Interactive Epistemology

Eric Pacuit

Stanford University ai.stanford.edu/~epacuit

April 23, 2009

Fundamental Question: What does it mean to say that the players in a strategic interactive situation are

Fundamental Question: What does it mean to say that the players in a strategic interactive situation are rational,

Fundamental Question: What does it mean to say that the players in a strategic interactive situation are rational, each thinks each other is rational,

Fundamental Question: What does it mean to say that the players in a strategic interactive situation are rational, each thinks each other is rational, each thinks each other thinks the others are rational,

Fundamental Question: What does it mean to say that the players in a strategic interactive situation are rational, each thinks each other is rational, each thinks each other thinks the others are rational, and so on?

Fundamental Question: What does it mean to say that the players in a strategic interactive situation are rational, each thinks each other is rational, each thinks each other thinks the others are rational, and so on?

Fundamental Question: What does it mean to say that the players in a strategic interactive situation are rational, each thinks each other is rational, each thinks each other thinks the others are rational, and so on?

Goal for this talk: explore frameworks (logical and probabilistic) that attempt to answer this question.

Plan for Today

Keep in mind: *rational outcome* vs. *being rational*

- Introduction and Motivation
  1. Games
  2. Models of Information in Games. Type Structures and Epistemic Plausibility models
  3. Rationality

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 talk, **solution concepts** are more of an *endpoint*.

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 $\langle N, \{A_i\}_{i \in N}, \{\succeq_i\}_{i \in N} \rangle$ where

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

Strategic Games: Comments on Preferences

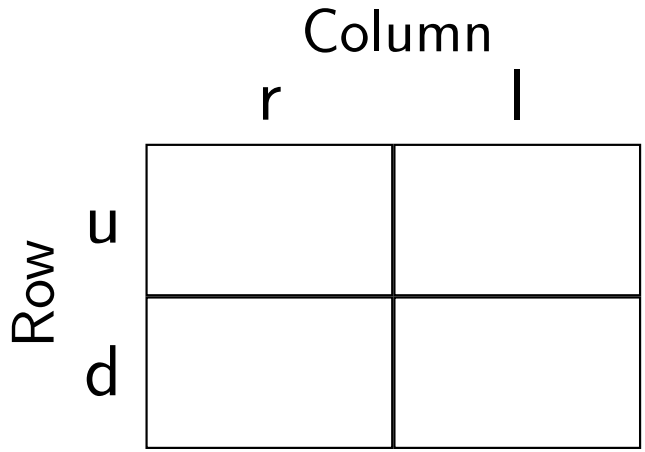
- Preferences may be over a set of consequences C. Assume $g : A \to C$ and $\{\succeq_i^* \mid i \in N\}$ a set of preferences on C. Then for $a, b \in A$,

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

- Consequences may be affected by exogenous random variable whose realization is not known before choosing actions. Let Ω be a set of states, then define $g : A \times \Omega \to C$. Where $g(a|\cdot)$ is interpreted as a *lottery*.

- Often $\succeq_i$ are represented by **utility functions** $u_i : A \to \mathbb{R}$

Strategic Games: Example



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

Strategic Games: Example

		Column: r	Column: 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

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

$$B_r(b_c) = \{b_r\} \qquad B_r(s_c) = \{s_r\}$$

Strategic Games Example: Bach or Stravinsky?

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

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

$$B_r(b_c) = \{b_r\} \qquad B_r(s_c) = \{s_r\}$$

$$B_c(b_r) = \{b_c\} \qquad B_c(s_r) = \{s_c\}$$

Strategic Games Example: Bach or Stravinsky?

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

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

$$B_r(b_c) = \{b_r\} \qquad B_r(s_c) = \{s_r\}$$

$$B_c(b_r) = \{b_c\} \qquad B_c(s_r) = \{s_c\}$$

(b_r, b_c) is a Nash Equilibrium (s_r, s_c) is a Nash Equilibrium

Other Game Forms

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

Other Game Forms

In strategic games, strategies are chosen *once and for all at the start of the game*, but usually there is more structure to a strategic interactive situation.

1. Other game forms: extensive games (represents the sequential nature of many situations), stochastic games, etc.
2. Infinite/finite number of actions
3. ....

Returning to the Conceptual Discussion

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

Returning to the Conceptual Discussion

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

What does it mean for Ann to **be rational**? What is the rational thing for Ann to do?

- It depends on what she *expects* Bob to do.
- But this depends on what she thinks Bob expects her to do.
- And so on...

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".

Information in games

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

Information in games

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.

Information in games

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.

Information in games

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.

Information in games

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. *When all is said and done, how should you play and what should you expect?*. Center for the Study of Rationality, 2005.

Information in games

- Various states of information disclosure.

Information in games

- Various states of information disclosure.
  - *Ex ante*, *ex interim*, *ex post*

Information in games

- Various states of information disclosure.
  - *Ex ante*, *ex interim*, *ex post*
- Various "types" of information:

Information in games

- Various states of information disclosure.
  - *Ex ante*, *ex interim*, *ex post*
- Various "types" of information:
  - (hard information) own preferences, own beliefs, structure of the game, (soft information) what the other agent will do, etc.

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-theortical (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.

Two Models

1. Harsanyi Type Space
2. Partition Model (Kripke Structure)

Harsanyi Type Space

John C. Harsanyi, nobel prize winner in economics, developed a theory of games with **incomplete information**.

J. Harsanyi. *Games with incomplete information played by "bayesian" players I-III. Management Science Theory* **14***: 159-182, 1967-68.*

Harsanyi Type Space

John C. Harsanyi, nobel prize winner in economics, developed a theory of games with **incomplete information**.

1. incomplete information: uncertainty about the *structure* of the game (outcomes, payoffs, strategy space)
2. imperfect information: uncertainty *within the game* about the previous moves of the players

J. Harsanyi. *Games with incomplete information played by "bayesian" players I-III. Management Science Theory* **14***: 159-182, 1967-68.*

Harsanyi Type Space

The problem: A natural question following any game-theoretic analysis is

Harsanyi Type Space

The problem: A natural question following any game-theoretic analysis is *how would the players react if some parameters of the model are not known to the players*?

Harsanyi Type Space

The problem: A natural question following any game-theoretic analysis is *how would the players react if some parameters of the model are not known to the players*? How do we completely specify such a model?

Harsanyi Type Space

The problem: A natural question following any game-theoretic analysis is *how would the players react if some parameters of the model are not known to the players*? How do we completely specify such a model?

1. Suppose there is a parameter that some player i does not know

Harsanyi Type Space

The problem: A natural question following any game-theoretic analysis is *how would the players react if some parameters of the model are not known to the players*? How do we completely specify such a model?

1. Suppose there is a parameter that some player i does not know
2. i's uncertainty about the parameter must be included in the model (first-order beliefs)

Harsanyi Type Space

The problem: A natural question following any game-theoretic analysis is *how would the players react if some parameters of the model are not known to the players*? How do we completely specify such a model?

1. Suppose there is a parameter that some player i does not know
2. i's uncertainty about the parameter must be included in the model (first-order beliefs)
3. this is a new parameter that the other players may not know, so we must specify the players beliefs about this parameter (second-order beliefs)

Harsanyi Type Space

The problem: A natural question following any game-theoretic analysis is *how would the players react if some parameters of the model are not known to the players*? How do we completely specify such a model?

1. Suppose there is a parameter that some player i does not know
2. i's uncertainty about the parameter must be included in the model (first-order beliefs)
3. this is a new parameter that the other players may not know, so we must specify the players beliefs about this parameter (second-order beliefs)
4. but this is a new parameter, and so on....

Harsanyi Type Space

A (game-theoretic) **type** of a player summarizes everything the player knows privately at the beginning of the game which could affect his beliefs about payoffs in the game and about all other players' types.

(Harsanyi argued that all uncertainty in a game can be equivalently modeled as uncertainty about payoff functions.)

Harsanyi Type Space: The Basic Model

$$\mathcal{T} = \langle \mathcal{A}, S, \{T_i\}_{\in \mathcal{A}}, \{\lambda_i\}_{i \in \mathcal{A}} \rangle$$

- $\mathcal{A}$ is a finite set of n agents
- S is the uncertainty domain
- T_i is a set of types
- $\lambda_i : T_i \rightarrow \Delta(S \times T_{-i})$

A state of the world is a tuple

$$(s, t_1, \ldots, t_n) \in S \times T_1 \times \cdots \times T_n$$

Example

$$T_1 = \{t_1, t_1'\},\ T_2 = \{t_2, t_2'\},\ S = \{a, b\}$$

Player 1: $\lambda_1(t_1)$

	a	b
t_2	1	0
t_2'	0	0

$\lambda_1(t_1')$

	a	b
t_2	0	0
t_2'	0.3	0.7

Player 2: $\lambda_2(t_2)$

	a	b
t_1	0	0.5
t_1'	0.5	0

$\lambda_2(t_2')$

	a	b
t_1	0	0
t_1'	0	1

Example

$$T_1 = \{t_1, t_1'\},\ T_2 = \{t_2, t_2'\},\ S = \{a, b\}$$

Player 1: $\lambda_1(t_1)$

	a	b
t_2	1	0
t_2'	0	0

$\lambda_1(t_1')$

	a	b
t_2	0	0
t_2'	0.3	0.7

Player 2: $\lambda_2(t_2)$

	a	b
t_1	0	0.5
t_1'	0.5	0

$\lambda_2(t_2')$

	a	b
t_1	0	0
t_1'	0	1

t_1 is **certain** the outcome is a $(o = a)$.

Example

$$T_1 = \{t_1, t_1'\},\ T_2 = \{t_2, t_2'\},\ S = \{a, b\}$$

Player 1: $\lambda_1(t_1)$

	a	b
t_2	1	0
t_2'	0	0

$\lambda_1(t_1')$

	a	b
t_2	0	0
t_2'	0.3	0.7

Player 2: $\lambda_2(t_2)$

	a	b
t_1	0	0.5
t_1'	0.5	0

$\lambda_2(t_2')$

	a	b
t_1	0	0
t_1'	0	1

t_2 assigns probability 0.5 to player 1 being **certain** $o = a$.

Example

$$T_1 = \{t_1, t_1'\},\ T_2 = \{t_2, t_2'\},\ S = \{a, b\}$$

Player 1: $\lambda_1(t_1)$

	a	b
t_2	1	0
t_2'	0	0

$\lambda_1(t_1')$

	a	b
t_2	0	0
t_2'	0.3	0.7

Player 2: $\lambda_2(t_2)$

	a	b
t_1	0	0.5
t_1'	0.5	0

$\lambda_2(t_2')$

	a	b
t_1	0	0
t_1'	0	1

t_2' is **certain** player 1 is certain that he is certain the $o = b$.

Example

$$T_1 = \{t_1, t_1'\},\ T_2 = \{t_2, t_2'\},\ S = \{a, b\}$$

Player 1: $\lambda_1(t_1)$

	a	b
t_2	1	0
t_2'	0	0

$\lambda_1(t_1')$

	a	b
t_2	0	0
t_2'	0.3	0.7

Player 2: $\lambda_2(t_2)$

	a	b
t_1	0	0.5
t_1'	0.5	0

$\lambda_2(t_2')$

	a	b
t_1	0	0
t_1'	0	1

t_2' is **certain** player 1 is certain that he is certain that $o = b$.

More on Types

For simplicity, we assume $S = \times_{i \in \mathcal{A}} S_i$, where each S_i is a strategy space for agent i in some fixed game G. In this case, $\lambda_i : T_i \rightarrow \Delta(S_{-i} \times T_{-i})$.

A fixed state $(s_1, t_1, s_2, t_2, \ldots, s_n, t_n)$ specifies the strategies and each player's *entire hierarchy of beliefs*:

More on Types

For simplicity, we assume $S = \times_{i \in \mathcal{A}} S_i$, where each S_i is a strategy space for agent i in some fixed game G. In this case, $\lambda_i : T_i \rightarrow \Delta(S_{-i} \times T_{-i})$.

A fixed state $(s_1, t_1, s_2, t_2, \ldots, s_n, t_n)$ specifies the strategies and each player's *entire hierarchy of beliefs*:

1. i's first-order beliefs: $T_i \mapsto \Delta(S_{-i} \times T_{-i}) \mapsto \Delta(S_{-i})$ (marginalizing)

More on Types

For simplicity, we assume $S = \times_{i \in \mathcal{A}} S_i$, where each S_i is a strategy space for agent i in some fixed game G. In this case, $\lambda_i : T_i \rightarrow \Delta(S_{-i} \times T_{-i})$.

A fixed state $(s_1, t_1, s_2, t_2, \ldots, s_n, t_n)$ specifies the strategies and each player's *entire hierarchy of beliefs*:

1. i's first-order beliefs: $T_i \mapsto \Delta(S_{-i} \times T_{-i}) \mapsto \Delta(S_{-i})$ (marginalizing)

2. i's second-order beliefs: $T_i \mapsto \Delta(S_{-i} \times T_{-i}) \mapsto \Delta(S^{-i} \times \times_{i \neq j} \Delta(S_{-j} \times T_{-j})) \mapsto \Delta(S_{-i} \times \times_{j \neq i} \Delta(S_{-j}))$ (marginalizing)

More on Types

- For any given set S of external states we can use a type space on S to provide consistent representations of the players' beliefs.

More on Types

- For any given set S of external states we can use a type space on S to provide consistent representations of the players' beliefs.

- Every state in a belief model or type space induces an infinite hierarchy of beliefs, but *not all consistent and coherent infinite hierarchies are in any finite model*. It is not obvious that even in an infinite model that all such hierarchies of beliefs can be represented.

More on this later...

Literature

R. Myerson. *Harsanyi's Games with Incomplete Information*. Special 50th anniversary issue of *Management Science*, 2004.

M. Siniscalchi. *Epistemic Game Theory: Beliefs and Types*. New Palgrave Dictionary of Economics (forthcoming).

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

Single-Agent Epistemic Logic: The Language

φ 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

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

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

- $p \in$ 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

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

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

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

Single-Agent Epistemic Logic: The Language

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

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

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

Single-Agent Epistemic Logic: The Language

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

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

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

Single-Agent Epistemic Logic: The Language

φ 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

φ 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

φ 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"

$L\varphi$: "φ is an epistemic possibility"

$KL\varphi$: "Ann knows that she thinks φ 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 "according to the agent's current information, w and v are indistinguishable"
- $V : \text{At} \rightarrow \wp(W)$ 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.

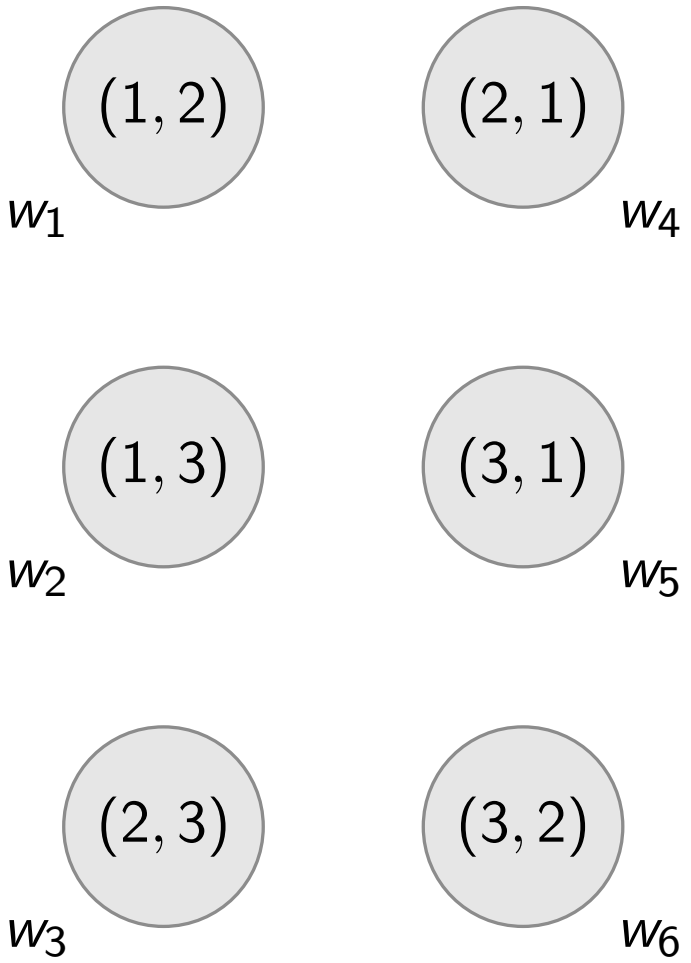
What are the relevant states?

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?

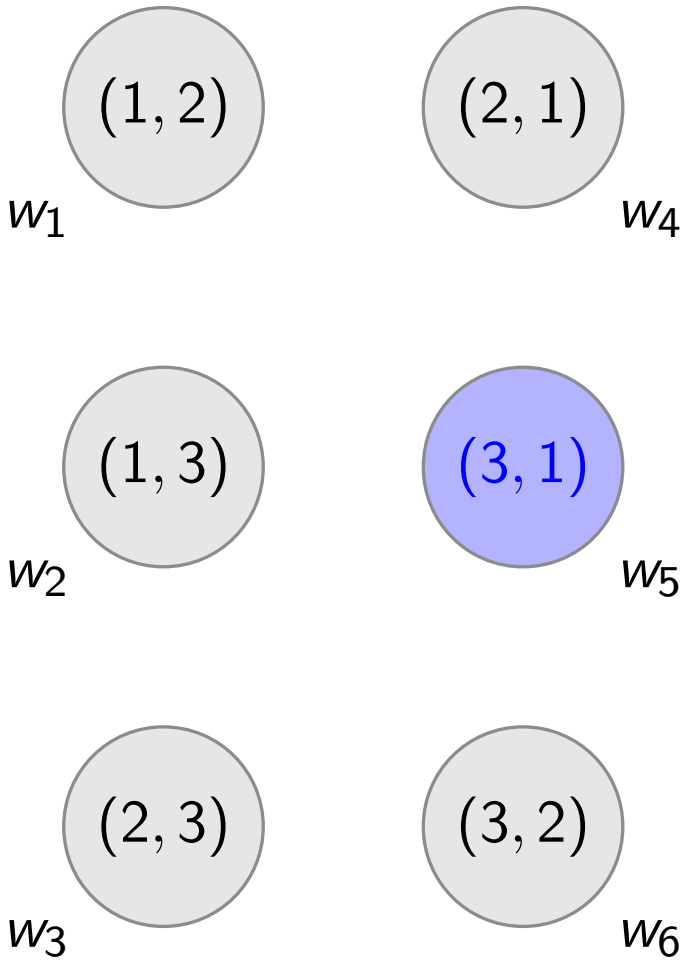


Example

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

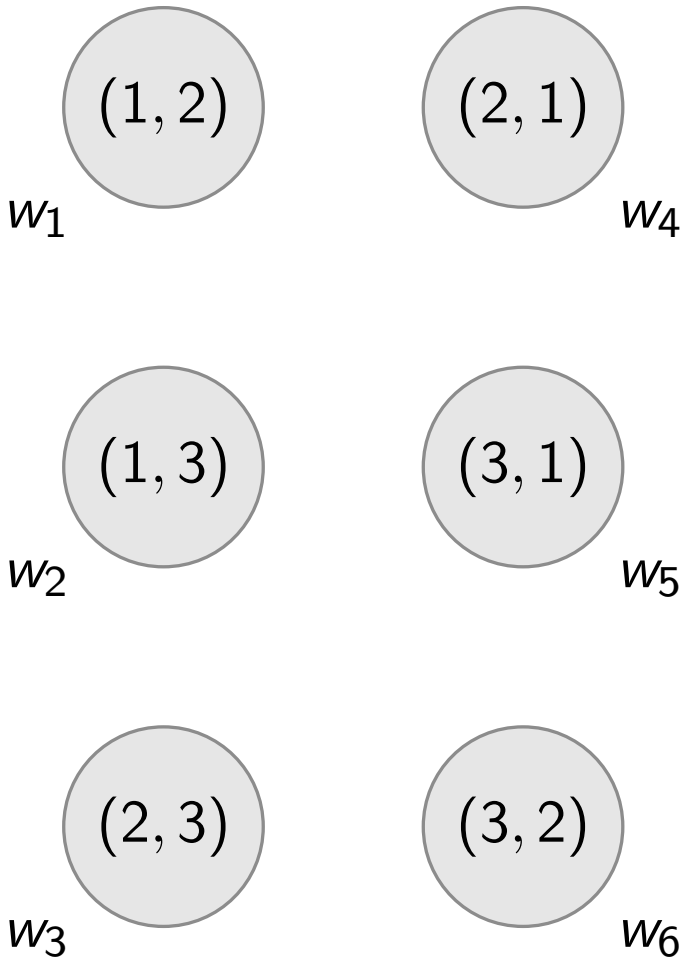


Example

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?

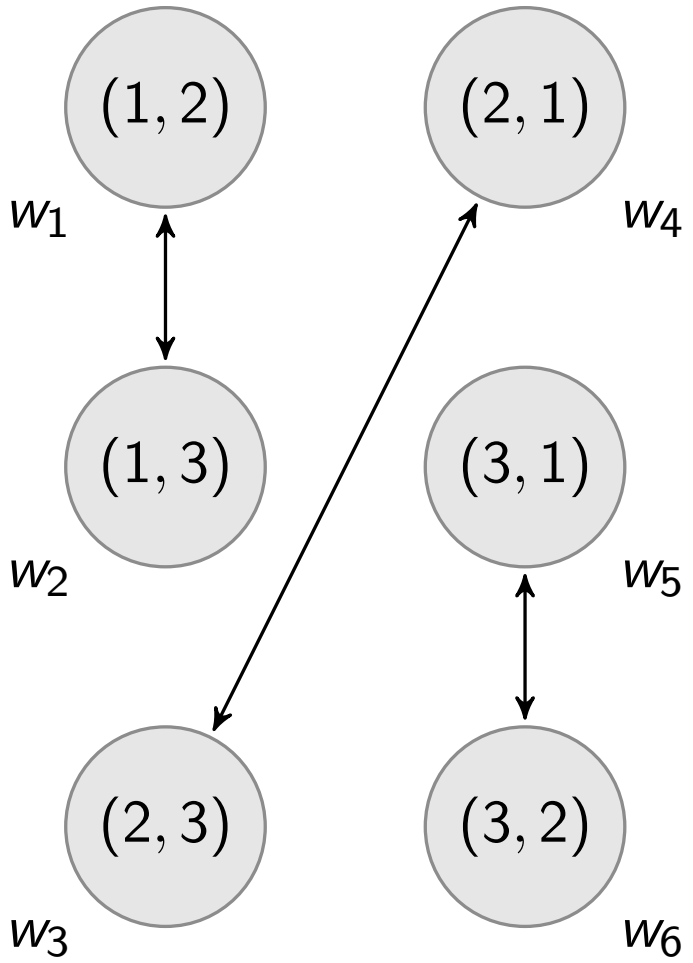


Example

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?

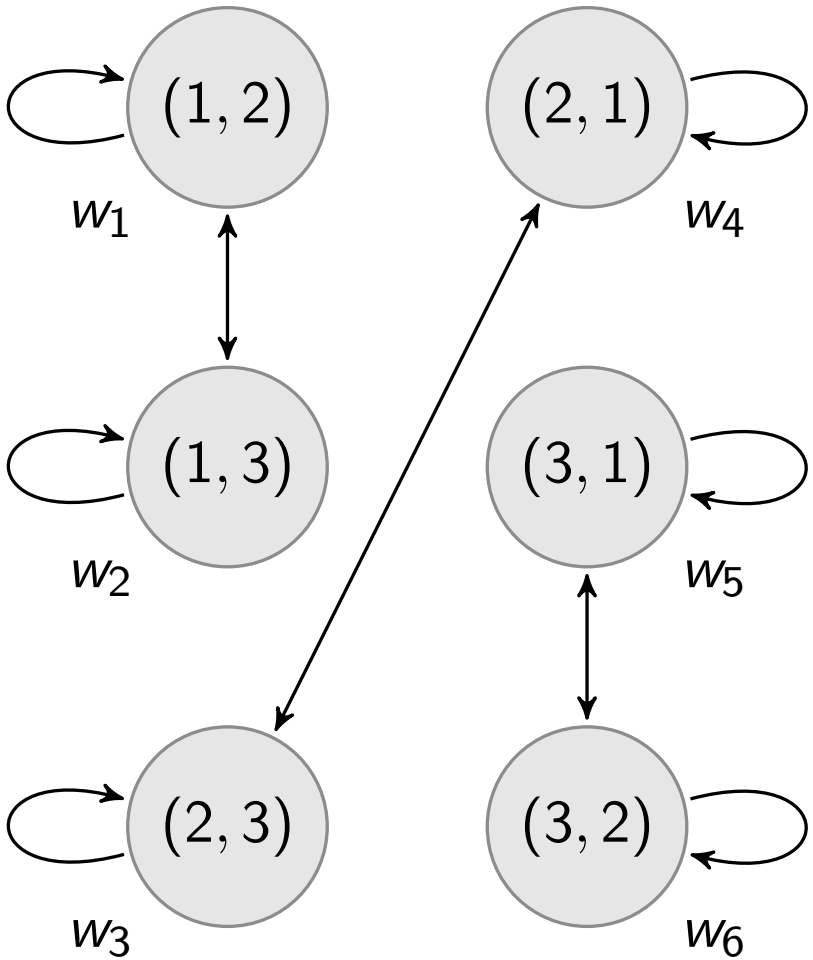


Example

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

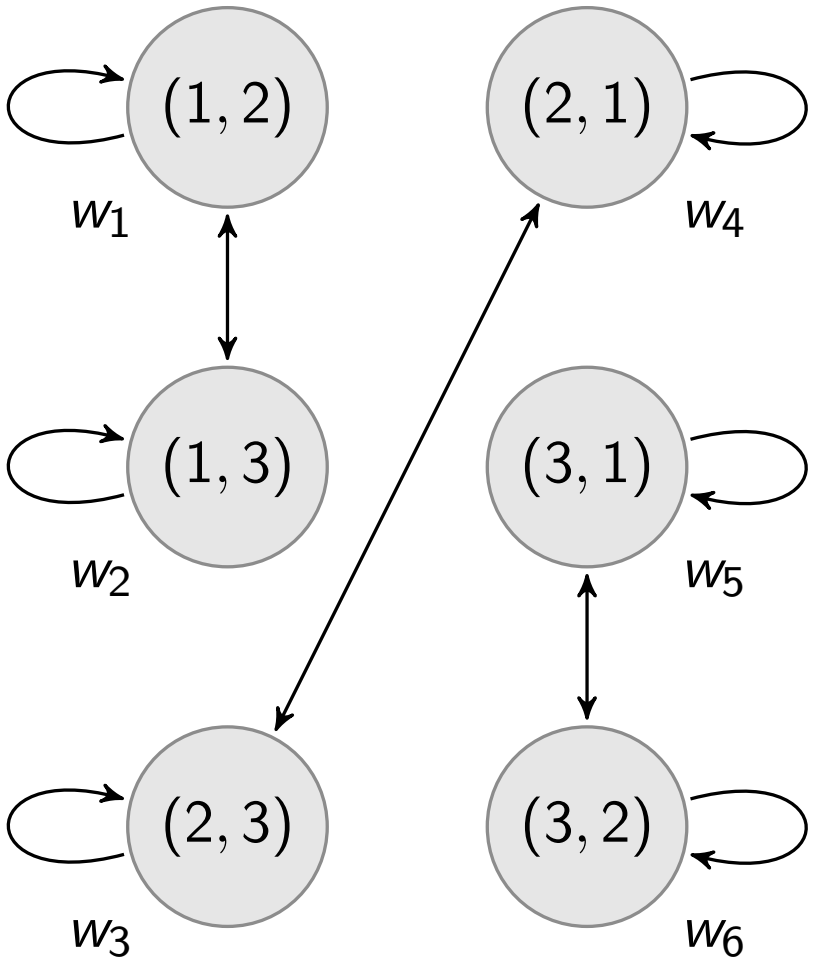
Suppose there are three cards: 1, 2 and 3.

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

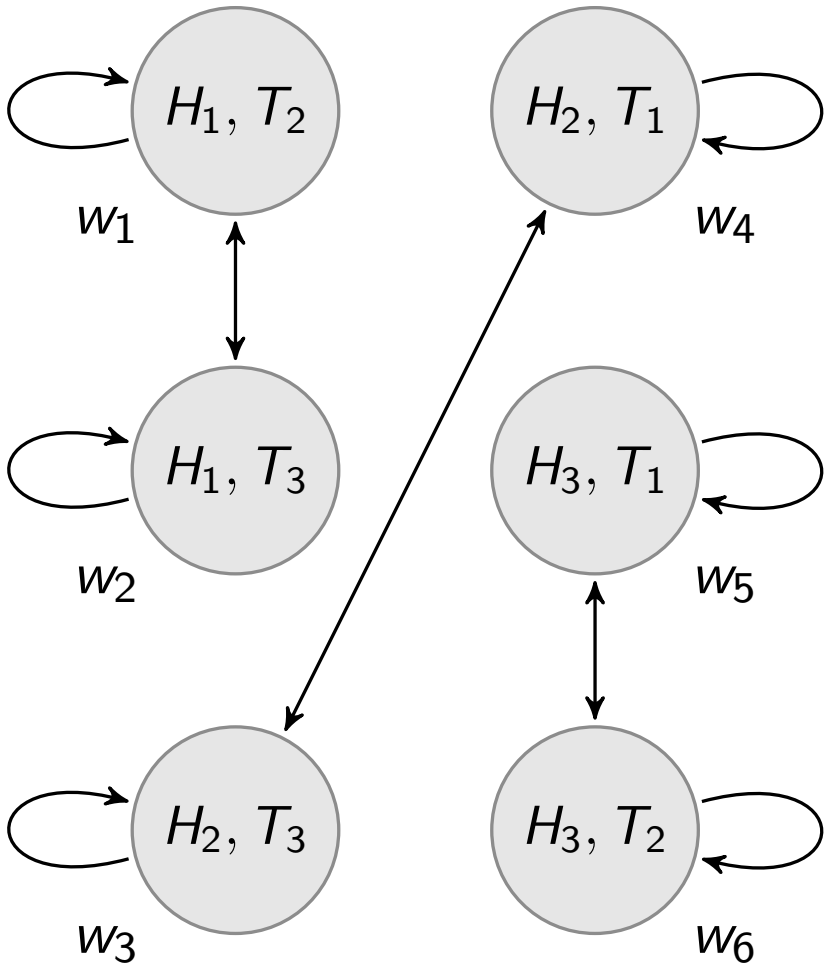
Suppose there are three cards: 1, 2 and 3.

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 φ 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 $w \in W$

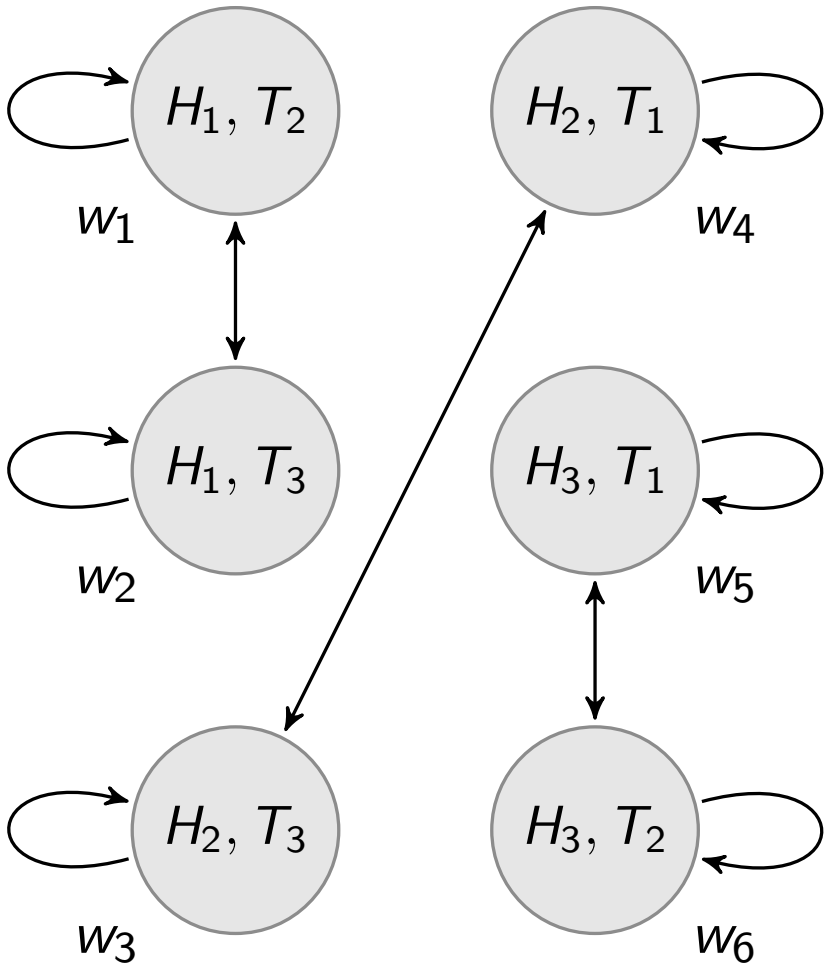
$\mathcal{M}, w \models \varphi$ is defined as follows:

- ✓ $\mathcal{M}, w \models p$ 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.

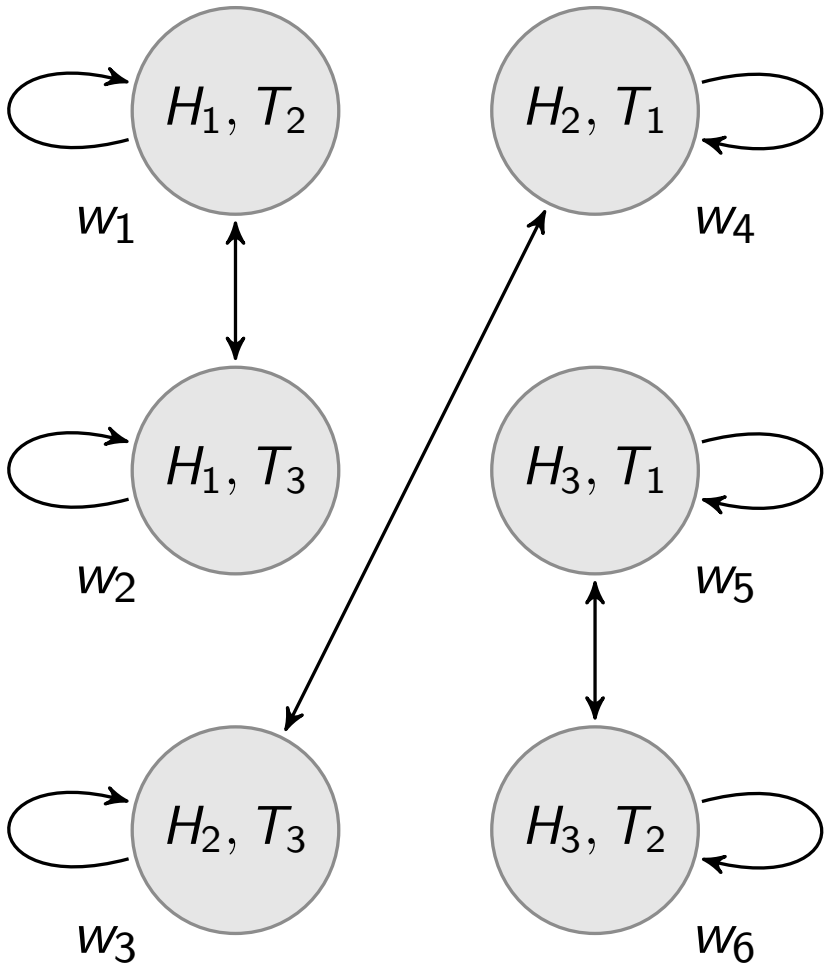


Example

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.

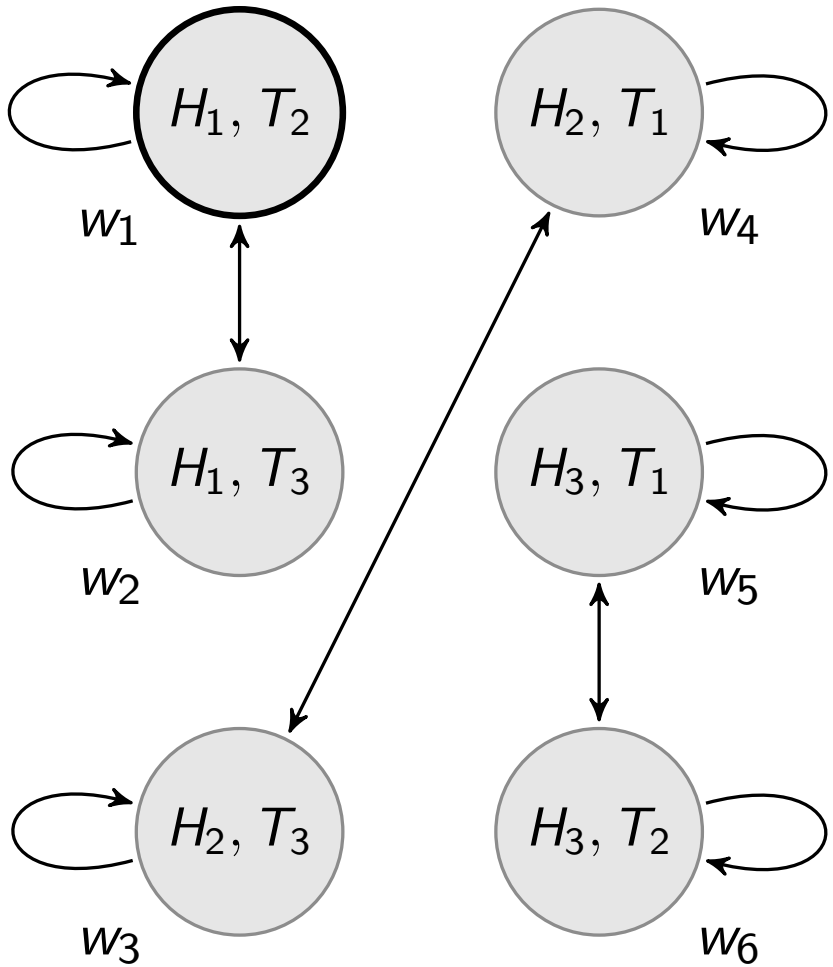


Example

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.

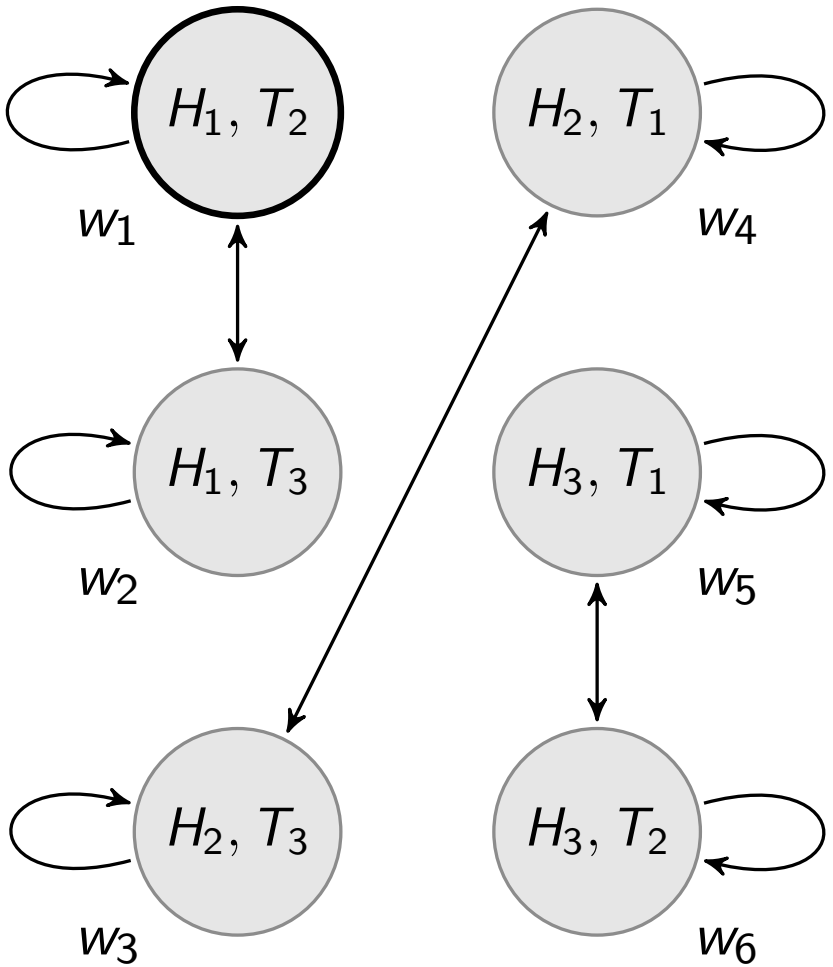


Example

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.

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

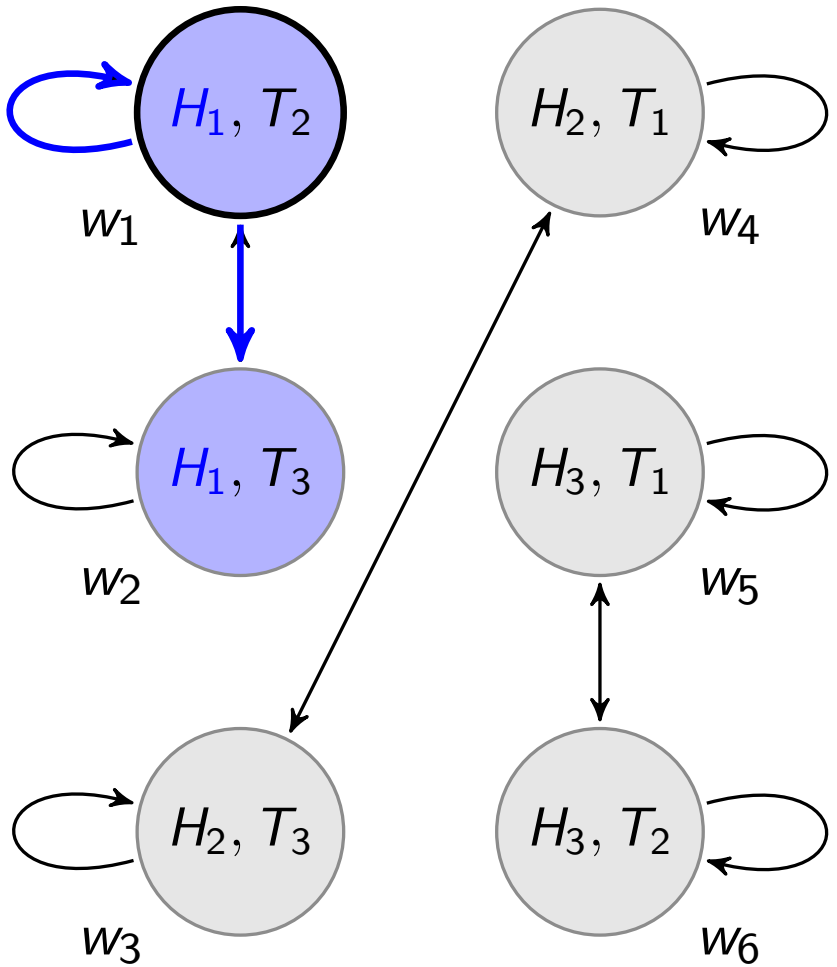


Example

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.

$\mathcal{M}, w_1 \models K H_1$



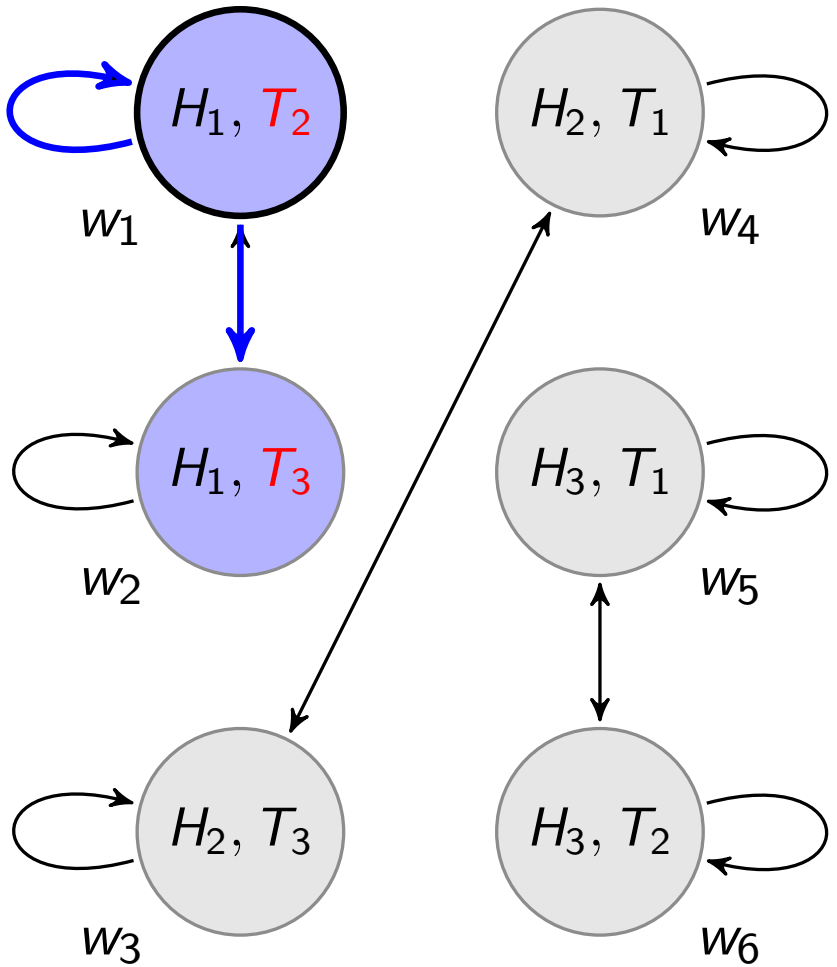
Example

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.

$\mathcal{M}, w_1 \models K H_1$

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

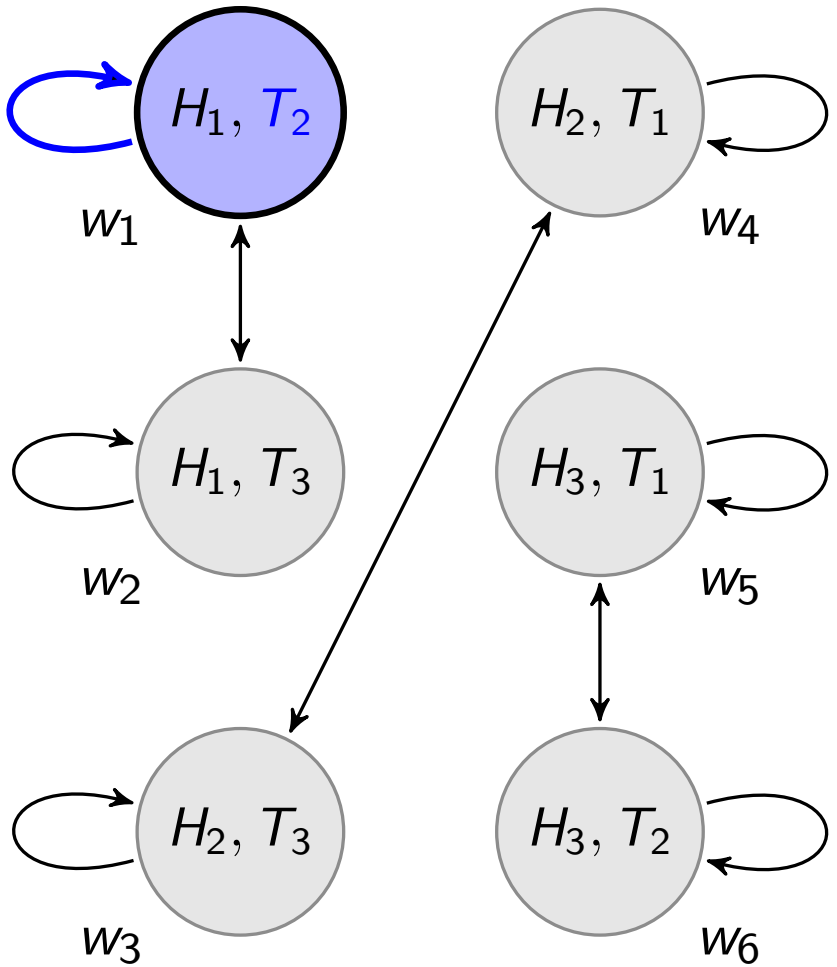


Example

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?

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
$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$	—	Logical Omniscience
$K\varphi \rightarrow \varphi$	Reflexive	Truth

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

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

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

Background: Multiagent Epistemic Logic

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

Kripke Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, V\rangle$ 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 $w \sim_i v$, then $\mathcal{M}, v \models \varphi$

Background: Multiagent Epistemic Logic

- $K_A K_B \varphi$: "Ann knows that Bob knows φ"
- $K_A(K_B \varphi \vee K_B \neg \varphi)$: "Ann knows that Bob knows whether φ
- $\neg K_B K_A K_B(\varphi)$: "Bob does not know that Ann knows that Bob knows that φ"

Background: Group Knowledge

$K_A P$: "Ann knows that P"

Background: Group Knowledge

$K_A P$: "Ann knows that P"

$K_B P$: "Bob knows that P"

Background: 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"

Background: 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".

Background: 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 $EP := K_A P \wedge K_B P$

Background: 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".

Background: 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".

Background: 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."

Background: Common Knowledge

CP: "It is common knowledge that P" — "Everyone knows that everyone knows that everyone knows that $\cdots$ P".

Three Views of Common Knowledge

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

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

Three Views of Common Knowledge

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

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

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

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

Three Views of Common Knowledge

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

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

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

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

3. There is a *shared situation* s such that
   - s entails φ
   - s entails i knows φ
   - s entails j knows φ

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

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

Background: Common Knowledge

Definition

The operator "everyone knows φ", denoted $E\varphi$, is defined as follows

$$E\varphi \quad := \quad \bigwedge_{i \in \mathcal{A}} K_i \varphi$$

Background: Common Knowledge

Definition

The operator "everyone knows φ", denoted $E\varphi$, is defined as follows

$$E\varphi \quad := \quad \bigwedge_{i \in \mathcal{A}} K_i \varphi$$

Definition

The multi-agent epistemic language with common knowledge is generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_i\varphi \mid C\varphi$$

where $p \in \text{At}$ and $i \in \mathcal{A}$.

Background: Common Knowledge

Definition

The truth of $C\varphi$ is:

$$\mathcal{M}, w \models C\varphi \quad \text{iff} \quad \text{for all } v \in W, \text{ if } wR^*v \text{ then } \mathcal{M}, v \models \varphi$$

where $R^* := (\bigcup_{i \in \mathcal{A}} R_i)^*$ is the reflexive transitive closure of the union of the R_i's.

Background: Common Knowledge

Definition

The truth of $C\varphi$ is:

$$\mathcal{M}, w \models C\varphi \quad \text{iff} \quad \text{for all } v \in W, \text{ if } wR^*v \text{ then } \mathcal{M}, v \models \varphi$$

where $R^* := (\bigcup_{i \in \mathcal{A}} R_i)^*$ is the reflexive transitive closure of the union of the R_i's.

$\mathcal{M}, w \models C\varphi$ iff every finite path starting at w ends with a state satisfying φ.

Bayesian Rationality

- Instrumental Rationality
- Decision Theory
  - Endogenous and Exogenous Uncertainty
  - Maximization of Expected Utility

...to understand the fundamental ideas of game theory, one should begin by studying decision theory.
-R. Myerson (recent Nobel Prize winner in Economics)

Rationality in Games

Definition

The *expected value* for player i of playing strategy s_i given that he is of type t_i is defined as follows.

$$EV_{t_i}(s_i) = \sum_{t'_{-i}} \sum_{\sigma'_{-i}} \lambda_i(t_i)(\sigma'_{-i}, t'_{-i}) v_i(s_i, \sigma'_{-i})$$

An Example

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

$\lambda_r(t_r)$

u_c	0	$1/2$
t_c	0	$1/2$
	L	R

$\lambda_r(u_r)$

u_c	$1/2$	0
t_c	0	$1/2$
	L	R

$\lambda_c(t_c)$

u_r	0	$1/2$
t_r	0	$1/2$
	U	D

$\lambda_c(u_c)$

u_r	$1/2$	0
t_r	0	$1/2$
	U	D

State: (D, t_r, R, t_c)

An Example

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

$\lambda_r(t_r)$

u_c	0	$1/2$
t_c	0	$1/2$
	L	R

$\lambda_r(u_r)$

u_c	$1/2$	0
t_c	0	$1/2$
	L	R

$\lambda_c(t_c)$

u_r	0	$1/2$
t_r	0	$1/2$
	U	D

$\lambda_c(u_c)$

u_r	$1/2$	0
t_r	0	$1/2$
	U	D

State: (D, t_r, R, t_c)

An Example

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

$\lambda_r(t_r)$

u_c	0	$1/2$
t_c	0	$1/2$
	L	R

$\lambda_r(u_r)$

u_c	$1/2$	0
t_c	0	$1/2$
	L	R

$\lambda_c(t_c)$

u_r	0	$1/2$
t_r	0	$1/2$
	U	D

$\lambda_c(u_c)$

u_r	$1/2$	0
t_r	0	$1/2$
	U	D

r is **correct** about c's strategy (similarly, for c).

An Example

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

$\lambda_r(t_r)$

u_c	0	1/2
t_c	0	1/2
	L	R

$\lambda_r(u_r)$

u_c	1/2	0
t_c	0	1/2
	L	R

$\lambda_c(t_c)$

u_r	0	1/2
t_r	0	1/2
	U	D

$\lambda_c(u_c)$

u_r	1/2	0
t_r	0	1/2
	U	D

r thinks it is possible c is wrong about her strategy

An Example

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

$\lambda_r(t_r)$

u_c	0	$1/2$
t_c	0	$1/2$
	L	R

$\lambda_r(u_r)$

u_c	$1/2$	0
t_c	0	$1/2$
	L	R

$\lambda_c(t_c)$

u_r	0	$1/2$
t_r	0	$1/2$
	U	D

$\lambda_c(u_c)$

u_r	$1/2$	0
t_r	0	$1/2$
	U	D

r is **rational**. (Similarly for c)

An Example

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

$\lambda_r(t_r)$

u_c	0	$1/2$
t_c	0	$1/2$
	L	R

$\lambda_r(u_r)$

u_c	$1/2$	0
t_c	0	$1/2$
	L	R

$\lambda_c(t_c)$

u_r	0	$1/2$
t_r	0	$1/2$
	U	D

$\lambda_c(u_c)$

u_r	$1/2$	0
t_r	0	$1/2$
	U	D

r thinks it is possible that c is **irrational**.

Where are we?

1. Formal definition of a strategic interactive situation
2. Formal models of informational attitudes
3. Formal definition of rationality

Where are we?

1. Formal definition of a strategic interactive situation
2. Formal models of informational attitudes
3. Formal definition of rationality

Where are we going?

1. Which informational attitudes should we focus on (knowledge, beliefs, certainty, assumption, etc.)?
2. Can we use these frameworks to understand what it means for players to *be rational*?

Expectation 1: Rationality and common belief of rationality

- What happens if all players are rational, believe that all players are rational, believe that all players believe that (...)?

Expectation 1: Rationality and common belief of rationality

- What happens if all players are rational, believe that all players are rational, believe that all players believe that (...)?

- "Classical" assumption about game-theoretic analysis. See e.g. Myerson (1991).

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

- B is a bad strategy for Bob.

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

- B is a bad strategy for Bob.
- It is *never* rational for him to choose B.

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

- A type t_B of Bob would be rational in choosing B iff:

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

- A type t_B of Bob would be rational in choosing B iff:

$$EV_{t_B}(B) \geq EV_{t_B}(A)$$

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

- A type t_B of Bob would be rational in choosing B iff:

$$v_{Bob}(aB)\lambda_{Bob}(t_{Bob})(a) + v_{Bob}(bB)\lambda_{Bob}(t_{Bob})(b) \geq$$
$$v_{Bob}(aA)\lambda_{Bob}(t_{Bob})(a) + v_{Bob}(bA)\lambda_{Bob}(t_{Bob})(b)$$

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

- A type t_B of Bob would be rational in choosing B iff:

$$1\lambda_{Bob}(t_{Bob})(a)+0\lambda_{Bob}(t_{Bob})(b) \geq 2\lambda_{Bob}(t_{Bob})(a)+1\lambda_{Bob}(t_{Bob})(b)$$

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

- A type t_B of Bob would be rational in choosing B iff:

$$\lambda_{Bob}(t_{Bob})(a) \geq 2\lambda_{Bob}(t_b)(a) + \lambda_{Bob}(t_{Bob})(b)$$

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

- A type t_B of Bob would be rational in choosing B iff:

$$0 \geq \lambda_{Bob}(t_{Bob})(a) + \lambda_{Bob}(t_{Bob})(b)$$

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

- A type t_B of Bob would be rational in choosing B iff:

$$0 \geq \lambda_{Bob}(t_{Bob})(a) + \lambda_{Bob}(t_{Bob})(b)$$

But $\lambda_{Bob}(t_{Bob})(a) + \lambda_{Bob}(t_{Bob})(b) = 0$!

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

- Bob never plays **B** at state (σ, t) if he is rational at that state.

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

- Bob never plays **B** at state (σ, t) if he is rational at that state.
- But then if Ann's type at that state believes that Bob is rational,

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

- Bob never plays **B** at state (σ, t) if he is rational at that state.
- But then if Ann's type at that state believes that Bob is rational, that type must assign probability 1 to Bob playing **A**.

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

- Bob never plays **B** at state (σ, t) if he is rational at that state.
- But then if Ann's type at that state believes that Bob is rational, that type must assign probability 1 to Bob playing **A**.
- Given this belief, **a** is her only rational strategy.

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

- If Ann and Bob are rational, and Ann believes that Bob is rational at state (σ, t),

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

- If Ann and Bob are rational, and Ann believes that Bob is rational at state (σ, t), then $\sigma = \text{aA}$.

Example

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

- If Ann and Bob are rational, and Ann believes that Bob is rational at state (σ, t), then $\sigma = \text{aA}$.
- This strategy profile is the only one that survives *iterated elimination of strictly dominated strategies*.

Strictly dominated strategies

Definition

A strategy s_i is *strictly dominated* by another strategy s_i' iff for all combinations of choices of the other players σ_{-i} :

$$v_i(s_i, \sigma_{-i}) < v_i(s_i', \sigma_{-i})$$

Iterated elimination of strictly dominated strategies

1. Start with a game;

Iterated elimination of strictly dominated strategies

1. Start with a game;
2. Eliminate all strictly dominated strategies;

Iterated elimination of strictly dominated strategies

1. Start with a game;
2. Eliminate all strictly dominated strategies;
3. Look at the reduced game;

Iterated elimination of strictly dominated strategies

1. Start with a game;
2. Eliminate all strictly dominated strategies;
3. Look at the reduced game;
4. Eliminate all strictly dominated strategies here;

Iterated elimination of strictly dominated strategies

1. Start with a game;
2. Eliminate all strictly dominated strategies;
3. Look at the reduced game;
4. Eliminate all strictly dominated strategies here;
5. Repeat 3 and 4 until you don't eliminate anything.

Iterated elimination of strictly dominated strategies

1. Start with a game;
2. Eliminate all strictly dominated strategies;
3. Look at the reduced game;
4. Eliminate all strictly dominated strategies here;
5. Repeat 3 and 4 until you don't eliminate anything.

	A	B
a	1, 2	0, 1
b	0, 1	1, 0

Iterated elimination of strictly dominated strategies

1. Start with a game;
2. Eliminate all strictly dominated strategies;
3. Look at the reduced game;
4. Eliminate all strictly dominated strategies here;
5. Repeat 3 and 4 until you don't eliminate anything.

	A
a	1, 2
b	0, 1

Iterated elimination of strictly dominated strategies

1. Start with a game;
2. Eliminate all strictly dominated strategies;
3. Look at the reduced game;
4. Eliminate all strictly dominated strategies here;
5. Repeat 3 and 4 until you don't eliminate anything.

	A
a	1, 2

Common knowledge of rational and elimination of strictly dominated strategies

- If Ann and Bob are rational, and Ann believes that Bob is rational at state (σ, t), then $\sigma = \text{aA}$.

Common knowledge of rational and elimination of strictly dominated strategies

- If Ann and Bob are rational, and Ann believes that Bob is rational at state (σ, t), then $\sigma = \text{aA}$.
- For this game we need rationality and only one level of higher-order information to conclude that aA will be played.

Common knowledge of rational and elimination of strictly dominated strategies

- If Ann and Bob are rational, and Ann believes that Bob is rational at state (σ, t), then $\sigma =$ aA.
- For this game we need rationality and only one level of higher-order information to conclude that aA will be played. But in the general case:

Theorem

For any state (σ, t) of a type structure for an arbitrary finite game $\mathbb{G}$, if all players are rational and it is common belief that all players are rational at (σ, t), then σ is a iteratively non-dominated strategy profile.

A. Brandenburger and E. Dekel. *Rationalizability and correlated equilibria*. *Econometrica*, 55:1391-1402, 1987.

Comments on characterization results

- If [such and such expectations] at state (σ, t), then [such and such *solution concept*] is played at that state.

Comments on characterization results

- If [such and such expectations] at state (σ, t), then [such and such *solution concept*] is played at that state.
- *Solution concepts*: proposal as to what is *rational* in a game. (traditionally)

Comments on characterization results

- If [such and such expectations] at state (σ, t), then [such and such *solution concept*] is played at that state.
- *Solution concepts*: proposal as to what is *rational* in a game. (traditionally)
- What about the converse?

Comments on characterization results

- If [such and such expectations] at state (σ, t), then [such and such *solution concept*] is played at that state.
- *Solution concepts*: proposal as to what is *rational* in a game. (traditionally)
- What about the converse?
  - If [such and such *solution concept*] then one can build a state in a model such that [such and such expectations] hold.

Epistemic Characterizations of Solutions Concepts

If the players all satisfy some **epistemic condition** involving some form of **rationality** (eg., common knowledge of rationality) then the players will play according to some solution concept (eg., Nash equilibrium, iterated removal of strongly dominated strategies, ...).

Epistemic Characterizations of Solutions Concepts

If the players all satisfy some **epistemic condition** involving some form of **rationality** (eg., common knowledge of rationality) then the players will play according to some solution concept (eg., Nash equilibrium, iterated removal of strongly dominated strategies, ...).

Two key assumptions about the rationality of players:

1. Common *knowledge* of *rationality* (i.e., common knowledge of choosing optimally)
2. Common prior

Removal of Strictly Dominated Strategies

We have seen that *common knowledge of rationality* implies that the players will follow the process of iteratively removing strictly dominated strategies.

Removal of Strictly Dominated Strategies

We have seen that *common knowledge of rationality* implies that the players will follow the process of iteratively removing strictly dominated strategies.

1. Players should not choose strictly dominated strategies (*it is never rational*)
2. Assuming the above statement is common knowledge is equivalent to assuming the players iteratively remove strictly dominated strategies.

Admissibility

Can the same be proven for *admissibility*, i.e., avoidance of *weakly* dominated strategies?

Admissibility

Can the same be proven for *admissibility*, i.e., avoidance of *weakly* dominated strategies?

	L	R
T	$1,1$	$0,0$
M	$1,1$	$2,1$
B	$0,0$	$2,1$

Admissibility

Can the same be proven for *admissibility*, i.e., avoidance of *weakly* dominated strategies?

	L	R
T	$1,1$	$0,0$
M	$1,1$	$2,1$
B	$0,0$	$2,1$

Admissibility

Does assuming that it is commonly known that players play only admissible strategies lead to a process of iterated removal of weakly dominated strategies?

Admissibility

Does assuming that it is commonly known that players play only admissible strategies lead to a process of iterated removal of weakly dominated strategies? no!

L. Samuelson. *Dominated Strategies and Common Knowledge*. Games and Economic Behavior (1992).

Results

1. Removal of weakly dominated strategies and common knowledge of admissibility diverge.

L. Samuelson. *Dominated Strategies and Common Knowledge*. Games and Economic Behavior (1992).

Results

1. Removal of weakly dominated strategies and common knowledge of admissibility diverge.

2. There exist games in which assuming that admissibility is common knowledge does not provide players with sufficient information to determine which strategies should be eliminated on admissibility grounds.

L. Samuelson. *Dominated Strategies and Common Knowledge*. Games and Economic Behavior (1992).

Results

1. Removal of weakly dominated strategies and common knowledge of admissibility diverge.

2. There exist games in which assuming that admissibility is common knowledge does not provide players with sufficient information to determine which strategies should be eliminated on admissibility grounds.

3. There exists games in which assuming that admissibility is common knowledge yields a contradiction

L. Samuelson. *Dominated Strategies and Common Knowledge*. Games and Economic Behavior (1992).

An Issue

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

An Issue

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

Suppose rationality incorporates *admissibility* (or *cautiousness*).

An Issue

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

Suppose rationality incorporates *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?

An Issue

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

Suppose rationality incorporates *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 are rational *seems to conflict with the condition that the players think the other players are rational*

An Issue

The argument for deletion of a weakly dominated strategy for player i is that he contemplates the possibility that every strategy combination of his rivals occurs with positive probability. However, this hypothesis clashes with the logic of iterated deletion, which assumes, precisely, that eliminated strategies are not expected to occur.

Mas-Colell, Whinston and Green. *Introduction to Microeconomics*. 1995.

Both Including and Excluding a Strategy

One solution is to assume that players consider some strategies *infinitely more likely than other strategies*.

Both Including and Excluding a Strategy

One solution is to assume that players consider some strategies *infinitely more likely than other strategies*.

Lexiographic Probability System: a sequence of probability distributions each infinitely more likely than the next.

Both Including and Excluding a Strategy

One solution is to assume that players consider some strategies *infinitely more likely than other strategies*.

Lexiographic Probability System: a sequence of probability distributions each infinitely more likely than the next.

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

Types of Irrationality?

1. A player is irrational if he does not optimize given its current beliefs

A. Brandenburger, A. Friedenberg, H. J. Keisler. *Admissibility in Games*. Econometrica (2008).

Types of Irrationality?

1. A player is irrational if he does not optimize given its current beliefs
2. A player is irrational if, although he optimizes, he does not consider *all possibilities*

A. Brandenburger, A. Friedenberg, H. J. Keisler. *Admissibility in Games*. Econometrica (2008).

Types of Irrationality?

1. A player is irrational if he does not optimize given its current beliefs
2. A player is irrational if, although he optimizes, he does not consider *all possibilities*

A. Brandenburger, A. Friedenberg, H. J. Keisler. *Admissibility in Games*. Econometrica (2008).

The General Question

Does such a space of all possible (interactive) beliefs exist?

Is there a "large" type space?

A **universal type space** is a types space to which every type space (on the same space of states of nature and same set of agents) can be mapped, preferably in a unique way, by a map that preserves the structure of the type space.

If such a space exists, then the any analysis of a game could be carried out in this space without the risk of missing any "relevant" states of affairs.

Yes, if . . .

The existence of a universal types space depends on the topological and/or measure theoretic assumptions being made about the underlying state space S.

Yes, if . . .

The existence of a universal types space depends on the topological and/or measure theoretic assumptions being made about the underlying state space S.

First shown by Mertens and Zamir (1985)

The problem is to define the set of all infinite hierarchies of beliefs satsifying the same consistency properties (coherency and common knowledge of coherency) as that of hierarchies obtained at some state in a type space.

Kolomogorov Extension Theorem

"Large" Spaces of Beliefs

1. **Universal models**: Start with a space of underlying uncertainty, players form beliefs over this space, believes over this space and the space of 0-th order beliefs, and so on inductively. The question is, does this process end?

2. **Complete models**: The "two-way subjectivity" models described later.

3. **Terminal models**: Given a category $\mathbf{C}$ of models of beliefs, call a model $\mathcal{M}$ in $\mathbf{C}$ terminal if for any other model $\mathcal{N}$ in $\mathbf{C}$, there is a unique belief preserving morphism from $\mathcal{M}$ to $\mathcal{N}$.

Some Literature

A. Brandenburger and E. Dekel. *Hierarchies of Beliefs and Common Knowledge*. Journal of Economic Theory (1993).

Some Literature

A. Brandenburger and E. Dekel. *Hierarchies of Beliefs and Common Knowledge*. Journal of Economic Theory (1993).

A. Heifetz and D. Samet. *Knoweldge Spaces with Arbitrarily High Rank*. Games and Economic Behavior (1998).

Some Literature

A. Brandenburger and E. Dekel. *Hierarchies of Beliefs and Common Knowledge*. Journal of Economic Theory (1993).

A. Heifetz and D. Samet. *Knoweldge Spaces with Arbitrarily High Rank*. Games and Economic Behavior (1998).

L. Moss and I. Viglizzo. *Harsanyi type spaces and final coalgebras constructed from satisfied theories*. EN in Theoretical Computer Science (2004).

Some Literature

A. Brandenburger and E. Dekel. *Hierarchies of Beliefs and Common Knowledge*. Journal of Economic Theory (1993).

A. Heifetz and D. Samet. *Knoweldge Spaces with Arbitrarily High Rank*. Games and Economic Behavior (1998).

L. Moss and I. Viglizzo. *Harsanyi type spaces and final coalgebras constructed from satisfied theories*. EN in Theoretical Computer Science (2004).

A. Friendenberg. *When do type structures contain all hierarchies of beliefs?*. working paper (2007).

Impossibility Result

Doesn't such talk of what Ann believes Bob believes about her, and so on, suggest that some kind of self-reference arises in games, similar to the well-known examples of self-reference in mathematical logic.

A. Brandenburger and H. J. Keisler. *An Impossibility Theorem on Beliefs in Games*. forthcoming in *Studia Logica*.

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong?

$*$ An **assumption** (or strongest belief) is a belief that implies all other beliefs.

A. Brandenburger and H. J. Keisler. *An Impossibility Theorem on Beliefs in Games*. *Studia Logica* (2006).

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong? Yes.

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong? Yes.

Then according to Ann, Bob's assumption is right.

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong? Yes.

Then according to Ann, Bob's assumption is right.

But then, Ann does not believe Bob's assumption is wrong.

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong? Yes.

Then according to Ann, Bob's assumption is right.

But then, Ann does not believe Bob's assumption is wrong.

So, the answer must be no.

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong? No.

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong? No.

Then Ann does not believe that Bob's assumption is wrong.

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong? No.

Then Ann does not believe that Bob's assumption is wrong.

Then, in Ann's view, Bob's assumption is wrong.

A Paradox

Ann believes that Bob assumes* that
Ann believes that Bob's assumption is wrong.

Does Ann believe that Bob's assumption is wrong? No.

Then Ann does not believe that Bob's assumption is wrong.

Then, in Ann's view, Bob's assumption is wrong.

So, the answer must be yes.

Main Result

Belief Model: a set of states for each player, and a relation for each player that specifies when a state of one player considers a state of the other player to be possible.

Main Result

Belief Model: a set of states for each player, and a relation for each player that specifies when a state of one player considers a state of the other player to be possible.

Language: the language used by the players to formulate their beliefs.

Main Result

Belief Model: a set of states for each player, and a relation for each player that specifies when a state of one player considers a state of the other player to be possible.

Language: the language used by the players to formulate their beliefs.

Complete: A belief model is complete for a language if every statement in a player's langauge which is possible (i.e. true for some states) can be assumed by the player.

Main Result

Belief Model: a set of states for each player, and a relation for each player that specifies when a state of one player considers a state of the other player to be possible.

Language: the language used by the players to formulate their beliefs.

Complete: A belief model is complete for a language if every statement in a player's langauge which is possible (i.e. true for some states) can be assumed by the player.

Theorem (Brandenburger and Keisler) No belief model can be complete for a language that contains first-order logic.

Discussion

- Which fragments of first-order logic have assumption complete models?
- Other logical considerations
- What role do "large" belief spaces play in the analysis of solution concepts?

Open Question

Can we find a logic $\mathcal{L}$ such that

1. Complete belief models for $\mathcal{L}$ exist for each game;
2. notions such as rationality, belief in rationality, etc. are expressible in $\mathcal{L}$; and
3. the ingredients in 1 and 2 can be combined to yield various well-known game-theoretic solution concepts.

We think of a particular incomplete structure as giving the "context" in which the game is played.

We think of a particular incomplete structure as giving the "context" in which the game is played. In line with Savage's Small-Worlds idea in decision theory [...], who the players are in the given game can be seen as a shorthand for their experiences before the game.

We think of a particular incomplete structure as giving the "context" in which the game is played. In line with Savage's Small-Worlds idea in decision theory [...], who the players are in the given game can be seen as a shorthand for their experiences before the game. The players' possible characteristics — including their possible types — then reflect the prior history or context. (Seen in this light, complete structures represent a special "context-free" case, in which there has been no narrowing down of types.) (pg. 319)

A. Brandenburger, A. Friedenberg, H. J. Keisler. *Admissibility in Games*. Econometrica (2008).

Literature

See, for example,

R. Aumann. *Interactive Epistemology I & II*. International Journal of Game Theory (1999).

Literature

See, for example,

R. Aumann. *Interactive Epistemology I & II*. International Journal of Game Theory (1999).

P. Battigalli and G. Bonanno. *Recent results on belief, knowledge and the epistemic foundations of game theory*. Research in Economics (1999).

Literature

See, for example,

R. Aumann. *Interactive Epistemology I & II*. International Journal of Game Theory (1999).

P. Battigalli and G. Bonanno. *Recent results on belief, knowledge and the epistemic foundations of game theory*. Research in Economics (1999).

B. de Bruin. *Explaining Games*. Ph.D. Thesis, ILLC (2004).

Literature

See, for example,

R. Aumann. *Interactive Epistemology I & II*. International Journal of Game Theory (1999).

P. Battigalli and G. Bonanno. *Recent results on belief, knowledge and the epistemic foundations of game theory*. Research in Economics (1999).

B. de Bruin. *Explaining Games*. Ph.D. Thesis, ILLC (2004).

A. Brandenburger. *The Power of Paradox: Some Recent Developments in Interactive Epistemology*. International Journal of Game Theory (2007).

Epistemic Characterizations of Solutions Concepts

If the players all satisfy some **epistemic condition** involving some form of **rationality** (eg., common knowledge of rationality) then the players will play according to some solution concept (eg., Nash equilibrium, iterated removal of strongly dominated strategies, ...).

Epistemic Characterizations of Solutions Concepts

If the players all satisfy some **epistemic condition** involving some form of **rationality** (eg., common knowledge of rationality) then the players will play according to some solution concept (eg., Nash equilibrium, iterated removal of strongly dominated strategies, ...).

The key "axioms" and assumptions:

1. Players know there own strategies (and types)
2. Players are expected utility maximizers
3. The above facts are common *knowledge*
4. Players do not completely rule out choices of the other players
5. The players do not have any (soft) information about the other players

The point of view of this model is not normative; it is not meant to advise the players what to do. The players do whatever they do; their strategies are taken as given.

The point of view of this model is not normative; it is not meant to advise the players what to do. The players do whatever they do; their strategies are taken as given. Neither is it meant as a description of what human beings actually do in interactive situations.

The point of view of this model is not normative; it is not meant to advise the players what to do. The players do whatever they do; their strategies are taken as given. Neither is it meant as a description of what human beings actually do in interactive situations. The most appropriate term is perhaps "analytic"; it asks, what are the implications *of rationality in interactive situations? Where does it lead?*

The point of view of this model is not normative; it is not meant to advise the players what to do. The players do whatever they do; their strategies are taken as given. Neither is it meant as a description of what human beings actually do in interactive situations. The most appropriate term is perhaps "analytic"; it asks, what are the implications *of rationality in interactive situations? Where does it lead? This question may be as important as, or even more important than, more direct "tests" of the relevance of the rationality hypothesis.*

R. Aumann. *Irrationality in Game Theory*. 1992.

Logics for Games

Recognize that (extensive, strategic) games form a class of modal models.

G. Bonanno. *Modal logic and game theory: Two alternative approaches*. Risk Decision and Policy 7 (2002).

J. van Benthem. *Extensive Games as Process Models*. JOLLI 11 (2002).

Logics for Games

Recognize that (extensive, strategic) games form a class of modal models.

G. Bonanno. *Modal logic and game theory: Two alternative approaches*. Risk Decision and Policy 7 (2002).

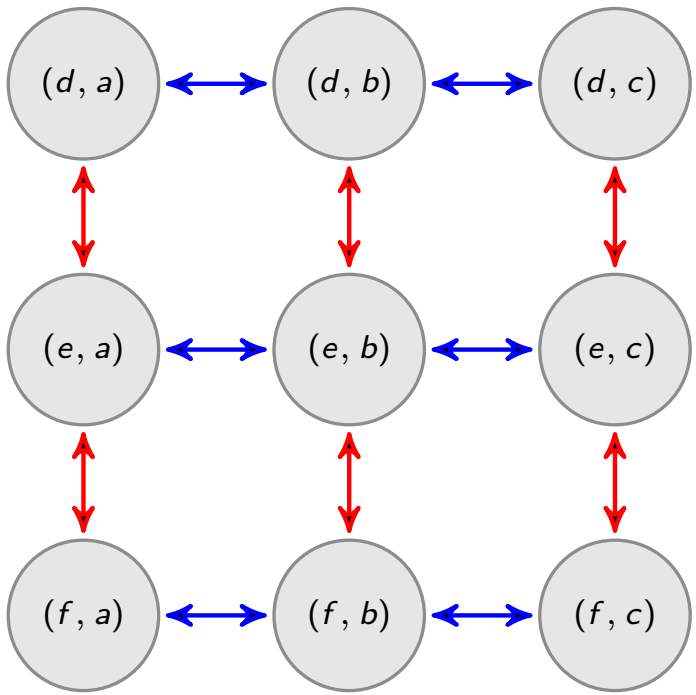
J. van Benthem. *Extensive Games as Process Models*. JOLLI 11 (2002).

What is a "good language" for expressing properties of these structures?

Epistemic Logic on Games

There is a Kripke structure "built in" a strategic game.

	a	b	c
d	(2,3)	(2,2)	(1,1)
e	(0,2)	(4,0)	(1,0)
f	(0,1)	(1,4)	(2,0)



Logical Characterization

- Define a logical language which can express actions, preferences, etc. and use game models (strategic or extensive) as *models* for the language. Then show there is a formula that **corresponds** to various solution concepts.

  G. Bonanno. *A Syntactic Approach to the Epistemic Foundations of Game Thoery*. Proceedings of LOFT, 2007.

- Focus on the iterative process of the solution concept (eg., backwards induction, iterated removal of strongly dominated strategies).

  J. van Benthem. *Rational Dynamics and Epistemic Logic in Games*. International Journal of Game Theory, 2005.

Example

- $[i]\varphi$: "φ holds in all states at least as preferable to the present one"
- $[\sigma]\varphi$: "if from here all players adhere to σ, then play will eventually end in a state in which φ holds"
- $[i, \sigma]\varphi$: "φ holds in all states that will be reached if all the players except possibly i play the strategy σ"

Example

- $[i]\varphi$: "φ holds in all states at least as preferable to the present one"
- $[\sigma]\varphi$: "if from here all players adhere to σ, then play will eventually end in a state in which φ holds"
- $[i, \sigma]\varphi$: "φ holds in all states that will be reached if all the players except possibly i play the strategy σ"

Fact: σ is a subgame perfect Nash equilibrium iff

$$\mathcal{F} \models \bigwedge_{i \in \mathcal{A}} (\langle i, \sigma \rangle [i]\varphi \rightarrow [\sigma]\varphi)$$

P. Harrenstein, W. van der Hoek, J-J. Meyer and C. Witteveen. *A Modal Characterization of Nash Equilibrium*. Fundamenta Informaticae 57 (2003).

More Examples

Nash Equilibrium:

$$w \models \bigwedge_{i \in \mathcal{A}} D_{-i} \Diamond_i^{\leq} w$$

More Examples

Nash Equilibrium:

$$w \models \bigwedge_{i \in \mathcal{A}} D_{-i} \Diamond_i^{\leq} w$$

Backwards Induction:

$Win_i := \mu P.(\text{end} \wedge \text{win}_\text{i}) \vee (\text{turn}_i \wedge \langle any(i) \rangle P) \vee (\text{turn}_j \wedge [any(j)]P)$

J. van Benthem. *Extensive Games as Process Models*. JOLLI 11 (2002).

More Examples

- (WR') $s_i^k \rightarrow \neg B_i(s_i^l \prec_i s_i^k)$
- (SR') $s_i^k \rightarrow \neg(B_i(s_i^l \preceq_i s_i^k) \wedge \neg B_i \neg(s_i^l \prec_i s_i^k)$

More Examples

- (WR') $s_i^k \rightarrow \neg B_i(s_i^l \prec_i s_i^k)$
- (SR') $s_i^k \rightarrow \neg(B_i(s_i^l \preceq_i s_i^k) \wedge \neg B_i \neg(s_i^l \prec_i s_i^k)$

Facts:

1. WR' characterizes Iterated Removal of Strongly Dominated Strategies (on $KD45$ frames).
2. SR' characterizes Iterated Removal of Weakly Dominated Strategies (on $S5$ frames).

with specific interpretations of the propositional variables

G. Bonanno. *A Syntactic Approach to Rationality in Games with Ordinal Payoffs*. 2007.

W. van der Hoek and M. Pauly. *Modal Logic for Games and Information*. in Handbook of Modal Logic (2007).

Dynamic Epistemic Analysis of Solution Concepts

From Games to Logic: Given some algorithmic algorithm defining a solution concept, try to find epistemic actions driving its dynamics

From Logic to Games: Any type of epistemic assertion defines and iterated solution process.

J. van Benthem. *Rational Dynamics and Epistemic Logic in Games*. IJGT, 2007.

Rationality Announcement

Weak Rationality: $w \models WR_j$ means $\bigwedge_{a \neq w(j)}$ 'j thinks that j's current action is at least as good for j as a.', where the a's run over the *current* model.

Theorem The following are equivalent for all states s in a full game model

1. s survives iterated removal of strongly dominated strategies
2. repeated successive **announcements** of WR for the players stabilizes at a submodel whose domain contains s.

Conclusions

1. This is just the beginning: there are many different epistemic/logical characterization results of solutions concepts (eg., common knowledge of "rationality" does/does not imply the backwards induction solution)
2. Many different style of analyses (probabilistic/logical)
3. ....

Thank You!