Logics of Action, Ability, Knowledge and Obligation

John F. Horty   Eric Pacuit

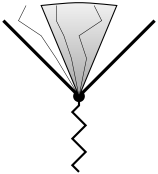
Department of Philosophy
University of Maryland

pacuit.org/esslli2019/epstit

August 8, 2019

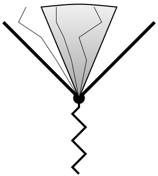
Actions

Actions *restrict* the set of possible future histories:

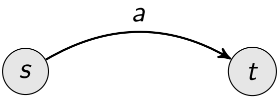


Actions

Actions *restrict* the set of possible future histories:



Actions as *transitions between states, or situations*:



Propositional Dynamic Logic

Language: The language of propositional dynamic logic is generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid [\pi]\varphi$$

where $p \in \mathsf{At}$ and α is generated by the following grammar:

$$a \mid \pi \cup \tau \mid \pi ; \tau \mid \pi^* \mid \varphi?$$

where $a \in \mathsf{Act}$ and φ is a formula.

Propositional Dynamic Logic

Language: The language of propositional dynamic logic is generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid [\pi]\varphi$$

where $p \in \mathsf{At}$ and α is generated by the following grammar:

$$a \mid \pi \cup \tau \mid \pi;\tau \mid \pi^* \mid \varphi?$$

where $a \in \mathsf{Act}$ and φ is a formula.

Semantics: $\mathcal{M} = \langle W, \{R_a \mid a \in \mathsf{P}\}, V\rangle$ where for each $a \in \mathsf{P}$, $R_a \subseteq W \times W$ and $V : \mathsf{At} \rightarrow \wp(W)$

Propositional Dynamic Logic

Language: The language of propositional dynamic logic is generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid [\pi]\varphi$$

where $p \in \mathrm{At}$ and α is generated by the following grammar:

$$a \mid \pi \cup \tau \mid \pi; \tau \mid \pi^* \mid \varphi?$$

where $a \in \mathrm{Act}$ and φ is a formula.

Semantics: $\mathcal{M} = \langle W, \{R_a \mid a \in \mathrm{P}\}, V\rangle$ where for each $a \in \mathrm{P}$, $R_a \subseteq W \times W$ and $V : \mathrm{At} \rightarrow \wp(W)$

$[\pi]\varphi$ means "after doing π, φ will be true"

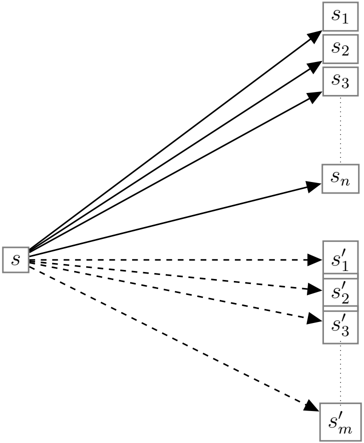
$\langle\pi\rangle\varphi$ means "after doing π, φ may be true"

$\mathcal{M}, w \models [\pi]\varphi$ iff for each v, if $wR_\pi v$ then $\mathcal{M}, v \models \varphi$

$\mathcal{M}, w \models \langle\pi\rangle\varphi$ iff there is a v such that $wR_\pi v$ and $\mathcal{M}, v \models \varphi$

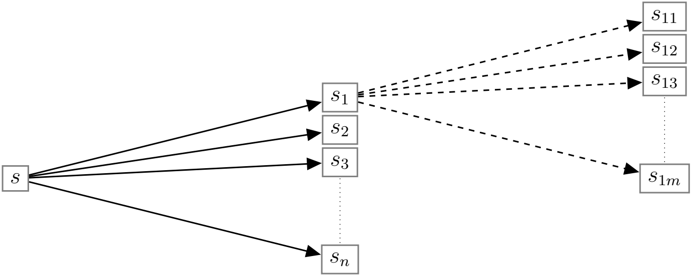
Union

$R_{\pi \cup \tau} := R_\pi \cup R_\tau$



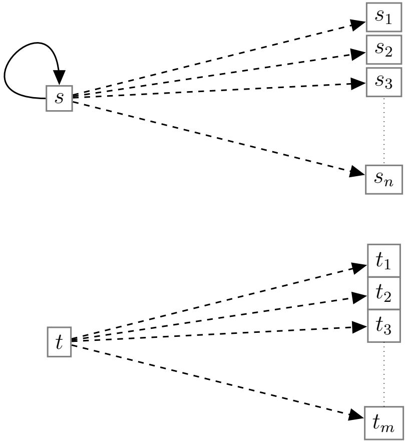
Sequence

$$R_{\pi;\tau} := R_\pi \circ R_\tau$$



Test

$$R_{\varphi?} = \{(w, w) \mid \mathcal{M}, w \models \varphi\}$$



Iteration

$$R_{\pi^*} := \cup_{n\geq 0} R_\pi^n$$

where $R_\pi^n = R \circ R_\pi^{n-1}$ and $R^0 = R$

Propositional Dynamic Logic

1. Axioms of propositional logic
2. $[\pi](\varphi \to \psi) \to ([\pi]\varphi \to [\pi]\psi)$
3. $[\pi \cup \tau]\varphi \leftrightarrow [\pi]\varphi \wedge [\tau]\varphi$
4. $[\pi; \tau]\varphi \leftrightarrow [\pi][\tau]\varphi$
5. $[\psi?]\varphi \leftrightarrow (\psi \to \varphi)$
6. $\varphi \wedge [\pi][\pi^*]\varphi \leftrightarrow [\pi^*]\varphi$
7. $\varphi \wedge [\pi^*](\varphi \to [\pi]\varphi) \to [\pi^*]\varphi$
8. Modus Ponens and Necessitation (for each program π)

Propositional Dynamic Logic

1. Axioms of propositional logic
2. $[\pi](\varphi \rightarrow \psi) \rightarrow ([\pi]\varphi \rightarrow [\pi]\psi)$
3. $[\pi \cup \tau]\varphi \leftrightarrow [\pi]\varphi \wedge [\tau]\varphi$
4. $[\pi; \tau]\varphi \leftrightarrow [\pi][\tau]\varphi$
5. $[\psi?]\varphi \leftrightarrow (\psi \rightarrow \varphi)$
6. $\varphi \wedge [\tau][\tau^*]\varphi \leftrightarrow [\tau^*]\varphi$ (Fixed-Point Axiom)
7. $\varphi \wedge [\tau^*](\varphi \rightarrow [\tau]\varphi) \rightarrow [\tau^*]\varphi$ (Induction Axiom)
8. Modus Ponens and Necessitation (for each program τ)

Actions and Ability

An early approach to interpret PDL as logic of actions was put forward by Krister Segerberg.

Segerberg adds an "agency" program to the PDL language δA where A is a formula.

K. Segerberg. *Bringing it about*. Journal of Philosophical Logic, 18(4), 327 - 347, 1989.

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

1. p is the computation according to some program π, and
2. π only terminates at states in which it is true that A

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

1. p is the computation according to some program π, and
2. π only terminates at states in which it is true that A

Interestingly, Segerberg also briefly considers a third condition:

3. p is optimal (in some sense: shortest, maximally efficient, most convenient, etc.) in the set of computations satisfying conditions (1) and (2).

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

1. p is the computation according to some program π, and
2. π only terminates at states in which it is true that A

Interestingly, Segerberg also briefly considers a third condition:

3. p is optimal (in some sense: shortest, maximally efficient, most convenient, etc.) in the set of computations satisfying conditions (1) and (2).

The axioms:

1. $[\delta A]A$
2. $[\delta A]B \to ([\delta B]C \to [\delta A]C)$

Coalitional Logic

M. Pauly. *A Modal Logic for Coalitional Powers in Games*. Journal of Logic and Computation, 12:1, pp. 149 - 166, 2002.

M. Pauly. *Logic for Social Software*. PhD Thesis, Institute for Logic, Language and Computation, 2001.

Strategic Game Forms

$$\langle N, \{S_i\}_{i \in N}, O, o \rangle$$

- N is a finite set of players;
- for each $i \in N$, S_i is a non-empty set (elements of which are called actions or strategies);
- O is a non-empty set (elements of which are called **outcomes**); and
- $o : \Pi_{i \in N} S_i \to O$ is a function assigning an outcome

		Bob	
		t_1	t_2
Ann	s_1	o_1	o_2
	s_2	o_2	o_3
	s_3	o_4	o_1

α-Effectivity

$S = \Pi_{i \in N} S_i$ are called **strategy profiles**. Given a strategy profile $s \in S$, let s_i denote i's component and s_{-i} the profile of strategies from s for all players except i.

A strategy for a coalition C is a sequence of strategies for each player in C, i.e., $s_C \in \Pi_{i \in C} S_i$ (similarly for $s_{\overline{C}}$, where $\overline{C}$ is $N - C$).

α-Effectivity

$S = \Pi_{i \in N} S_i$ are called **strategy profiles**. Given a strategy profile $s \in S$, let s_i denote i's component and s_{-i} the profile of strategies from s for all players except i.

A strategy for a coalition C is a sequence of strategies for each player in C, i.e., $s_C \in \Pi_{i \in C} S_i$ (similarly for $s_{\overline{C}}$, where $\overline{C}$ is $N - C$).

Suppose that $G = \langle N, \{S_i\}_{i \in N}, O, o \rangle$ be a strategic game form. An α-**effectivity function** is a map $E^{\alpha}_G : \wp(N) \to \wp(\wp(O))$ defined as follows: For all $C \subseteq N$, $X \in E^{\alpha}_G(C)$ iff there exists a strategy profile s_C such that for all $s_{\overline{C}} \in \Pi_{i \in N-C} S_i$, $o(s_C, s_{\overline{C}}) \in X$.

α-Effectivity vs. β-Effectivity

$\exists$ "something a player/a coalition *can* do" such that $\forall$ "actions of the other players/nature"...

α-Effectivity vs. β-Effectivity

$\exists$ "something a player/a coalition *can* do" such that $\forall$ "actions of the other players/nature"...

$\forall$ "(joint) actions of the other players", $\exists$ "something the agent/coalition can do"...

Bob

Ann	t_1	t_2
s_1	o_1	o_2
s_2	o_2	o_3
s_3	o_4	o_1

Bob

Ann	t_1	t_2
s_1	o_1	o_2
s_2	o_2	o_3
s_3	o_4	o_1

$$E^{\alpha}_{G_0}(\{A\}) = sup(\{\{o_1, o_2\}, \{o_2, o_3\}, \{o_1, o_4\}\})$$

Ann \ Bob	t_1	t_2
s_1	o_1	o_2
s_2	o_2	o_3
s_3	o_4	o_1

$$E^{\alpha}_{G_0}(\{A\}) = sup(\{\{o_1, o_2\}, \{o_2, o_3\}, \{o_1, o_4\}\})$$

$$E^{\alpha}_{G_0}(\{B\}) = sup(\{\{o_1, o_2, o_4\}, \{o_1, o_2, o_3\}\})$$

		Bob	
		t_1	t_2
Ann	s_1	o_1	o_2
	s_2	o_2	o_3
	s_3	o_4	o_1

$$E^{\alpha}_{G_0}(\{A\}) = sup(\{\{o_1, o_2\}, \{o_2, o_3\}, \{o_1, o_4\}\})$$

$$E^{\alpha}_{G_0}(\{B\}) = sup(\{\{o_1, o_2, o_4\}, \{o_1, o_2, o_3\}\})$$

$$E^{\alpha}_{G_0}(\{A, B\}) = sup(\{\{o_1\}, \{o_2\}, \{o_3\}, \{o_4\}\}) = \wp(O) - \emptyset$$

Bob

Ann	t_1	t_2
s_1	o_1	o_2
s_2	o_2	o_3
s_3	o_4	o_1

$$E^{\alpha}_{G_0}(\{A\}) = sup(\{\{o_1, o_2\}, \{o_2, o_3\}, \{o_1, o_4\}\})$$

$$E^{\alpha}_{G_0}(\{B\}) = sup(\{\{o_1, o_2, o_4\}, \{o_1, o_2, o_3\}\})$$

$$E^{\alpha}_{G_0}(\{A, B\}) = sup(\{\{o_1\}, \{o_2\}, \{o_3\}, \{o_4\}\}) = \wp(O) - \emptyset$$

$$E^{\alpha}_{G_0}(\emptyset) = \{\{o_1, o_2, o_3, o_4, o_5, o_6\}\}$$

Playable Effectivity Functions

1. (*Liveness*) For all $C \subseteq N$, $\emptyset \notin E(C)$

Playable Effectivity Functions

1. (*Liveness*) For all $C \subseteq N$, $\emptyset \notin E(C)$
2. (*Safety*) For all $C \subseteq N$, $O \in E(C)$

Playable Effectivity Functions

1. (*Liveness*) For all $C \subseteq N$, $\emptyset \not\in E(C)$
2. (*Safety*) For all $C \subseteq N$, $O \in E(C)$
3. (*N-maximality*) For all $X \subseteq O$, if $X \in E(N)$ then $\overline{X} \not\in E(\emptyset)$

Playable Effectivity Functions

1. (*Liveness*) For all $C \subseteq N$, $\emptyset \notin E(C)$
2. (*Safety*) For all $C \subseteq N$, $O \in E(C)$
3. (*N-maximality*) For all $X \subseteq O$, if $X \in E(N)$ then $\overline{X} \notin E(\emptyset)$
4. (*Outcome-monotonicity*) For all $X \subseteq X' \subseteq O$, and $C \subseteq N$, if $X \in E(C)$ then $X' \in E(C)$

Playable Effectivity Functions

1. (*Liveness*) For all $C \subseteq N$, $\emptyset \notin E(C)$

2. (*Safety*) For all $C \subseteq N$, $O \in E(C)$

3. (*N-maximality*) For all $X \subseteq O$, if $X \in E(N)$ then $\overline{X} \notin E(\emptyset)$

4. (*Outcome-monotonicity*) For all $X \subseteq X' \subseteq O$, and $C \subseteq N$, if $X \in E(C)$ then $X' \in E(C)$

5. (*Superadditivity*) For all subsets X_1, X_2 of O and sets of agents C_1, C_2, if $C_1 \cap C_2 = \emptyset$, $X_1 \in E(C_1)$ and $X_2 \in E(C_2)$, then $X_1 \cap X_2 \in E(C_1 \cup C_2)$

Characterizing Playable Effectivity Functions

Theorem (Pauly 2001; Goranko, Jamorga and Turrini 2013). If $E : \wp(N) \to \wp(\wp(O))$ is a function that satisfies the conditions 1-6 given above, then $E = E^{\alpha}_{G}$ for some strategic game form.

V. Goranko, W. Jamroga, and P. Turrini. *Strategic Games and Truly Playable Effectivity Functions*. Journal of Autonomous Agents and Multiagent Systems, 26(2), pgs. 288 - 314, 2013.

M. Pauly. *Logic for Social Software*. PhD Thesis, Institute for Logic, Language and Computation, 2001.

Coalitional Models

A coalitional logic model is a tuple $\mathcal{M} = \langle W, E, V\rangle$ where W is a set of states, $E : W \to (\wp(N) \to \wp(\wp(W)))$ assigns to each state a playable effectivity function, and $V : \text{At} \to \wp(W)$ is a valuation function.

$$\mathcal{M}, w \models [C]\varphi \text{ iff } [\![\varphi]\!]_{\mathcal{M}} = \{w \mid \mathcal{M}, w \models \varphi\} \in E(w)(C)$$

Coalitional Logic: Axiomatics

1. (*Liveness*) For all $C \subseteq N$, $\emptyset \notin E(C)$

2. (*Safety*) For all $C \subseteq N$, $O \in E(C)$

3. (*N-maximality*) For all $X \subseteq O$, if $X \in E(N)$ then $\overline{X} \notin E(\emptyset)$

4. (*Outcome-monotonicity*) For all $X \subseteq X' \subseteq O$, and $C \subseteq N$, if $X \in E(C)$ then $X' \in E(C)$

5. (*Superadditivity*) For all subsets X_1, X_2 of O and sets of agents C_1, C_2, if $C_1 \cap C_2 = \emptyset$, $X_1 \in E(C_1)$ and $X_2 \in E(C_2)$, then $X_1 \cap X_2 \in E(C_1 \cup C_2)$

Coalitional Logic: Axiomatics

1. (*Liveness*) $\neg[C]\bot$

2. (*Safety*) For all $C \subseteq N$, $O \in E(C)$

3. (*N-maximality*) For all $X \subseteq O$, if $X \in E(N)$ then $\overline{X} \notin E(\emptyset)$

4. (*Outcome-monotonicity*) For all $X \subseteq X' \subseteq O$, and $C \subseteq N$, if $X \in E(C)$ then $X' \in E(C)$

5. (*Superadditivity*) For all subsets X_1, X_2 of O and sets of agents C_1, C_2, if $C_1 \cap C_2 = \emptyset$, $X_1 \in E(C_1)$ and $X_2 \in E(C_2)$, then $X_1 \cap X_2 \in E(C_1 \cup C_2)$

Coalitional Logic: Axiomatics

1. (*Liveness*) $\neg[C]\bot$

2. (*Safety*) $[C]\top$

3. (*N-maximality*) For all $X \subseteq O$, if $X \in E(N)$ then $\overline{X} \notin E(\emptyset)$

4. (*Outcome-monotonicity*) For all $X \subseteq X' \subseteq O$, and $C \subseteq N$, if $X \in E(C)$ then $X' \in E(C)$

5. (*Superadditivity*) For all subsets X_1, X_2 of O and sets of agents C_1, C_2, if $C_1 \cap C_2 = \emptyset$, $X_1 \in E(C_1)$ and $X_2 \in E(C_2)$, then $X_1 \cap X_2 \in E(C_1 \cup C_2)$

Coalitional Logic: Axiomatics

1. (*Liveness*) $\neg[C]\bot$

2. (*Safety*) $[C]\top$

3. (*N-maximality*) $[N]\varphi \to \neg[\emptyset]\neg\varphi$

4. (*Outcome-monotonicity*) For all $X \subseteq X' \subseteq O$, and $C \subseteq N$, if $X \in E(C)$ then $X' \in E(C)$

5. (*Superadditivity*) For all subsets X_1, X_2 of O and sets of agents C_1, C_2, if $C_1 \cap C_2 = \emptyset$, $X_1 \in E(C_1)$ and $X_2 \in E(C_2)$, then $X_1 \cap X_2 \in E(C_1 \cup C_2)$

Coalitional Logic: Axiomatics

1. (*Liveness*) $\neg[C]\bot$

2. (*Safety*) $[C]\top$

3. (*N-maximality*) $[N]\varphi \rightarrow \neg[\emptyset]\neg\varphi$

4. (*Outcome-monotonicity*) $[C](\varphi \wedge \psi) \rightarrow ([C]\varphi \wedge [C]\psi)$

5. (*Superadditivity*) For all subsets X_1, X_2 of O and sets of agents C_1, C_2, if $C_1 \cap C_2 = \emptyset$, $X_1 \in E(C_1)$ and $X_2 \in E(C_2)$, then $X_1 \cap X_2 \in E(C_1 \cup C_2)$

Coalitional Logic: Axiomatics

1. (*Liveness*) $\neg[C]\bot$

2. (*Safety*) $[C]\top$

3. (*N-maximality*) $[N]\varphi \to \neg[\emptyset]\neg\varphi$

4. (*Outcome-monotonicity*) $[C](\varphi \wedge \psi) \to ([C]\varphi \wedge [C]\psi)$

5. (*Superadditivity*) $([C_1]\varphi_1 \wedge [C_2]\varphi_2) \to [C_1 \cup C_2](\varphi_1 \wedge \varphi_2)$, where $C_1 \cap C_2 = \emptyset$

From Temporal Logic to Strategy Logic

From Temporal Logic to Strategy Logic

- *Linear Time Temporal Logic*: Reasoning about computation paths:

$F\varphi$: φ is true some time in *the* future.

A. Pnuelli. *A Temporal Logic of Programs*. in *Proc. 18th IEEE Symposium on Foundations of Computer Science* (1977).

From Temporal Logic to Strategy Logic

- *Linear Time Temporal Logic*: Reasoning about computation paths:

  $F\varphi$: φ is true some time in *the* future.

A. Pnuelli. *A Temporal Logic of Programs*. in *Proc. 18th IEEE Symposium on Foundations of Computer Science* (1977).

- *Branching Time Temporal Logic*: Allows quantification over paths:

  $\exists F\varphi$: there is a path in which φ is eventually true.

E. M. Clarke and E. A. Emerson. *Design and Synthesis of Synchronization Skeletons using Branching-time Temproal-logic Specifications*. In *Proceedings Workshop on Logic of Programs*, LNCS (1981).

From Temporal Logic to Strategy Logic

- *Coalitional Logic*: Reasoning about (local) group power.

  $[C]\varphi$: coalition C has a **joint action** to bring about φ.

M. Pauly. *A Modal Logic for Coalition Powers in Games*. *Journal of Logic and Computation* **12** (2002).

From Temporal Logic to Strategy Logic

- *Coalitional Logic*: Reasoning about (local) group power.

  $[C]\varphi$: coalition C has a **joint action** to bring about φ.

M. Pauly. *A Modal Logic for Coalition Powers in Games*. *Journal of Logic and Computation* **12** (2002).

- *Alternating-time Temporal Logic*: Reasoning about (local and global) group power:

  $\langle\langle A\rangle\rangle G\varphi$: The coalition A has a **joint action** to ensure that φ will remain true.

R. Alur, T. Henzinger and O. Kupferman. *Alternating-time Temproal Logic*. *Jouranl of the ACM* (2002).

Example: Suppose that there are two agents: a server (s) and a client (c). The client asks to set the value of x and the server can either grant or deny the request. Assume the agents make simultaneous moves.

Example: Suppose that there are two agents: a server (s) and a client (c). The client asks to set the value of x and the server can either grant or deny the request. Assume the agents make simultaneous moves.

	deny	*grant*
*set*1		
*set*2		

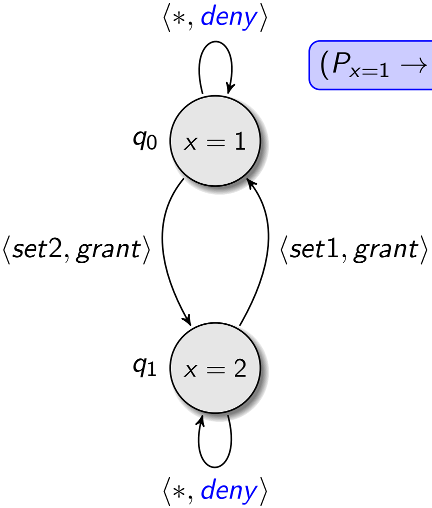
Example: Suppose that there are two agents: a server (s) and a client (c). The client asks to set the value of x and the server can either grant or deny the request. Assume the agents make simultaneous moves.

	deny	*grant*
*set*1		$q_0 \Rightarrow q_0,\ q_1 \Rightarrow q_0$
*set*2		$q_0 \Rightarrow q_1,\ q_1 \Rightarrow q_1$

Example: Suppose that there are two agents: a server (s) and a client (c). The client asks to set the value of x and the server can either grant or deny the request. Assume the agents make simultaneous moves.

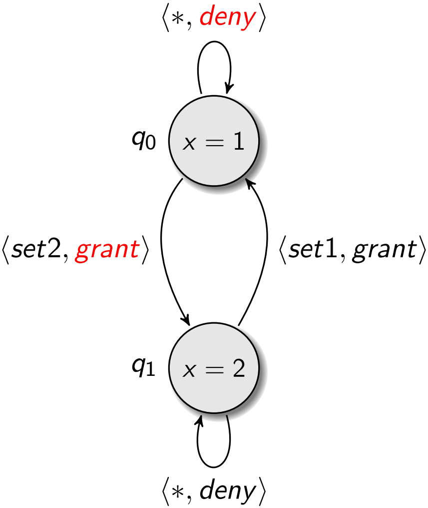
	deny	*grant*
$set1$	$q \Rightarrow q$	$q_0 \Rightarrow q_0,\ q_1 \Rightarrow q_0$
$set2$	$q \Rightarrow q$	$q_0 \Rightarrow q_1,\ q_1 \Rightarrow q_1$

Multi-agent Transition Systems



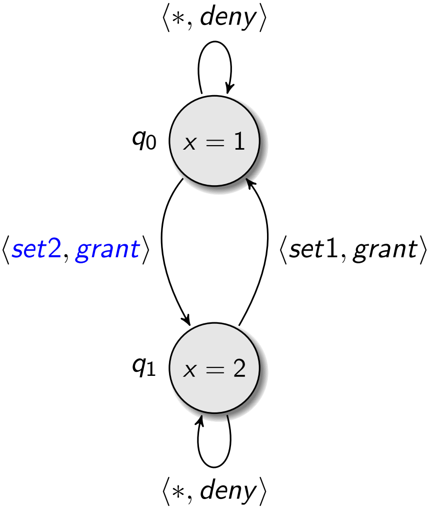
$(P_{x=1} \rightarrow [s]P_{x=1}) \wedge (P_{x=2} \rightarrow [s]P_{x=2})$

Multi-agent Transition Systems



$$P_{x=1} \rightarrow \neg[s]P_{x=2}$$

Multi-agent Transition Systems



$P_{x=1} \rightarrow [s, c]P_{x=2}$

Axioms:

(TAUT) Enough propositional tautologies.
($\bot$) $\neg\langle\langle A\rangle\rangle\bigcirc\bot$
($\top$) $\langle\langle A\rangle\rangle\bigcirc\top$
(Σ) $\neg\langle\langle\varnothing\rangle\rangle\bigcirc\neg\varphi\to\langle\langle\Sigma\rangle\rangle\bigcirc\varphi$
(S) $\langle\langle A_1\rangle\rangle\bigcirc\varphi_1\wedge\langle\langle A_2\rangle\rangle\bigcirc\varphi_2\to\langle\langle A_1\cup A_2\rangle\rangle\bigcirc(\varphi_1\wedge\varphi_2)$ for disjoint A_1 and A_2
(FP$_\Box$) $\langle\langle A\rangle\rangle\Box\varphi\leftrightarrow\varphi\wedge\langle\langle A\rangle\rangle\bigcirc\langle\langle A\rangle\rangle\Box\varphi$
(GFP$_\Box$) $\langle\langle\varnothing\rangle\rangle\Box(\theta\to(\varphi\wedge\langle\langle A\rangle\rangle\bigcirc\theta))\to\langle\langle\varnothing\rangle\rangle\Box(\theta\to\langle\langle A\rangle\rangle\Box\varphi)$
(FP$_\mathcal{U}$) $\langle\langle A\rangle\rangle\varphi_1\mathcal{U}\varphi_2\leftrightarrow\varphi_2\vee(\varphi_1\wedge\langle\langle A\rangle\rangle\bigcirc\langle\langle A\rangle\rangle\varphi_1\mathcal{U}\varphi_2)$
(LFP$_\mathcal{U}$) $\langle\langle\varnothing\rangle\rangle\Box((\varphi_2\vee(\varphi_1\wedge\langle\langle A\rangle\rangle\bigcirc\theta))\to\theta)\to\langle\langle\varnothing\rangle\rangle\Box(\langle\langle A\rangle\rangle\varphi_1\mathcal{U}\varphi_2\to\theta)$

Rules of inference:

(Modus Ponens) $\dfrac{\varphi_1,\varphi_1\to\varphi_2}{\varphi_2}$

($\langle\langle A\rangle\rangle\bigcirc$-Monotonicity) $\dfrac{\varphi_1\to\varphi_2}{\langle\langle A\rangle\rangle\bigcirc\varphi_1\to\langle\langle A\rangle\rangle\bigcirc\varphi_2}$

($\langle\langle\varnothing\rangle\rangle\Box$-Necessitation) $\dfrac{\varphi}{\langle\langle\varnothing\rangle\rangle\Box\varphi}$

V. Goranko and G. van Drimmelen. *Complete Axiomatization and Decidability of the Alternating-time Temporal Logic*. Theoretical Computer Science, 353(1-3), 93 - 117, 2006.

Comparing Semantics

V. Goranko and W. Jamroga. *Comparing Semantics for Logics of Multi-agent Systems*. Synthese 139(2), 241 - 280, 2004.

J. Broersen, A. Herzig and N. Troquard. *From Coalition Logic to STIT*. Proceedings of the Third International Workshop on Logic and Communication in Multi-Agent Systems (LCMAS 2005).

J. Broersen, A. Herzig and N. Troquard. *Embedding alternating-time temporal logic in strategic STIT logic of agency*. Journal of Logic and Computation, 16(5), 559 - 578, 2006.

R. Ciuni and E. Lorini. *Comparing semantics for temporal STIT logic*. Logique & Analyse, 61(243), 2018.

Logic of Knowledge (and Belief)

Epistemic Logic

Let $K_a P$ informally mean "agent *a* knows that *P* (is true)".

Epistemic Logic

Let $K_a P$ informally mean "agent a knows that P (is true)".

$K_a(P \rightarrow Q)$: "Ann knows that P implies Q"

Epistemic Logic

Let K_aP informally mean "agent a knows that P (is true)".

$K_a(P \to Q)$: "Ann knows that P implies Q"

$K_aP \vee \neg K_aP$: "either Ann does or does not know P"

Epistemic Logic

Let K_aP informally mean "agent a knows that P (is true)".

$K_a(P \to Q)$: "Ann knows that P implies Q"

$K_aP \vee \neg K_aP$: "either Ann does or does not know P"

$K_aP \vee K_a\neg P$: "Ann knows whether P is true"

Epistemic Logic

Let K_aP informally mean "agent a knows that P (is true)".

$K_a(P \to Q)$: "Ann knows that P implies Q"

$K_aP \vee \neg K_aP$: "either Ann does or does not know P"

$K_aP \vee K_a\neg P$: "Ann knows whether P is true"

$\neg K_a \neg P$: "P is an epistemic possibility for Ann"

Epistemic Logic

Let $K_a P$ informally mean "agent *a* knows that P (is true)".

$K_a(P \to Q)$: "Ann knows that P implies Q"

$K_a P \vee \neg K_a P$: "either Ann does or does not know P"

$K_a P \vee K_a \neg P$: "Ann knows whether P is true"

$\neg K_a \neg P$: "P is an epistemic possibility for Ann"

$K_a K_a P$: "Ann knows that she knows that P"

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.

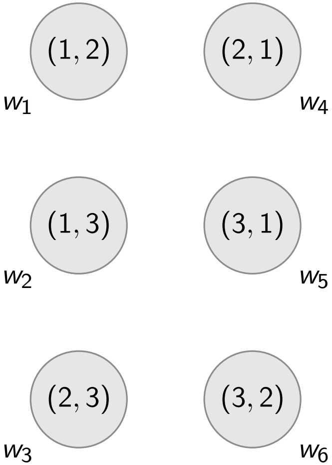
What are the relevant states?

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?

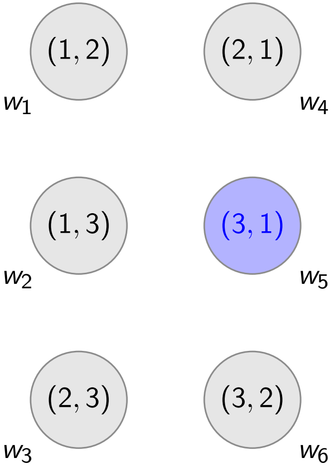


Example

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

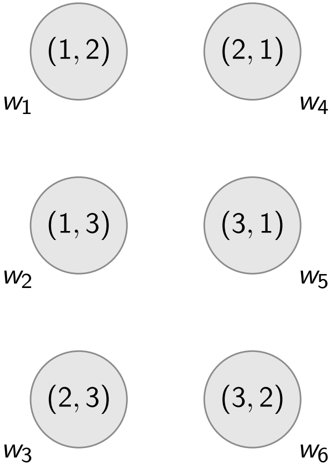


Example

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?

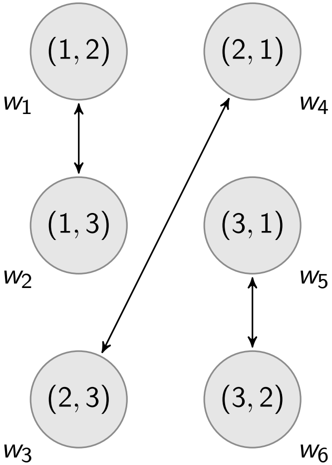


Example

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?

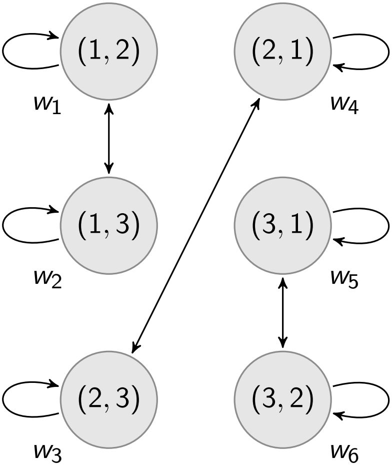


Example

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

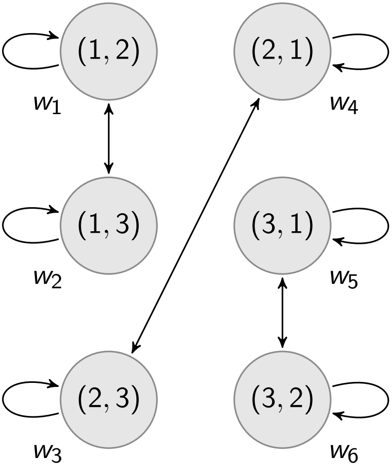
Suppose there are three cards: 1, 2 and 3.

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

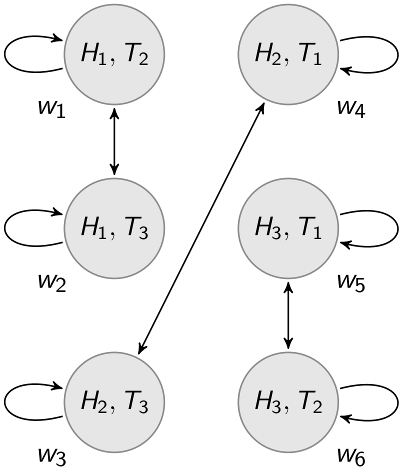
Suppose there are three cards: 1, 2 and 3.

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"

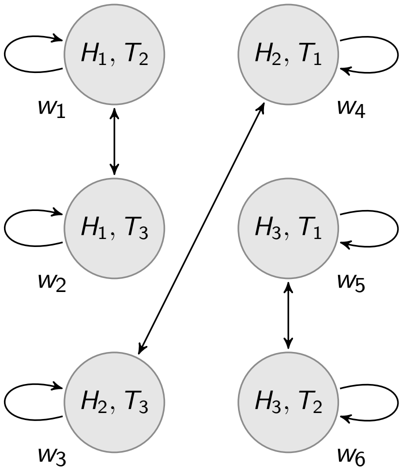
Eg., $V(H_1) = \{w_1, w_2\}$



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.

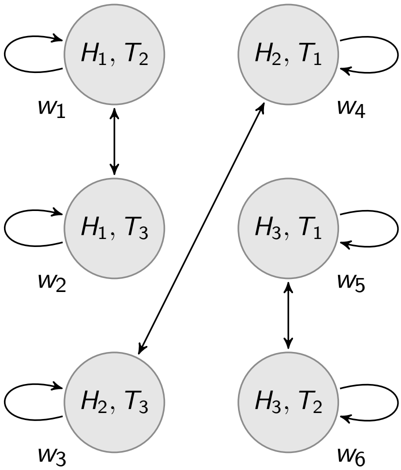


Example

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.

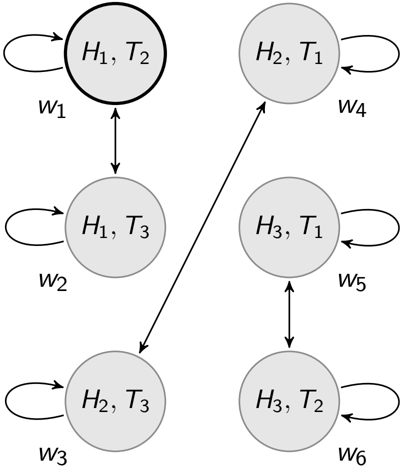


Example

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.

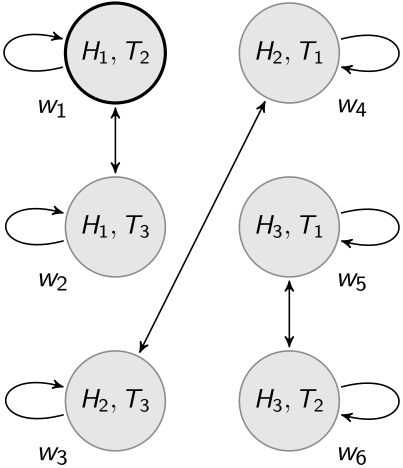


Example

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.

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

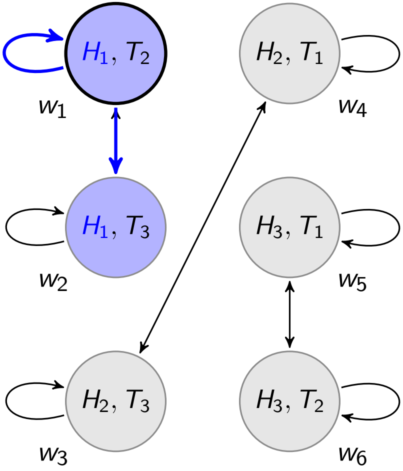


Example

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.

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



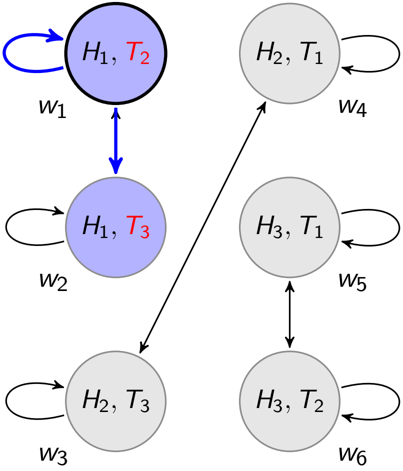
Example

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.

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

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

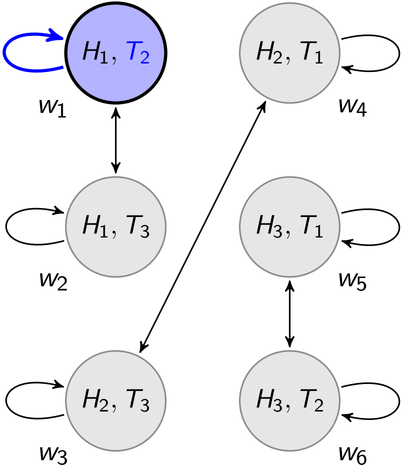


Example

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.

$\mathcal{M}, w_1 \models \neg K_a \neg T_2$

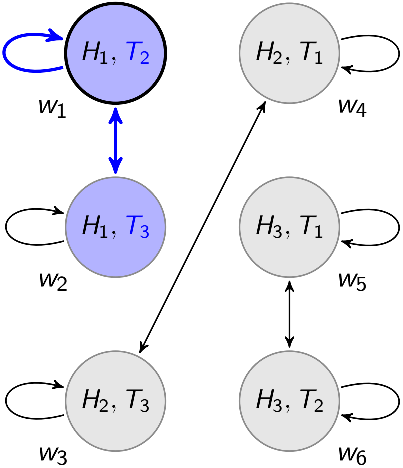


Example

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.

$\mathcal{M}, w_1 \models K_a(T_2 \vee T_3)$



Multiagent Epistemic Logic

Many of the examples we are interested in involve more than one agent!

Multiagent Epistemic Logic

Many of the examples we are interested in involve more than one agent!

$K_a P$ means "Ann knows P"

$K_b P$ means "Bob knows P"

Multiagent Epistemic Logic

Many of the examples we are interested in involve more than one agent!

$K_a P$ means "Ann knows P"

$K_b P$ means "Bob knows P"

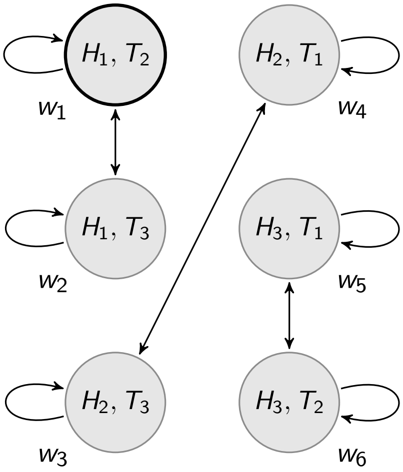
- $K_a K_b \varphi$: "Ann knows that Bob knows φ"
- $K_a(K_b\varphi \vee K_b\neg\varphi)$: "Ann knows that Bob knows whether φ
- $\neg K_b K_a K_b(\varphi)$: "Bob does not know that Ann knows that Bob knows that φ"

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.

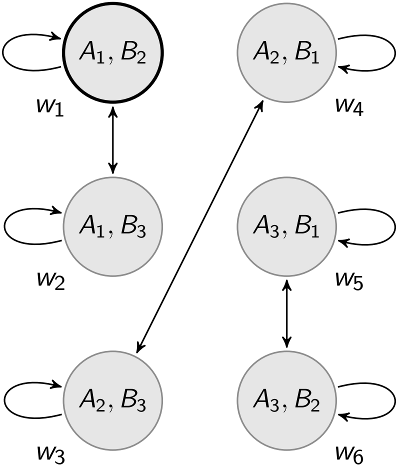


Example

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.

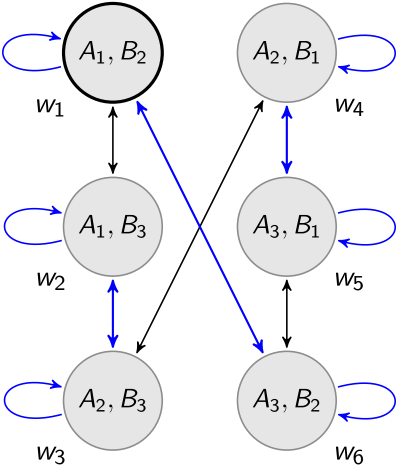


Example

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.

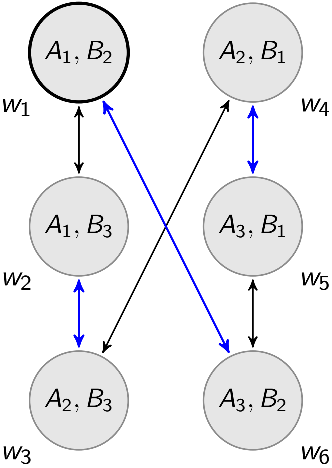


Example

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.



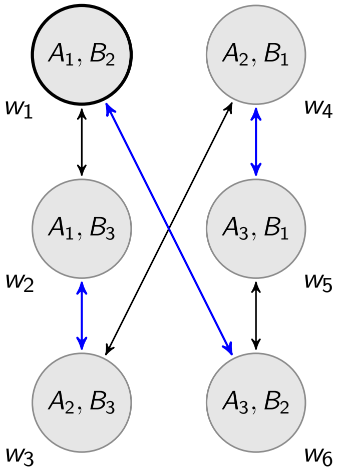
Example

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.

$\mathcal{M}, w_1 \models K_b(K_a A_1 \vee K_a \neg A_1)$



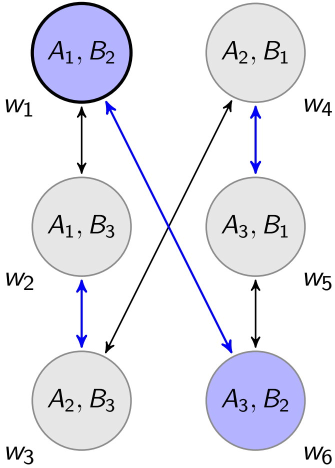
Example

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.

$\mathcal{M}, w_1 \models K_b(K_a A_1 \vee K_a \neg A_1)$



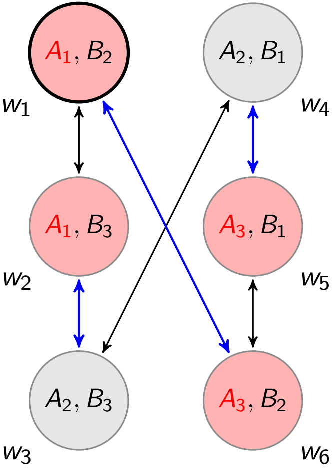
Example

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.

$\mathcal{M}, w_1 \models K_b(K_a A_1 \vee K_a \neg A_1)$



College Park and Amsterdam

Suppose agent c, who lives in College Park, knows that agent a lives in Amsterdam. Let r stand for 'it's raining in Amsterdam'. Although c doesn't know whether it's raining in Amsterdam, c knows that a knows whether it's raining there:

College Park and Amsterdam

Suppose agent c, who lives in College Park, knows that agent a lives in Amsterdam. Let r stand for 'it's raining in Amsterdam'. Although c doesn't know whether it's raining in Amsterdam, c knows that a knows whether it's raining there:

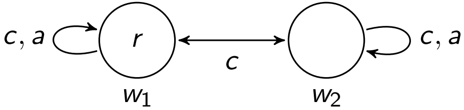
$$\neg(K_c r \vee K_c \neg r) \wedge K_c(K_a r \vee K_a \neg r).$$

College Park and Amsterdam

Suppose agent c, who lives in College Park, knows that agent a lives in Amsterdam. Let r stand for 'it's raining in Amsterdam'. Although c doesn't know whether it's raining in Amsterdam, c knows that a knows whether it's raining there:

$$\neg(K_c r \vee K_c \neg r) \wedge K_c(K_a r \vee K_a \neg r).$$

The following picture depicts a situation in which this is true, where an arrow represents *compatibility with one's knowledge*:

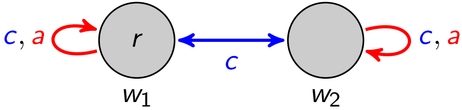


College Park and Amsterdam

Suppose agent c, who lives in College Park, knows that agent a lives in Amsterdam. Let r stand for 'it's raining in Amsterdam'. Although c doesn't know whether it's raining in Amsterdam, c knows that a knows whether it's raining there:

$$\neg(K_c r \vee K_c \neg r) \wedge K_c(K_a r \vee K_a \neg r).$$

The following picture depicts a situation in which this is true, where an arrow represents *compatibility with one's knowledge*:

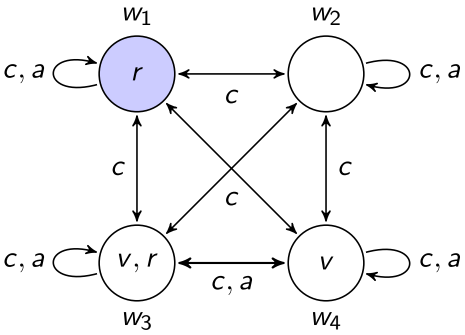


Now suppose that agent c doesn't know whether agent a has left Amsterdam for a vacation. (Let v stand for 'a has left Amsterdam on vacation'.) Agent c knows that if a is not on vacation, then a knows whether it's raining in Amsterdam; but if a is on vacation, then a won't bother to follow the weather.

$$K_c(\neg v \to (K_a r \vee K_a \neg r)) \wedge K_c(v \to \neg(K_a r \vee K_a \neg r)).$$

Now suppose that agent c doesn't know whether agent a has left Amsterdam for a vacation. (Let v stand for 'a has left Amsterdam on vacation'.) Agent c knows that if a is not on vacation, then a knows whether it's raining in Amsterdam; but if a is on vacation, then a won't bother to follow the weather.

$$K_c(\neg v \rightarrow (K_a r \vee K_a \neg r)) \wedge K_c(v \rightarrow \neg(K_a r \vee K_a \neg r)).$$



Epistemic Logic: The Language

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

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

Epistemic Logic: The Language

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

$$\varphi \; := \; p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_a\varphi$$

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

Epistemic Logic: The Language

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

$$\varphi \;:=\; p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_a\varphi$$

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

Epistemic Logic: The Language

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

$$\varphi \;:=\; p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_a\varphi$$

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

Epistemic Logic: The Language

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

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

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

Epistemic Logic: The Language

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

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

$K_a(p \rightarrow q)$: "Ann knows that p implies q"

$K_a p \vee \neg K_a p$:

$K_a p \vee K_a \neg p$:

$L_a\varphi$:

$K_a L_a \varphi$:

Epistemic Logic: The Language

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

$$\varphi \;:=\; p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_a\varphi$$

$K_a(p \to q)$: "Ann knows that p implies q"

$K_a p \vee \neg K_a p$: "either Ann does or does not know p"

$K_a p \vee K_a \neg p$: "Ann knows whether p is true"

$L_a\varphi$:

$K_a L_a\varphi$:

Epistemic Logic: The Language

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

$$\varphi \;:=\; p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_a\varphi$$

$K_a(p \rightarrow q)$: "Ann knows that p implies q"

$K_a p \vee \neg K_a p$: "either Ann does or does not know p"

$K_a p \vee K_a \neg p$: "Ann knows whether p is true"

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

$K_a L_a\varphi$: "Ann knows that she thinks φ is possible"

Epistemic Logic: Kripke Models

$$\mathcal{M} = \langle W, \{R_a\}_{a \in \mathcal{A}}, V \rangle$$

Epistemic Logic: Kripke Models

$$\mathcal{M} = \langle W, \{R_a\}_{a\in\mathcal{A}}, V\rangle$$

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

Epistemic Logic: Kripke Models

$$\mathcal{M} = \langle W, \{R_a\}_{a \in \mathcal{A}}, V \rangle$$

- $W \neq \emptyset$ is the set of all relevant situations (states of affairs, possible worlds)
- $R_a \subseteq W \times W$ *represents* the agent a's knowledge

Epistemic Logic: Kripke Models

$$\mathcal{M} = \langle W, \{R_a\}_{a\in\mathcal{A}}, V\rangle$$

- $W \neq \emptyset$ is the set of all relevant situations (states of affairs, possible worlds)
- $R_a \subseteq W \times W$ *represents* the agent a's knowledge
- $V : \mathsf{At} \rightarrow \wp(W)$ is a valuation function assigning propositional variables to worlds

Epistemic Logic: Truth in a Model

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

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

Epistemic Logic: Truth in a Model

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

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

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

Epistemic Logic: Truth in a Model

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

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

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

Epistemic Logic: Truth in a Model

Given $\varphi \in \mathcal{L}$, a Kripke model $\mathcal{M} = \langle W, \{R_a\}_{a \in \mathcal{A}}, V \rangle$ 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 $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K_a\varphi$ if for each $v \in W$, if wR_av, then $\mathcal{M}, v \models \varphi$

Epistemic Logic: Truth in a Model

Given $\varphi \in \mathcal{L}$, a Kripke model $\mathcal{M} = \langle W, \{R_a\}_{a \in \mathcal{A}}, V \rangle$ 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 $\mathcal{M}, w \models \psi$
- ✓ $\mathcal{M}, w \models K_a\varphi$ if for each $v \in W$, if wR_av, then $\mathcal{M}, v \models \varphi$

Epistemic Logic: Truth in a Model

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

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

- ✓ $\mathcal{M}, w \models p$ iff $w \in V(p)$ (with $p \in \mathsf{At}$)
- ✓ $\mathcal{M}, w \models \neg\varphi$ if $\mathcal{M}, w \not\models \varphi$
- ✓ $\mathcal{M}, w \models \varphi \wedge \psi$ if $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- ✓ $\mathcal{M}, w \models K_a\varphi$ if for each $v \in W$, if wR_av, then $\mathcal{M}, v \models \varphi$
- ✓ $\mathcal{M}, w \models L_a\varphi$ if there exists a $v \in W$ such that wR_av and $\mathcal{M}, v \models \varphi$

$K_a\varphi$: "Agent *a* is *informed* that φ", "Agent *a knows* that φ"

$K_a\varphi$: "Agent a is *informed* that φ", "Agent a *knows* that φ"

$\mathcal{M}, w \models K_a\varphi$ iff for all $v \in W$, if wR_av then $\mathcal{M}, v \models \varphi$

I.e., $R_a(w) = \{v \mid wR_av\} \subseteq [\![\varphi]\!]_{\mathcal{M}} = \{v \mid \mathcal{M}, v \models \varphi\}$:

$K_a\varphi$: "Agent a is *informed* that φ", "Agent a *knows* that φ"

$\mathcal{M}, w \models K_a\varphi$ iff for all $v \in W$, if wR_av then $\mathcal{M}, v \models \varphi$
I.e., $R_a(w) = \{v \mid wR_av\} \subseteq [\![\varphi]\!]_{\mathcal{M}} = \{v \mid \mathcal{M}, v \models \varphi\}$:

- wR_av if "everything a knows in state w is true in v

$K_a\varphi$: "Agent a is *informed* that φ", "Agent a *knows* that φ"

$\mathcal{M}, w \models K_a\varphi$ iff for all $v \in W$, if wR_av then $\mathcal{M}, v \models \varphi$

I.e., $R_a(w) = \{v \mid wR_av\} \subseteq [\![\varphi]\!]_{\mathcal{M}} = \{v \mid \mathcal{M}, v \models \varphi\}$:

- wR_av if "everything a knows in state w is true in v
- wR_av if "agent a has the same experiences and memories in both w and v"

$K_a\varphi$: "Agent a is *informed* that φ", "Agent a *knows* that φ"

$\mathcal{M}, w \models K_a\varphi$ iff for all $v \in W$, if wR_av then $\mathcal{M}, v \models \varphi$
I.e., $R_a(w) = \{v \mid wR_av\} \subseteq [\![\varphi]\!]_{\mathcal{M}} = \{v \mid \mathcal{M}, v \models \varphi\}$:

- wR_av if "everything a knows in state w is true in v
- wR_av if "agent a has the same experiences and memories in both w and v"
- wR_av if "agent a has cannot *rule-out* v, given her evidence and observations (at state w)"

$K_a\varphi$: "Agent a is *informed* that φ", "Agent a *knows* that φ"

$\mathcal{M}, w \models K_a\varphi$ iff for all $v \in W$, if wR_av then $\mathcal{M}, v \models \varphi$
I.e., $R_a(w) = \{v \mid wR_av\} \subseteq [\![\varphi]\!]_{\mathcal{M}} = \{v \mid \mathcal{M}, v \models \varphi\}$:

- wR_av if "everything a knows in state w is true in v
- wR_av if "agent a has the same experiences and memories in both w and v"
- wR_av if "agent a has cannot *rule-out* v, given her evidence and observations (at state w)"
- wR_av if "agent a is in the same *local state* in w and v"

$L_a\varphi$ iff there is a $v \in W$ such that $\mathcal{M}, v \models \varphi$

I.e., $R_a(w) = \{v \mid wR_av\} \cap [\![\varphi]\!]_{\mathcal{M}} = \{v \mid \mathcal{M}, v \models \varphi\} \neq \emptyset$

$L_a\varphi$ iff there is a $v \in W$ such that $\mathcal{M}, v \models \varphi$

I.e., $R_a(w) = \{v \mid wR_a v\} \cap [\![\varphi]\!]_{\mathcal{M}} = \{v \mid \mathcal{M}, v \models \varphi\} \neq \emptyset$

- $L_a\varphi$: "Agent a thinks that φ might be true."
- $L_a\varphi$: "Agent a considers φ possible."

$L_a\varphi$ iff there is a $v \in W$ such that $\mathcal{M}, v \models \varphi$

I.e., $R_a(w) = \{v \mid wR_av\} \cap [\![\varphi]\!]_{\mathcal{M}} = \{v \mid \mathcal{M}, v \models \varphi\} \neq \emptyset$

- ~~$L_a\varphi$: "Agent a thinks that φ might be true."~~
- ~~$L_a\varphi$: "Agent a considers φ possible."~~
- $L_a\varphi$: "(*according to the model*), φ is consistent with what a knows ($\neg K_a \neg\varphi$)".

Taking Stock

Multi-agent language: $\varphi \;:=\; p \mid \neg\varphi \mid \varphi\wedge\psi \mid \Box_i\varphi$

- $\Box_i\varphi$: "agent i knows that φ" (write $K_i\varphi$ for $\Box_i\varphi$)
- $\Box_i\varphi$: "agent i believes that φ" (write $B_i\varphi$ for $\Box_i\varphi$)

Kripke Models: $\mathcal{M} = \langle W, \{R_i\}_{i\in Agt}, V\rangle$

Truth: $\mathcal{M}, w \models \Box_i\varphi$ iff for all $v \in W$, if wR_iv then $\mathcal{M}, v \models \varphi$

Modal Formula	Corresponding Property

Modal Formula	Corresponding Property
$\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$	—

Modal Formula	Corresponding Property
$\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$	—
$\Box\varphi \rightarrow \varphi$	Reflexive

Modal Formula	Corresponding Property
$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$	—
$\Box\varphi \to \varphi$	Reflexive
$\Box\varphi \to \Box\Box\varphi$	Transitive

Modal Formula	Corresponding Property
$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$	—
$\Box\varphi \to \varphi$	Reflexive
$\Box\varphi \to \Box\Box\varphi$	Transitive
$\neg\Box\varphi \to \Box\neg\Box\varphi$	Euclidean

Modal Formula	Corresponding Property
$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$	—
$\Box\varphi \to \varphi$	Reflexive
$\Box\varphi \to \Box\Box\varphi$	Transitive
$\neg\Box\varphi \to \Box\neg\Box\varphi$	Euclidean
$\neg\Box\bot$	Serial

The Logic **S5**

The logic **S5** contains the following axioms and rules:

Pc	Axiomatization of Propositional Calculus
K	$K(\varphi \to \psi) \to (K\varphi \to K\psi)$
T	$K\varphi \to \varphi$
4	$K\varphi \to KK\varphi$
5	$\neg K\varphi \to K\neg K\varphi$
MP	$\dfrac{\varphi \qquad \varphi \to \psi}{\psi}$
Nec	$\dfrac{\varphi}{K\psi}$

The Logic **S5**

The logic **S5** contains the following axioms and rules:

Pc	Axiomatization of Propositional Calculus
K	$K(\varphi \to \psi) \to (K\varphi \to K\psi)$
T	$K\varphi \to \varphi$
4	$K\varphi \to KK\varphi$
5	$\neg K\varphi \to K\neg K\varphi$
MP	$\dfrac{\varphi \qquad \varphi \to \psi}{\psi}$
Nec	$\dfrac{\varphi}{K\psi}$

Theorem

S5 *is sound and strongly complete with respect to the class of Kripke frames with equivalence relations.*

The Logic **KD45**

The logic **S5** contains the following axioms and rules:

Pc	Axiomatization of Propositional Calculus
K	$B(\varphi \to \psi) \to (B\varphi \to B\psi)$
D	$\neg B\bot \quad (B\varphi \to \neg B\neg\varphi)$
4	$B\varphi \to BB\varphi$
5	$\neg B\varphi \to B\neg B\varphi$
MP	$\dfrac{\varphi \qquad \varphi \to \psi}{\psi}$
Nec	$\dfrac{\varphi}{B\psi}$

The Logic **KD45**

The logic **S5** contains the following axioms and rules:

Pc	Axiomatization of Propositional Calculus
K	$B(\varphi \to \psi) \to (B\varphi \to B\psi)$
D	$\neg B\bot \quad (B\varphi \to \neg B\neg\varphi)$
4	$B\varphi \to BB\varphi$
5	$\neg B\varphi \to B\neg B\varphi$
MP	$\dfrac{\varphi \qquad \varphi \to \psi}{\psi}$
Nec	$\dfrac{\varphi}{B\psi}$

Theorem

KD45 *is sound and strongly complete with respect to the class of Kripke frames with pseudo-equivalence relations (reflexive, transitive and serial).*

Truth Axiom/Consistency

$$K\varphi \rightarrow \varphi$$

$$\neg B \bot$$

Negative Introspection

$$\neg\Box\varphi \to \Box\neg\Box\varphi$$

$(\Box = K, B)$

Why would an agent not know some fact φ? (i.e., why would $\neg K_i\varphi$ be true?)

Why would an agent not know some fact φ? (i.e., why would $\neg K_i\varphi$ be true?)

- The agent may or may not believe φ, but has not ruled out all the $\neg\varphi$-worlds

Why would an agent not know some fact φ? (i.e., why would $\neg K_i\varphi$ be true?)

- The agent may or may not believe φ, but has not ruled out all the $\neg\varphi$-worlds
- The agent may believe φ and ruled-out the $\neg\varphi$-worlds, but this was based on “bad” evidence, or was not justified, or the agent was “epistemically lucky” (e.g., Gettier cases),...

Why would an agent not know some fact φ? (i.e., why would $\neg K_i\varphi$ be true?)

- The agent may or may not believe φ, but has not ruled out all the $\neg\varphi$-worlds
- The agent may believe φ and ruled-out the $\neg\varphi$-worlds, but this was based on "bad" evidence, or was not justified, or the agent was "epistemically lucky" (e.g., Gettier cases),...
- The agent has not yet entertained possibilities relevant to the truth of φ (the agent is unaware of φ).

Positive Introspection

$$\Box\varphi \rightarrow \Box\Box\varphi$$

$(\Box = K, B)$

How Many Modalities?

Fact. In **S5** and **KD45**, there are only three modalities ($\Box$, $\Diamond$, and the "empty modality")

Y. Ding, W. Holliday, and C. Zhang. *When Do Introspection Axioms Matter for Multi-Agent Epistemic Reasoning?*. Proceedings of TARK 2019.

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

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

It is not Common Knowledge who "defined" Common Knowledge!

The first formal definition of common knowledge?

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

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

The first formal definition of common knowledge?

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

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

The first rigorous analysis of common knowledge

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

The first formal definition of common knowledge?

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

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

The first rigorous analysis of common knowledge

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

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

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

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

The first formal definition of common knowledge?

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

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

The first rigorous analysis of common knowledge

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

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

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

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

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

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

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

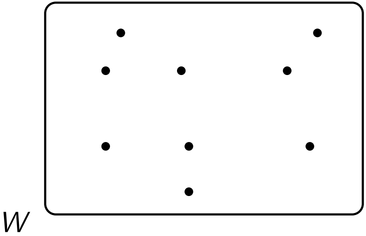
P. Vanderschraaf and G. Sillari. *"Common Knowledge", The Stanford Encyclopedia of Philosophy (2009)*.
http://plato.stanford.edu/entries/common-knowledge/.

The "Standard" Account

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

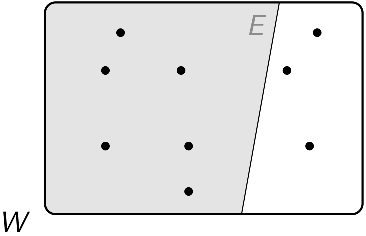
R. Fagin, J. Halpern, Y. Moses and M. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.

The “Standard” Account



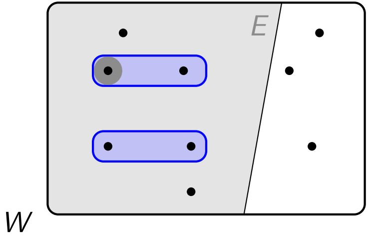
W is a set of **states** or **worlds**.

The "Standard" Account



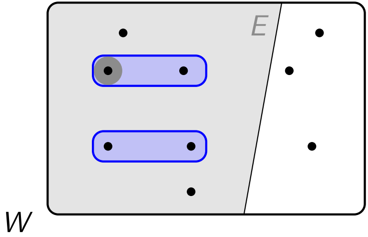
An **event/proposition** is any (definable) subset $E \subseteq W$

The "Standard" Account



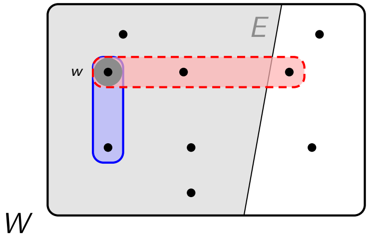
At each state, agents are assigned a set of states they *consider possible* (according to their information).
The information may be (in)correct, partitional, ...

The "Standard" Account



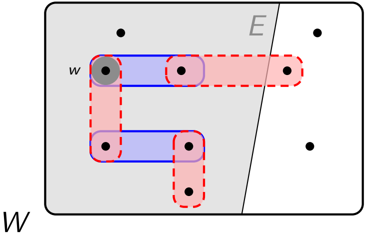
Knowledge Function: $K_i : \wp(W) \to \wp(W)$ where $K_i(E) = \{w \mid R_i(w) \subseteq E\}$

The "Standard" Account



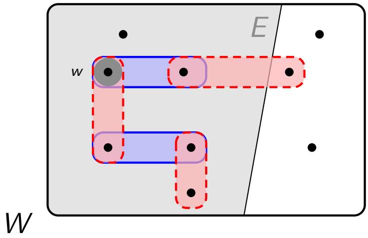
$w \in K_A(E)$ and $w \notin K_B(E)$

The “Standard” Account



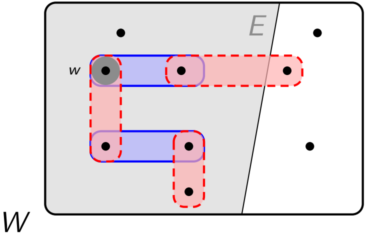
The model also describes the agents' **higher-order knowledge/beliefs**

The "Standard" Account



Everyone Knows: $K(E) = \bigcap_{i \in \mathcal{A}} K_i(E)$, $K^0(E) = E$, $K^m(E) = K(K^{m-1}(E))$

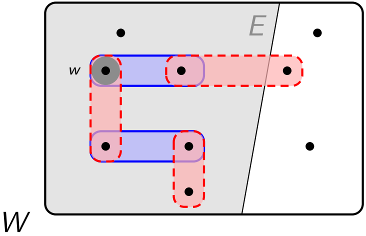
The "Standard" Account



Common Knowledge: $C : \wp(W) \rightarrow \wp(W)$ with

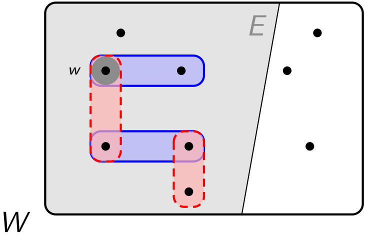
$$C(E) = \bigcap_{m \geq 0} K^m(E)$$

The "Standard" Account



$w \in K(E)$ $\qquad$ $w \notin C(E)$

The "Standard" Account



$w \in C(E)$

Fact. For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

Fact. For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

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

Fact. For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

An event F is **self-evident** if $K_i(F) = F$ for all $i \in \mathcal{A}$.

Fact. An event E is commonly known iff some self-evident event that entails E obtains.

Fact. For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

An event F is **self-evident** if $K_i(F) = F$ for all $i \in \mathcal{A}$.

Fact. An event E is commonly known iff some self-evident event that entails E obtains.

Fact. $w \in C(E)$ if every finite path starting at w ends in a state in E

The following axiomatize common knowledge:

- $C(\varphi \to \psi) \to (C\varphi \to C\psi)$
- $C\varphi \to (\varphi \wedge EC\varphi)$ (Fixed-Point)
- $C(\varphi \to E\varphi) \to (\varphi \to C\varphi)$ (Induction)

The Fixed-Point Definition

Separating the fixed-point/iteration definition of common knowledge/belief:

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

J. van Benthem and D. Saraenac. *The Geometry of Knowledge*. Aspects of Universal Logic (2004).

A. Heifetz. *Iterative and Fixed Point Common Belief*. Journal of Philosophical Logic (1999).

Some Issues

- What *does* a group know/believe/accept? vs. what *can* a group (come to) know/believe/accept?

C. List. *Group knowledge and group rationality: a judgment aggregation perspective*. Episteme (2008).

- Other "group informational attitudes": distributed knowledge, common belief, ...

- Where does common knowledge come from?

Distributed Knowledge

$$\mathcal{M}, w \models D_G\varphi \text{ iff for all } v \text{ if } w \bigcap_{i \in G} R_i v \text{ then } \mathcal{M}, v \models \varphi\}$$

Distributed Knowledge

$$\mathcal{M}, w \models D_G\varphi \text{ iff for all } v \text{ if } w \bigcap_{i\in G} R_i v \text{ then } \mathcal{M}, v \models \varphi\}$$

- $K_i(p) \wedge K_j(p \to q) \to D_{\{i,j\}}(q)$
- $D_G(\varphi) \to \bigwedge_{i\in G} K_i\varphi$

Distributed Knowledge

$$\mathcal{M}, w \models D_G\varphi \text{ iff for all } v \text{ if } w \bigcap_{i\in G} R_i v \text{ then } \mathcal{M}, v \models \varphi\}$$

- $K_i(p) \wedge K_j(p \to q) \to D_{\{i,j\}}(q)$
- $D_G(\varphi) \to \bigwedge_{i\in G} K_i\varphi$

F. Roelofsen. *Distributed Knowledge*. Journal of Applied Nonclassical Logic (2006).

- Logics of knowledge and belief: $KX \supset BX$, $BX \supset BKX$, $BX \supset KBX$, ...

- Logical omniscience: from $X \supset Y$, infer $KX \supset KY$; $K(X \supset Y) \supset (KX \supset KY)$, $(KX \wedge KY) \equiv K(X \wedge Y)$; from $X \equiv Y$ infer $KX \equiv KY$, ...

- Awareness logics, justification logic

- Dynamic epistemic logic: $[B]KX$, $\neg[X \wedge \neg KX]KX$, $[X]CX$

- Logics of belief: Plausibility structures, probabilistic beliefs