives card  
1 and card 2 is on the table.

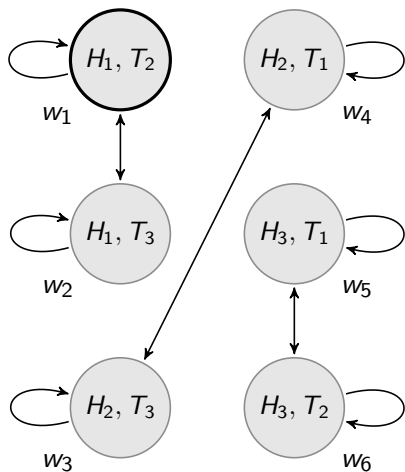


## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

Suppose that Ann receives card  
1 and card 2 is on the table.

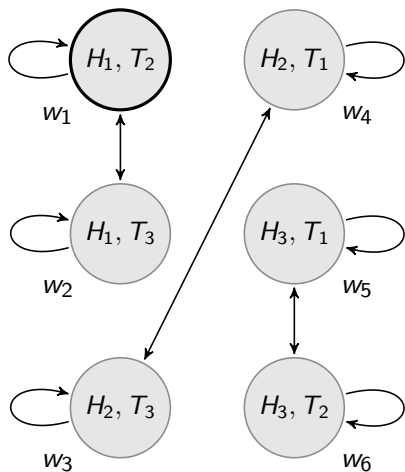


## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

$$\mathcal{M}, w_1 \models KH_1$$



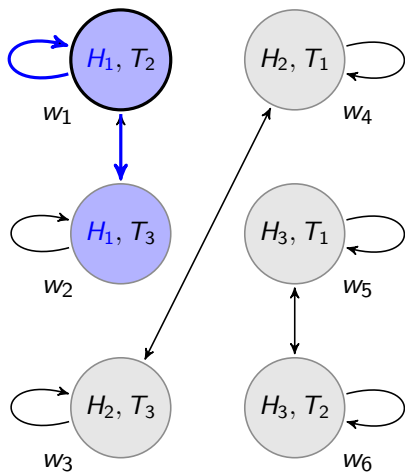


## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

$$\mathcal{M}, w_1 \models KH_1$$



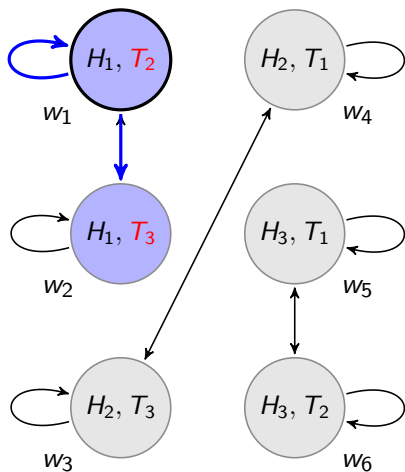
## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

$$\mathcal{M}, w_1 \models KH_1$$

$$\mathcal{M}, w_1 \models K\neg T_1$$

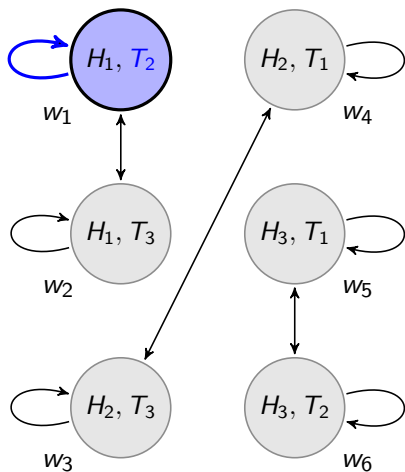


## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

$$\mathcal{M}, w_1 \models LT_2$$



## Some Questions

Should we make additional assumptions about  $R$  (i.e., reflexive, transitive, etc.)

What idealizations have we made?

## Some Notation

A **Kripke Frame** is a tuple  $\langle W, R \rangle$  where  $R \subseteq W \times W$ .

$\varphi$  is **valid in a Kripke model**  $\mathcal{M}$  if  $\mathcal{M}, w \models \varphi$  for all states  $w$  (we write  $\mathcal{M} \models \varphi$ ).

$\varphi$  is **valid on a Kripke frame**  $\mathcal{F}$  if  $\mathcal{M} \models \varphi$  for all models  $\mathcal{M}$  based on  $\mathcal{F}$ .

---

---

Modal Formula

Property

Philosophical Assumption

Modal Formula	Property	Philosophical Assumption
$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$	—	Logical Omniscience

Modal Formula	Property	Philosophical Assumption
$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$ $K\varphi \rightarrow \varphi$	<p>—</p> <p>Reflexive</p>	<p>Logical Omniscience</p> <p>Truth</p>



Modal Formula	Property	Philosophical Assumption
$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$ $K\varphi \rightarrow \varphi$ $K\varphi \rightarrow KK\varphi$	— Reflexive Transitive	Logical Omniscience Truth Positive Introspection

Modal Formula	Property	Philosophical Assumption
$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$	—	Logical Omniscience
$K\varphi \rightarrow \varphi$	Reflexive	Truth
$K\varphi \rightarrow KK\varphi$	Transitive	Positive Introspection
$\neg K\varphi \rightarrow K\neg K\varphi$	Euclidean	Negative Introspection

Modal Formula	Property	Philosophical Assumption
$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$ $K\varphi \rightarrow \varphi$ $K\varphi \rightarrow KK\varphi$ $\neg K\varphi \rightarrow K\neg K\varphi$ $\neg K\perp$	— Reflexive Transitive Euclidean Serial	Logical Omniscience Truth Positive Introspection Negative Introspection Consistency

## Multi-agent Epistemic Logic

**The Language:**  $\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi$

**Kripke Models:**  $\mathcal{M} = \langle W, R, V \rangle$  and  $w \in W$

**Truth:**  $\mathcal{M}, w \models \varphi$  is defined as follows:

- ▶  $\mathcal{M}, w \models p$  iff  $w \in V(p)$  (with  $p \in \text{At}$ )
- ▶  $\mathcal{M}, w \models \neg\varphi$  if  $\mathcal{M}, w \not\models \varphi$
- ▶  $\mathcal{M}, w \models \varphi \wedge \psi$  if  $\mathcal{M}, w \models \varphi$  and  $\mathcal{M}, w \models \psi$
- ▶  $\mathcal{M}, w \models K\varphi$  if for each  $v \in W$ , if  $wRv$ , then  $\mathcal{M}, v \models \varphi$

## Multi-agent Epistemic Logic

**The Language:**  $\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_i\varphi$  with  $i \in \mathcal{A}$

**Kripke Models:**  $\mathcal{M} = \langle W, \{R_i\}_{i \in \mathcal{A}}, V \rangle$  and  $w \in W$

**Truth:**  $\mathcal{M}, w \models \varphi$  is defined as follows:

- ▶  $\mathcal{M}, w \models p$  iff  $w \in V(p)$  (with  $p \in \text{At}$ )
- ▶  $\mathcal{M}, w \models \neg\varphi$  if  $\mathcal{M}, w \not\models \varphi$
- ▶  $\mathcal{M}, w \models \varphi \wedge \psi$  if  $\mathcal{M}, w \models \varphi$  and  $\mathcal{M}, w \models \psi$
- ▶  $\mathcal{M}, w \models K_i\varphi$  if for each  $v \in W$ , if  $wR_iv$ , then  $\mathcal{M}, v \models \varphi$

## Multi-agent Epistemic Logic

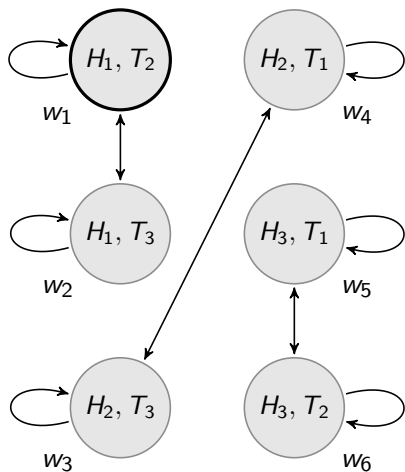
- ▶  $K_A K_B \varphi$ : “Ann knows that Bob knows  $\varphi$ ”
- ▶  $K_A (K_B \varphi \vee K_B \neg \varphi)$ : “Ann knows that Bob knows whether  $\varphi$ ”
- ▶  $\neg K_B K_A K_B (\varphi)$ : “Bob does not know that Ann knows that Bob knows that  $\varphi$ ”

## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
one of the cards is placed face  
down on the table and the third  
card is put back in the deck.

Suppose that Ann receives card  
1 and card 2 is on the table.

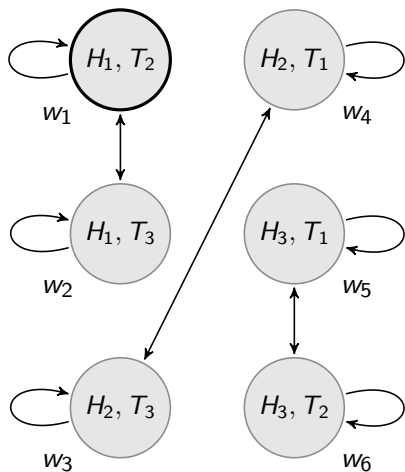


## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
Bob is given one of the cards  
and the third card is put back  
in the deck.

Suppose that Ann receives card  
1 and Bob receives card 2.



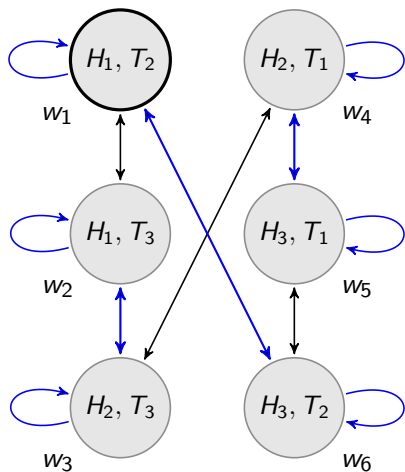


## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
Bob is given one of the cards  
and the third card is put back  
in the deck.

Suppose that Ann receives card  
1 and Bob receives card 2.

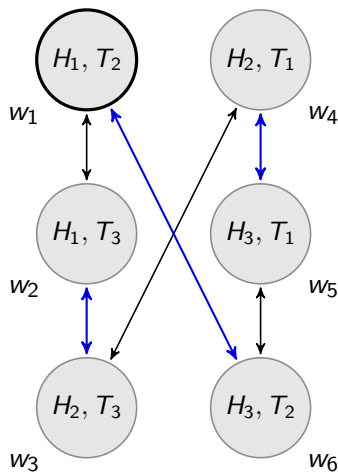


## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
Bob is given one of the cards  
and the third card is put back  
in the deck.

Suppose that Ann receives card  
1 and Bob receives card 2.



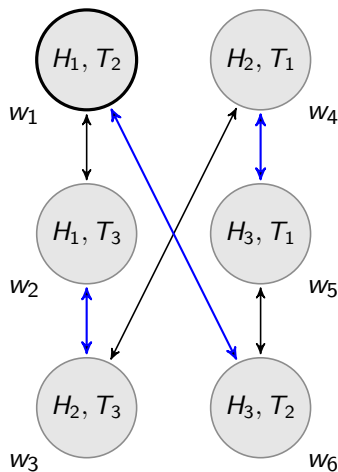
## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
Bob is given one of the cards  
and the third card is put back  
in the deck.

Suppose that Ann receives card  
1 and Bob receives card 2.

$$\mathcal{M}, w \models K_B(K_A H_1 \vee K_A \neg H_1)$$



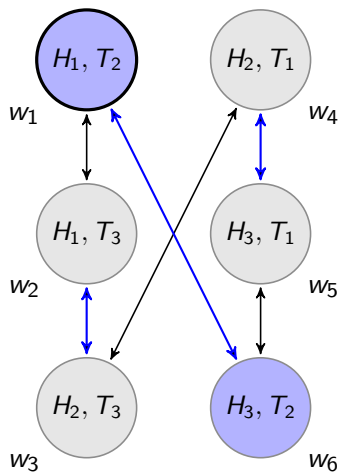
## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
Bob is given one of the cards  
and the third card is put back  
in the deck.

Suppose that Ann receives card  
1 and Bob receives card 2.

$$\mathcal{M}, w \models K_B(K_A H_1 \vee K_A \neg H_1)$$



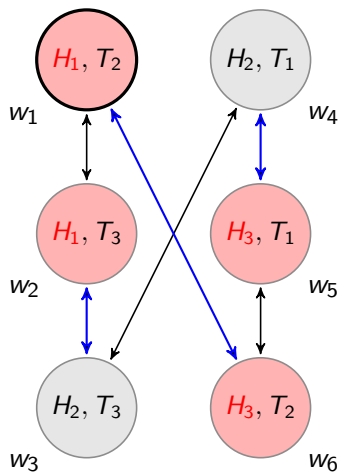
## Example

Suppose there are three cards:  
1, 2 and 3.

Ann is dealt one of the cards,  
Bob is given one of the cards  
and the third card is put back  
in the deck.

Suppose that Ann receives card  
1 and Bob receives card 2.

$$\mathcal{M}, w \models K_B(K_A H_1 \vee K_A \neg H_1)$$



## Example (1)

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.

## Three Views of Common Knowledge

1.  $\gamma := i$  knows that  $\varphi$ ,  $j$  knows that  $\varphi$ ,  $i$  knows that  $j$  knows that  $\varphi$ ,  $j$  knows that  $i$  knows that  $\varphi$ ,  $i$  knows that  $j$  knows that  $i$  knows that  $\varphi$ , ...

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

## Three Views of Common Knowledge

1.  $\gamma := i$  knows that  $\varphi$ ,  $j$  knows that  $\varphi$ ,  $i$  knows that  $j$  knows that  $\varphi$ ,  $j$  knows that  $i$  knows that  $\varphi$ ,  $i$  knows that  $j$  knows that  $i$  knows that  $\varphi$ , ...

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

2.  $\gamma := i$  and  $j$  know that ( $\varphi$  and  $\gamma$ )

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



## Three Views of Common Knowledge

1.  $\gamma := i$  knows that  $\varphi$ ,  $j$  knows that  $\varphi$ ,  $i$  knows that  $j$  knows that  $\varphi$ ,  $j$  knows that  $i$  knows that  $\varphi$ ,  $i$  knows that  $j$  knows that  $i$  knows that  $\varphi$ , ...

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

2.  $\gamma := i$  and  $j$  know that ( $\varphi$  and  $\gamma$ )

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

3. There is a *shared situation*  $s$  such that

- $s$  entails  $\varphi$
- $s$  entails  $i$  knows  $\varphi$
- $s$  entails  $j$  knows  $\varphi$

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

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

G. Sillari and P. Vanderscraaf. *Common Knowledge*. Stanford Encyclopedia of Philosophy Entry.

R. Cubitt and R. Sugden. *Common knowledge, salience and convention: a reconstruction of David Lewis' game theory*. *Economics and Philosophy* **19** (2003) pgs 175 - 210.

## Group Knowledge

$K_A P$ : “Ann knows that  $P$ ”

## Group Knowledge

$K_A P$ : “Ann knows that  $P$ ”

$K_B P$ : “Bob knows that  $P$ ”

## Group Knowledge

$K_A P$ : “Ann knows that  $P$ ”

$K_B P$ : “Bob knows that  $P$ ”

$K_A K_B P$ : “Ann knows that Bob knows that  $P$ ”

## Group Knowledge

$K_A P$ : “Ann knows that  $P$ ”

$K_B P$ : “Bob knows that  $P$ ”

$K_A K_B P$ : “Ann knows that Bob knows that  $P$ ”

$K_A P \wedge K_B P$ : “Every one knows  $P$ ”.

## Group Knowledge

$K_A P$ : “Ann knows that  $P$ ”

$K_B P$ : “Bob knows that  $P$ ”

$K_A K_B P$ : “Ann knows that Bob knows that  $P$ ”

$K_A P \wedge K_B P$ : “Every one knows  $P$ ”. let  $EP := K_A P \wedge K_B P$

## Group Knowledge

$K_A P$ : “Ann knows that  $P$ ”

$K_B P$ : “Bob knows that  $P$ ”

$K_A K_B P$ : “Ann knows that Bob knows that  $P$ ”

$K_A P \wedge K_B P$ : “Every one knows  $P$ ”. let  $EP := K_A P \wedge K_B P$

$K_A EP$ : “Ann knows that everyone knows that  $P$ ”.



## Group Knowledge

$K_A P$ : “Ann knows that  $P$ ”

$K_B P$ : “Bob knows that  $P$ ”

$K_A K_B P$ : “Ann knows that Bob knows that  $P$ ”

$K_A P \wedge K_B P$ : “Every one knows  $P$ ”. let  $EP := K_A P \wedge K_B P$

$K_A EP$ : “Ann knows that everyone knows that  $P$ ”.

$EEP$ : “Everyone knows that everyone knows that  $P$ ”.

## Group Knowledge

$K_A P$ : “Ann knows that  $P$ ”

$K_B P$ : “Bob knows that  $P$ ”

$K_A K_B P$ : “Ann knows that Bob knows that  $P$ ”

$K_A P \wedge K_B P$ : “Every one knows  $P$ ”. let  $EP := K_A P \wedge K_B P$

$K_A EP$ : “Ann knows that everyone knows that  $P$ ”.

$EEP$ : “Everyone knows that everyone knows that  $P$ ”.

$EEEP$ : “Everyone knows that everyone knows that everyone knows that  $P$ .”

## Common Knowledge

*CP*: "It is common knowledge that  $P$ "

## Common Knowledge

*CP*: “It is **common knowledge** that  $P$ ” — “Everyone knows that everyone knows that everyone knows that  $\dots P$ ”.

## Common Knowledge

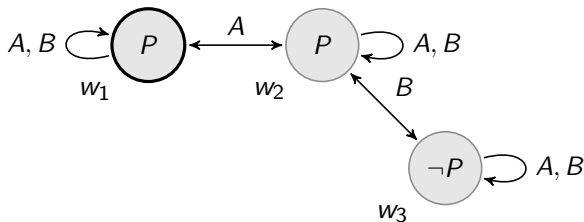
*CP*: “It is **common knowledge** that  $P$ ” — “Everyone knows that everyone knows that everyone knows that  $\dots P$ ”.

Is *common knowledge* different from *everyone knows*?

## Common Knowledge

$CP$ : “It is **common knowledge** that  $P$ ” — “Everyone knows that everyone knows that everyone knows that  $\dots P$ ”.

Is *common knowledge* different from *everyone knows*?

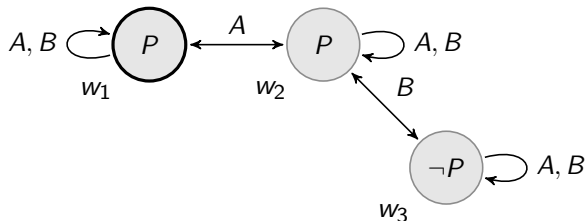


$$w_1 \models EP \wedge \neg CP$$

## Common Knowledge

*CP*: “It is **common knowledge** that  $P$ ” — “Everyone knows that everyone knows that everyone knows that  $\dots P$ ”.

Is *common knowledge* different from *everyone knows*?

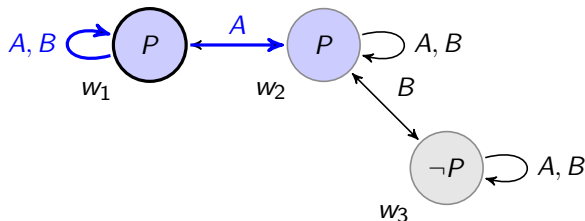


$$w_1 \models EP \wedge \neg CP$$

## Common Knowledge

$CP$ : “It is **common knowledge** that  $P$ ” — “Everyone knows that everyone knows that everyone knows that  $\dots P$ ”.

Is *common knowledge* different from *everyone knows*?



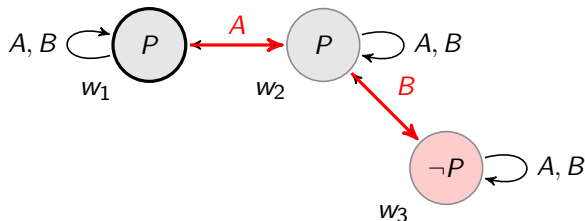
$$w_1 \models EP \wedge \neg CP$$



## Common Knowledge

*CP*: “It is **common knowledge** that  $P$ ” — “Everyone knows that everyone knows that everyone knows that  $\dots P$ ”.

Is *common knowledge* different from *everyone knows*?



$$w_1 \models EP \wedge \neg CP$$

## Common Knowledge

The operator “everyone knows  $P$ ”, denoted  $EP$ , is defined as follows

$$EP := \bigwedge_{i \in \mathcal{A}} K_i P$$

$w \models CP$  iff every finite path starting at  $w$  ends with a state satisfying  $P$ .

$$CP \rightarrow ECP$$

$$CP \rightarrow ECP$$

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  $P$  — is common knowledge if and only if some event — call it  $Q$  — happened that entails  $P$  and also entails all players’ knowing  $Q$  (like all players met Ann and Bob at an intimate party). (*Robert Aumann*)

$$P \wedge C(P \rightarrow EP) \rightarrow CP$$

## An Example

Two players Ann and Bob are told that the following will happen. Some positive integer  $n$  will be chosen and *one* of  $n$ ,  $n + 1$  will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

## An Example

Two players Ann and Bob are told that the following will happen. Some positive integer  $n$  will be chosen and *one* of  $n$ ,  $n + 1$  will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

Suppose the number are (2,3).

## An Example

Two players Ann and Bob are told that the following will happen. Some positive integer  $n$  will be chosen and *one* of  $n$ ,  $n + 1$  will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

Suppose the number are (2,3).

Do the agents know there numbers are less than 1000?



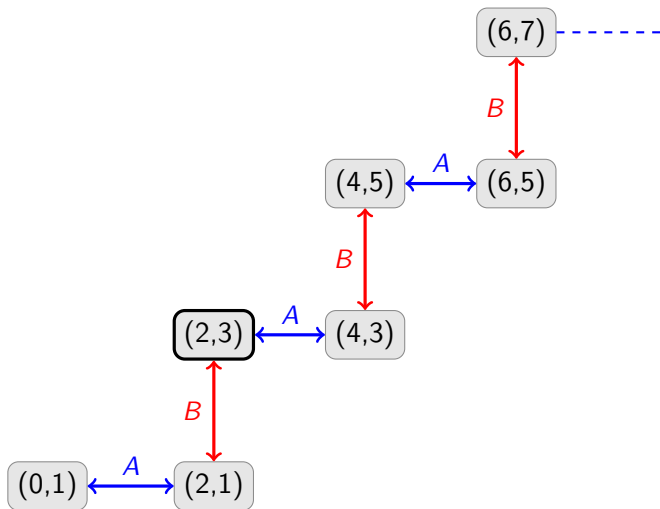
## An Example

Two players Ann and Bob are told that the following will happen. Some positive integer  $n$  will be chosen and *one* of  $n$ ,  $n + 1$  will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

Suppose the numbers are (2,3).

Do the agents know their numbers are less than 1000?

Is it common knowledge that their numbers are less than 1000?



## Epistemic Foundations of Game Theory

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

K. Binmore. *Rational Decisions*. Princeton University Press, 2009.

J. van Benthem. *Logic in Games*. Texts in Logic and Games, University of Amsterdam Press (forthcoming in 2010).

EP and O. Roy. *Interactive Rationality*. Book manuscript (forthcoming).

## 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?

### Reason to Believe

$R_i(\varphi)$ : “agent  $i$  has reason to believe  $\varphi$ ” (deductive, inductive, norm of practical reason)

## Reason to Believe

$R_i(\varphi)$ : “agent  $i$  has reason to believe  $\varphi$ ” (deductive, inductive, norm of practical reason)

If  $i$  is the **subject of the proposition**  $\varphi_i$ , consider

$$R_i(\varphi_i) \text{ vs. } R_j(\varphi_i)$$

## Reason to Believe

$R_i(\varphi)$ : “agent  $i$  has reason to believe  $\varphi$ ” (deductive, inductive, norm of practical reason)

If  $i$  is the **subject of the proposition**  $\varphi_i$ , consider

$$R_i(\varphi_i) \text{ vs. } R_j(\varphi_i)$$

$$\text{inf}(R) : x, y \rightarrow z$$

$$\text{inf}(R) : x_1, \dots, x_n, \neg(x_1, \dots, x_n, \neg z) \rightarrow z$$

## Common Reason to Believe

*Awareness of Common Reason:* for all  $i \in N$  and all propositions  $x$ ,

$$R^N(x) \Rightarrow R_i[R^N(x)]$$

## Common Reason to Believe

*Awareness of Common Reason:* for all  $i \in N$  and all propositions  $x$ ,

$$R^N(x) \Rightarrow R_i[R^N(x)]$$

*Authority of Common Reason:* for all  $i \in N$  and all propositions  $x$  for which  $i$  is not the subject

$$\text{inf}(R_i) : R^N(x) \rightarrow x$$

## Common Reason to Believe

*Awareness of Common Reason:* for all  $i \in N$  and all propositions  $x$ ,

$$R^N(x) \Rightarrow R_i[R^N(x)]$$

*Authority of Common Reason:* for all  $i \in N$  and all propositions  $x$  for which  $i$  is not the subject

$$\text{inf}(R_i) : R^N(x) \rightarrow x$$

*Common Attribution of Common Reason:* for all  $i \in N$ , for all propositions  $x$  for which  $i$  is not the subject

$$\text{inf}(R^N) : x \rightarrow R_i(x)$$



## Common Reason to Believe to Common Belief

**Theorem** The three previous properties can generate any hierarchy of belief ( $i$  has reason to believe that  $j$  has reason to believe that... that  $x$ ) for any  $x$  with  $R^N(x)$ .

## Team Maximising

$inf(R_i) : R^N[opt(v, N, s^N)],$

$R^N[ \text{each } i \in N \text{ endorses team maximising with respect to } N \text{ and } v ],$

$R^N[ \text{each member of } N \text{ acts on reasons } ] \rightarrow ought(i, s_i)$

## Team Maximising

$inf(R_i) : R^N[opt(v, N, s^N)],$

$R^N[ \text{each } i \in N \text{ endorses team maximising with respect to } N \text{ and } v ],$

$R^N[ \text{each member of } N \text{ acts on reasons } ] \rightarrow \textit{ought}(i, s_i)$

$R_i[\textit{ought}(i, s_i)]: i \text{ has reason to choose } s_i$

## Team Maximising

$inf(R_i) : R^N[opt(v, N, s^N)],$

$R^N[ \text{each } i \in N \text{ endorses team maximising with respect to } N \text{ and } v ],$

$R^N[ \text{each member of } N \text{ acts on reasons } ] \rightarrow ought(i, s_i)$

$i \text{ acts on reasons if for all } s_i, R[ought(i, s_i)] \Rightarrow choice(i, s_i)$

## Team Maximising

$inf(R_i) : R^N[opt(v, N, s^N)],$

$R^N[ \text{each } i \in N \text{ endorses team maximising with respect to } N \text{ and } v ],$

$R^N[ \text{each member of } N \text{ acts on reasons } ] \rightarrow ought(i, s_i)$

$opt(v, N, s^N): s^N$  is maximal for the group  $N$  w.r.t.  $v$

## Team Maximising

$inf(R_i) : R^N[opt(v, N, s^N)],$   
 $R^N[ \text{each } i \in N \text{ endorses team maximising with respect to } N \text{ and } v ],$   
 $R^N[ \text{each member of } N \text{ acts on reasons } ] \rightarrow ought(i, s_i)$

---

Thank you!