Logics of Rational Agency

Lecture 3

Eric Pacuit

Tilburg Institute for Logic and Philosophy of Science
Tilburg Univeristy
ai.stanford.edu/~epacuit

June 24, 2010

Ingredients of a Logical Analysis of Rational Agency

- ✓ informational attitudes (eg., knowledge, belief, certainty)
- ✓ group notions (eg., common knowledge and coalitional ability)
- ✓ motivational attitudes (eg., preferences)
- ✓ time, actions and ability
- ✓ normative attitudes (eg., obligations)

All Things Being Equal...

Let Γ be a set of (preference) formulas. Write $w \equiv_\Gamma v$ if for all $\varphi \in \Gamma$, $w \models \varphi$ iff $v \models \varphi$.

1. $\mathcal{M}, w \models \langle \Gamma \rangle \varphi$ iff there is a $v \in W$ such that $w \equiv_\Gamma v$ and $\mathcal{M}, v \models \varphi$.
2. $\mathcal{M}, w \models \langle \Gamma \rangle^{\leq} \varphi$ iff there is a $v \in W$ such that $w(\equiv_\Gamma \cap \leq)v$ and $\mathcal{M}, v \models \varphi$.
3. $\mathcal{M}, w \models \langle \Gamma \rangle^{<} \varphi$ iff there is a $v \in W$ such that $w(\equiv_\Gamma \cap <)v$ and $\mathcal{M}, v \models \varphi$.

All Things Being Equal...

Let Γ be a set of (preference) formulas. Write $w \equiv_\Gamma v$ if for all $\varphi \in \Gamma$, $w \models \varphi$ iff $v \models \varphi$.

1. $\mathcal{M}, w \models \langle \Gamma \rangle \varphi$ iff there is a $v \in W$ such that $w \equiv_\Gamma v$ and $\mathcal{M}, v \models \varphi$.
2. $\mathcal{M}, w \models \langle \Gamma \rangle^{\leq} \varphi$ iff there is a $v \in W$ such that $w(\equiv_\Gamma \cap \leq)v$ and $\mathcal{M}, v \models \varphi$.
3. $\mathcal{M}, w \models \langle \Gamma \rangle^{<} \varphi$ iff there is a $v \in W$ such that $w(\equiv_\Gamma \cap <)v$ and $\mathcal{M}, v \models \varphi$.

Key Principles:

- $\langle \Gamma' \rangle \varphi \rightarrow \langle \Gamma \rangle \varphi$ if $\Gamma \subseteq \Gamma'$
- $\pm\varphi \wedge \langle \Gamma \rangle (\alpha \wedge \pm\varphi) \rightarrow \langle \Gamma \cup \{\varphi\} \rangle \alpha$

All Things Being Equal...

Let Γ be a set of (preference) formulas. Write $w \equiv_\Gamma v$ if for all $\varphi \in \Gamma$, $w \models \varphi$ iff $v \models \varphi$.

1. $\mathcal{M}, w \models \langle\Gamma\rangle\varphi$ iff there is a $v \in W$ such that $w \equiv_\Gamma v$ and $\mathcal{M}, v \models \varphi$.
2. $\mathcal{M}, w \models \langle\Gamma\rangle^{\leq}\varphi$ iff there is a $v \in W$ such that $w(\equiv_\Gamma \cap \leq)v$ and $\mathcal{M}, v \models \varphi$.
3. $\mathcal{M}, w \models \langle\Gamma\rangle^{<}\varphi$ iff there is a $v \in W$ such that $w(\equiv_\Gamma \cap <)v$ and $\mathcal{M}, v \models \varphi$.

Key Principles:

- $\langle\Gamma'\rangle^{\leq}\varphi \rightarrow \langle\Gamma\rangle^{\leq}\varphi$ if $\Gamma \subseteq \Gamma'$
- $\pm\varphi \wedge \langle\Gamma\rangle^{\leq}(\alpha \wedge \pm\varphi) \rightarrow \langle\Gamma \cup \{\varphi\}\rangle^{\leq}\alpha$

From Worlds to Sets and Back

Lifting

From Worlds to Sets and Back

Lifting

- $X \geq_{\forall\exists} Y$ if $\forall y \in Y\ \exists x \in X$: $x \succeq y$

From Worlds to Sets and Back

Lifting

- $X \geq_{\forall\exists} Y$ if $\forall y \in Y$ $\exists x \in X$: $x \succeq y$
  $A(\varphi \rightarrow \langle\succeq\rangle\psi)$

From Worlds to Sets and Back

Lifting

- $X \geq_{\forall\exists} Y$ if $\forall y \in Y\ \exists x \in X$: $x \succeq y$
  $A(\varphi \rightarrow \langle\succeq\rangle\psi)$
- $X \geq_{\forall\forall} Y$ if $\forall y \in Y\ \forall x \in X$: $x \succeq y$
  $A(\varphi \rightarrow [\succ]\neg\psi)$

From Worlds to Sets and Back

Lifting

- $X \geq_{\forall\exists} Y$ if $\forall y \in Y\ \exists x \in X$: $x \succeq y$
  $A(\varphi \to \langle\succeq\rangle\psi)$
- $X \geq_{\forall\forall} Y$ if $\forall y \in Y\ \forall x \in X$: $x \succeq y$
  $A(\varphi \to [\succ]\neg\psi)$

Deriving

$P_1 \gg P_2 \gg P_3 \gg \cdots \gg P_n$
$x > y$ iff x and y differ in at least one P_i and the first P_i where this happens is one with $P_i x$ and $\neg P_i y$

F. Liu and D. De Jongh. *Optimality, belief and preference*. 2006.

The Logic of Group Decisions

The Logic of Group Decisions

Fundamental Problem: groups are inconsistent!

The Logic of Group Decisions: The Doctrinal "Paradox" (Kornhauser and Sager 1993)

p: a valid contract was in place

q: there was a breach of contract

r: the court is required to find the defendant liable.

	p	q	$(p \wedge q) \leftrightarrow r$	r
1	yes	yes	yes	yes
2	yes	no	yes	no
3	no	yes	yes	no

The Logic of Group Decisions: The Doctrinal "Paradox" (Kornhauser and Sager 1993)

Should we accept r?

	p	q	$(p \wedge q) \leftrightarrow r$	r
1	yes	yes	yes	yes
2	yes	no	yes	no
3	no	yes	yes	no

The Logic of Group Decisions: The Doctrinal "Paradox" (Kornhauser and Sager 1993)

Should we accept r? No, a simple majority votes no.

	p	q	$(p \wedge q) \leftrightarrow r$	r
1	yes	yes	yes	yes
2	yes	no	yes	no
3	no	yes	yes	no

The Logic of Group Decisions: The Doctrinal "Paradox" (Kornhauser and Sager 1993)

Should we accept r? Yes, a majority votes yes for p and q and $(p \wedge q) \leftrightarrow r$ is a legal doctrine.

	p	q	$(p \wedge q) \leftrightarrow r$	r
1	yes	yes	yes	yes
2	yes	no	yes	no
3	no	yes	yes	no

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"
$a \rightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"

$a \rightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"

b "There will be global warming"

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"
$a \longrightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"
b "There will be global warming"

	a	$a \longrightarrow b$	b

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"

$a \longrightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"

b "There will be global warming"

	a	$a \longrightarrow b$	b
1	True	True	

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"
$a \longrightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"
b "There will be global warming"

	a	$a \longrightarrow b$	b
1	True	True	True

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"
$a \longrightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"
b "There will be global warming"

	a	$a \longrightarrow b$	b
1	True	True	True
2	True	False	

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"

$a \longrightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"

b "There will be global warming"

	a	$a \longrightarrow b$	b
1	True	True	True
2	True	False	False

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"
$a \longrightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"
b "There will be global warming"

	a	$a \longrightarrow b$	b
1	True	True	True
2	True	False	False
3	False	True	

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"
$a \longrightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"
b "There will be global warming"

	a	$a \longrightarrow b$	b
1	True	True	True
2	True	False	False
3	False	True	False
Majority			

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"

$a \rightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"

b "There will be global warming"

	a	$a \rightarrow b$	b
1	True	True	True
2	True	False	False
3	False	True	False
Majority	True		

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"
$a \rightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"
b "There will be global warming"

	a	$a \rightarrow b$	b
1	True	True	True
2	True	False	False
3	False	True	False
Majority	True	True	

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"
$a \longrightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"
b "There will be global warming"

	a	$a \longrightarrow b$	b
1	True	True	True
2	True	False	False
3	False	True	False
Majority	True	True	False

Discursive Dilemma

a: "Carbon dioxide emissions are above the threshold x"

$a \rightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming"

b "There will be global warming"

	a	$a \rightarrow b$	b
1	True	True	True
2	True	False	False
3	False	True	False
Majority	True	True	False

Conclusion: Groups are inconsistent, difference between 'premise-based' and 'conclusion-based' decision making, ...

The Judgement Aggregation Model: The Propositions

Propositions: Let $\mathcal{L}$ be a logical language (called **propositions** in the literature) with the usual boolean connectives.

The Judgement Aggregation Model: The Propositions

Propositions: Let $\mathcal{L}$ be a logical language (called **propositions** in the literature) with the usual boolean connectives.

Consistency: The standard notion of logical consistency.

The Judgement Aggregation Model: The Propositions

Propositions: Let $\mathcal{L}$ be a logical language (called **propositions** in the literature) with the usual boolean connectives.

Consistency: The standard notion of logical consistency.

Aside: We actually need

1. $\{p, \neg p\}$ are inconsistent
2. all subsets of a consistent set are consistent
3. $\emptyset$ is consistent and each $S \subseteq \mathcal{L}$ has a consistent maximal extension (not needed in all cases)

The Judgement Aggregation Model: The Propositions

Definition A set $Y \subseteq \mathcal{L}$ is **minimally inconsistent** if it is inconsistent and every proper subset $X \subsetneq Y$ is consistent.

The Judgement Aggregation Model: The Agenda

Definition The **agenda** is a non-empty set $X \subseteq \mathcal{L}$, interpreted as the set of propositions on which judgments are made, with X is a union of proposition-negation pairs $\{p, \neg p\}$.

The Judgement Aggregation Model: The Agenda

Definition The **agenda** is a non-empty set $X \subseteq \mathcal{L}$, interpreted as the set of propositions on which judgments are made, with X is a union of proposition-negation pairs $\{p, \neg p\}$.

Example: In the discursive dilemma:
$X = \{a, \neg a, b, \neg b, a \rightarrow b, \neg(a \rightarrow b)\}$.

The Judgement Aggregation Model: The Judgement Sets

Definition: Given an agenda X, each individual i's judgement set is a subset $A_i \subseteq X$.

The Judgement Aggregation Model: The Judgement Sets

Definition: Given an agenda X, each individual i's judgement set is a subset $A_i \subseteq X$.

Rationality Assumptions:

1. A_i is **consistent**
2. A_i is **complete**, if for each $p \in X$, either $p \in A_i$ or $\neg p \in A_i$

The Judgement Aggregation Model: Aggregation Rules

Let X be an agenda, $N = \{1, \ldots, n\}$ a set of voters, a **profile** is a tuple $(A_i, \ldots, A_n)$ where each A_i is a judgement set. An **aggregation function** is a map from profiles to judgment sets. I.e., $F(A_1, \ldots, A_n)$ is a judgement set.

The Judgement Aggregation Model: Aggregation Rules

Let X be an agenda, $N = \{1, \ldots, n\}$ a set of voters, a **profile** is a tuple $(A_i, \ldots, A_n)$ where each A_i is a judgement set. An **aggregation function** is a map from profiles to judgment sets. I.e., $F(A_1, \ldots, A_n)$ is a judgement set.

Examples:

- **Propositionwise majority voting**: for each $(A_1, \ldots, A_n)$,

$$F(A_1, \ldots, A_n) = \{p \in X \mid |\{i \mid p \in A_i\}| \geq |\{i \mid p \notin A_i\}|\}$$

- **Dictator of** i: $F(A_1, \ldots, A_n) = A_i$
- **Reverse Dictator of** i: $F(A_1, \ldots, A_n) = \{\neg p \mid p \in A_i\}$

The Judgement Aggregation Model: Aggregation Rules

Universal Domain: The domain of F is the set of all possible profiles of consistent and complete judgement sets.

Collective Rationality: F generates consistent and complete collective judgment sets.

The Judgement Aggregation Model: Aggregation Rules

Unanimity: For all profiles $(A_1, \ldots, A_n)$ if $p \in A_i$ for each i then $p \in F(A_1, \ldots, A_n)$

The Judgement Aggregation Model: Aggregation Rules

Unanimity: For all profiles $(A_1, \ldots, A_n)$ if $p \in A_i$ for each i then $p \in F(A_1, \ldots, A_n)$

Independence: For any $p \in X$ and all $(A_1, \ldots, A_n)$ and $(A_1^*, \ldots, A_n^*)$ in the domain of F,

$$\text{if [for all } i \in N,\ p \in A_i \text{ iff } p \in A_i^*] \\ \text{then } [p \in F(A_1, \ldots, A_n) \text{ iff } p \in F(A_1^*, \ldots A_n^*)\ ].$$

The Judgement Aggregation Model: Aggregation Rules

Systematicity: For any $p, q \in X$ and all $(A_1, \ldots, A_n)$ and $(A_1^*, \ldots, A_n^*)$ in the domain of F,

$$\text{if [for all } i \in N\text{, } p \in A_i \text{ iff } q \in A_i^*]$$
$$\text{then } [p \in F(A_1, \ldots, A_n) \text{ iff } q \in F(A_1^*, \ldots A_n^*)\ ].$$

The Judgement Aggregation Model: Aggregation Rules

Systematicity: For any $p, q \in X$ and all $(A_1, \ldots, A_n)$ and $(A_1^*, \ldots, A_n^*)$ in the domain of F,

$$\text{if [for all } i \in N,\ p \in A_i \text{ iff } q \in A_i^*] \\ \text{then } [p \in F(A_1, \ldots, A_n) \text{ iff } q \in F(A_1^*, \ldots A_n^*)\ ].$$

Monotonicity: For any $p \in X$ and all $(A_1, \ldots A_i, \ldots, A_n)$ and $(A_1, \ldots, A_i^*, \ldots, A_n)$ in the domain of F,

$$\text{if } [p \notin A_i,\ p \in A_i^* \text{ and } p \in F(A_1, \ldots, A_i, \ldots A_n)] \\ \text{then } [p \in F(A_1, \ldots, A_i^*, \ldots A_n)].$$

The Judgement Aggregation Model: Aggregation Rules

Anonymity: If $(A_1, \ldots, A_n)$ and $(A_1^*, \ldots, A_n^*)$ are permutations of each other, then

$$F(A_1, \ldots, A_n) = F(A_1^*, \ldots, A_n^*)$$

Non-dictatorship: There exists no $i \in N$ such that, for any profile $(A_1, \ldots, A_n)$, $F(A_1, \ldots, A_n) = A_i$

Baseline Result

Theorem (List and Pettit, 2001) If $X \subseteq \{a, b, a \wedge b\}$, there exists no aggregation rule satisfying universal domain, collective rationality, systematicity and anonymity.

Baseline Result

Theorem (List and Pettit, 2001) If $X \subseteq \{a, b, a \wedge b\}$, there exists no aggregation rule satisfying universal domain, collective rationality, systematicity and anonymity.

See `personal.lse.ac.uk/LIST/doctrinalparadox.htm` for many generalizations!

Agenda Richness

Whether or not judgment aggregation gives rise to serious impossibility results depends on how the propositions in the agenda are *interconnected*.

Agenda Richness

Whether or not judgment aggregation gives rise to serious impossibility results depends on how the propositions in the agenda are *interconnected*.

Definition An agenda X is **minimally connected** if

1. it has a minimal inconsistent subset $Y \subseteq X$ with $|Y| \geq 3$
2. it has a minimal inconsistent subset $Y \subseteq X$ such that

$$Y - Z \cup \{\neg z \mid z \in Z\} \text{ is consistent}$$

for some subset $Z \subseteq Y$ of even size.

Theorem (Dietrich and List, 2007) For a minimally connected agenda X, an aggregation rule F satisfies universal domain, collective rationality, systematicity and the unanimity principle iff it is a dictatorship.

Theorem (Dietrich and List, 2007) For a minimally connected agenda X, an aggregation rule F satisfies universal domain, collective rationality, systematicity and the unanimity principle iff it is a dictatorship.

Pause for proof

Many Variants!

Christian List and Clemens Puppe. *Judgement Aggregation: A Survey*. 2010.

Group Preference Logics

H. Andréka, M. Ryan and P Yves Schobbens. *Operators and laws for combining preference relations*. Journal of Logic and Computation, 2002.

P. Girard. *Modal Logic for Lexicographic Preference Aggregation*. Manuscript, 2008.

Ingredients of a Logical Analysis of Rational Agency

✓ Logics of Informational Attitudes and Informative Actions

✓ Logics of Motivational Attitudes (Preferences)

- Time, Action and Agency

Time

One of the most “successful” applications of modal logic is in the “logic of time”.

Time

One of the most "successful" applications of modal logic is in the "logic of time".

Many variations

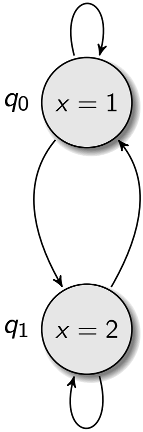
- discrete or continuous
- branching or linear
- point based or interval based

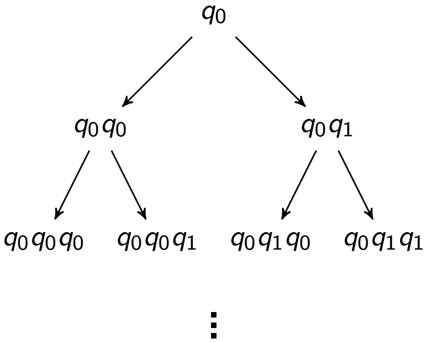
See, for example,

Antony Galton. *Temporal Logic*. Stanford Encyclopedia of Philosophy: `http://plato.stanford.edu/entries/logic-temporal/`.

I. Hodkinson and M. Reynolds. *Temporal Logic*. Handbook of Modal Logic, 2008.

Computational vs. Behavioral Structures





Temporal Logics

Temporal Logics

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

$$\Diamond\varphi\text{: } \varphi \text{ is true some time in } \textit{the} \text{ future.}$$

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

Temporal Logics

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

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

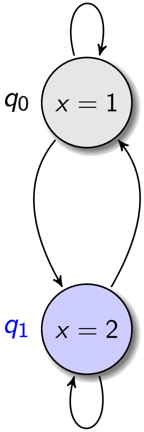
A. Pnuelli. *A Temporal Logic of Programs*. in *Proc. 18th IEEE Symposium on Foundations of Computer Science* (1977).

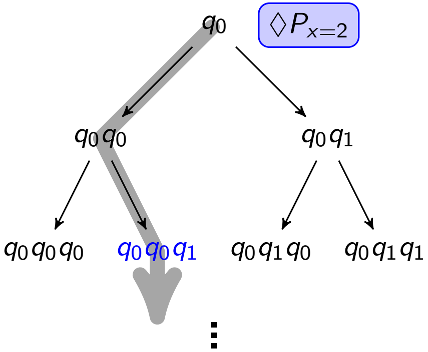
- *Branching Time Temporal Logic*: Allows quantification over paths:

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

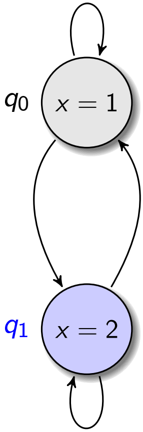
E. M. Clarke and E. A. Emerson. *Design and Synthesis of Synchronization Skeletons using Branching-time Temproal-logic Specifications*. In *Proceedings Workshop on Logic of Programs*, LNCS (1981).

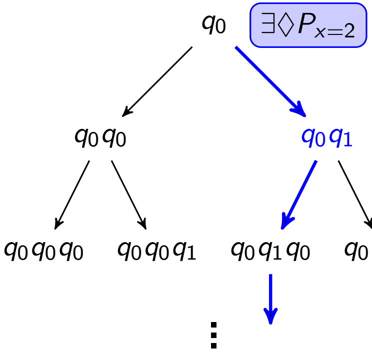
Temporal Logics



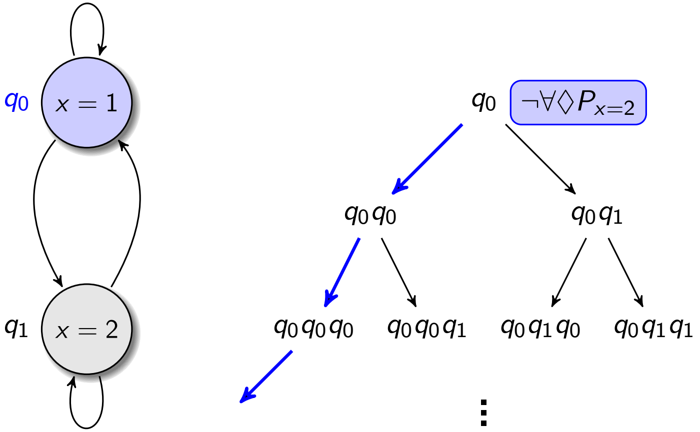


Temporal Logics





Temporal Logics



Abilities

$Abl_i\varphi$: agent i has the ability to bring about (see to it that) φ is true

What are core logical principles?

Abilities

$Abl_i\varphi$: agent i has the ability to bring about (see to it that) φ is true

What are core logical principles?

1. $Abl_i\varphi \rightarrow \varphi$ (or $\varphi \rightarrow Abl_i\varphi$)

Abilities

$Abl_i\varphi$: agent i has the ability to bring about (see to it that) φ is true

What are core logical principles?

1. $Abl_i\varphi \rightarrow \varphi$ (or $\varphi \rightarrow Abl_i\varphi$)
2. $\neg Abl_i\top$

Abilities

$Abl_i\varphi$: agent i has the ability to bring about (see to it that) φ is true

What are core logical principles?

1. $Abl_i\varphi \to \varphi$ (or $\varphi \to Abl_i\varphi$)
2. $\neg Abl_i\top$
3. $(Abl_i\varphi \wedge Abl_i\psi) \to Abl_i(\varphi \wedge \psi)$

Abilities

$Abl_i\varphi$: agent i has the ability to bring about (see to it that) φ is true

What are core logical principles?

1. $Abl_i\varphi \rightarrow \varphi$ (or $\varphi \rightarrow Abl_i\varphi$)
2. $\neg Abl_i\top$
3. $(Abl_i\varphi \wedge Abl_i\psi) \rightarrow Abl_i(\varphi \wedge \psi)$
4. $Abl_i(\varphi \vee \psi) \rightarrow (Abl_i\varphi \vee Abl_i\psi)$

Abilities

$Abl_i\varphi$: agent i has the ability to bring about (see to it that) φ is true

What are core logical principles?

1. $Abl_i\varphi \rightarrow \varphi$ (or $\varphi \rightarrow Abl_i\varphi$)
2. $\neg Abl_i\top$
3. $(Abl_i\varphi \wedge Abl_i\psi) \rightarrow Abl_i(\varphi \wedge \psi)$
4. $Abl_i(\varphi \vee \psi) \rightarrow (Abl_i\varphi \vee Abl_i\psi)$
5. $Abl_i(\varphi \wedge \psi) \rightarrow (Abl_i\varphi \wedge Abl_i\psi)$

Abilities

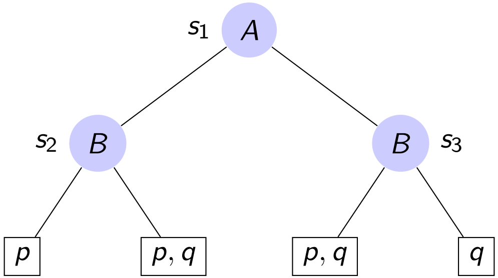
$Abl_i\varphi$: agent i has the ability to bring about (see to it that) φ is true

What are core logical principles?

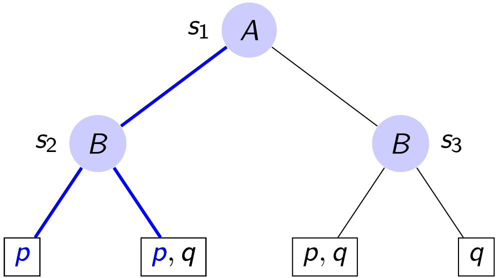
1. $Abl_i\varphi \to \varphi$ (or $\varphi \to Abl_i\varphi$)
2. $\neg Abl_i\top$
3. $(Abl_i\varphi \wedge Abl_i\psi) \to Abl_i(\varphi \wedge \psi)$
4. $Abl_i(\varphi \vee \psi) \to (Abl_i\varphi \vee Abl_i\psi)$
5. $Abl_i(\varphi \wedge \psi) \to (Abl_i\varphi \wedge Abl_i\psi)$
6. $Abl_iAbl_j\varphi \to Abl_i\varphi$

Games: $(Abl_i\varphi \wedge Abl_i\psi) \not\rightarrow Abl_i(\varphi \wedge \psi)$

Games: $(Abl_i\varphi \wedge Abl_i\psi) \not\to Abl_i(\varphi \wedge \psi)$

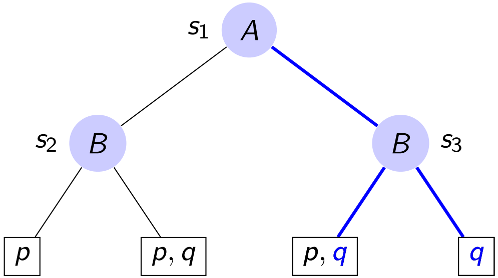


Games: $(Abl_i\varphi \wedge Abl_i\psi) \not\to Abl_i(\varphi \wedge \psi)$



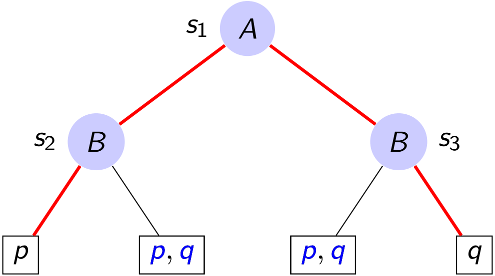
$$s_1 \models \Box_A p$$

Games: $(Abl_i\varphi \wedge Abl_i\psi) \not\to Abl_i(\varphi \wedge \psi)$



$$s_1 \models \Box_A p \wedge \Box_A q$$

Games: $(Abl_i\varphi \wedge Abl_i\psi) \not\to Abl_i(\varphi \wedge \psi)$



$$s_1 \models \Box_A p \wedge \Box_A q \wedge \neg\Box_A(p \wedge q)$$

Games: $(Abl_i\varphi \wedge Abl_i\psi) \not\to Abl_i(\varphi \wedge \psi)$

R. Parikh. *The Logic of Games and its Applications*. Annals of Discrete Mathematics (1985).

M. Pauly and R. Parikh. *Game Logic — An Overview*. Studia Logica (2003).

J. van Benthem. *Logic and Games*. Course notes (2007).

$\varphi \not\to Abl_i\varphi$

Suppose an agent (call her Ann) is throwing a dart and she is not a very good dart player, but she just happens to throw a bull's eye.

$\varphi \not\to Abl_i\varphi$

Suppose an agent (call her Ann) is throwing a dart and she is not a very good dart player, but she just happens to throw a bull's eye.

Intuitively, we do not want to say that Ann has the *ability* to throw a bull's eye even though it happens to be true.

$Abl_i(\varphi \vee \psi) \not\to Abl_i\varphi \vee Abl_i\psi$

Continuing with this example, suppose that Ann has the ability to hit the dart board, but has no other control over the placement of the dart.

$Abl_i(\varphi \vee \psi) \not\to Abl_i\varphi \vee Abl_i\psi$

Continuing with this example, suppose that Ann has the ability to hit the dart board, but has no other control over the placement of the dart.

Now, when she throws the dart, as a matter of fact, it will either hit the top half of the board or the bottom half of the board.

$Abl_i(\varphi \vee \psi) \not\to Abl_i\varphi \vee Abl_i\psi$

Continuing with this example, suppose that Ann has the ability to hit the dart board, but has no other control over the placement of the dart.

Now, when she throws the dart, as a matter of fact, it will either hit the top half of the board or the bottom half of the board.

Since, Ann has the ability to hit the dart board, she has the ability to either hit the top half of the board or the bottom half of the board.

$Abl_i(\varphi \vee \psi) \not\rightarrow Abl_i\varphi \vee Abl_i\psi$

Continuing with this example, suppose that Ann has the ability to hit the dart board, but has no other control over the placement of the dart.

Now, when she throws the dart, as a matter of fact, it will either hit the top half of the board or the bottom half of the board.

Since, Ann has the ability to hit the dart board, she has the ability to either hit the top half of the board or the bottom half of the board.

However, intuitively, it seems true that Ann does not have the ability to hit the top half of the dart board, and also she does not have the ability to hit the bottom half of the dart board.

Abilities

$Abl_i\varphi$: agent i has the ability to bring about (see to it that) φ is true

What are core logical principles? Depends very much on the intended "application" and how actions are represented...

1. $Abl_i\varphi \rightarrow \varphi$ (or $\varphi \rightarrow Abl_i\varphi$)
2. $\neg Abl_i\top$
3. $(Abl_i\varphi \wedge Abl_i\psi) \rightarrow Abl_i(\varphi \wedge \psi)$
4. $Abl_i(\varphi \vee \psi) \rightarrow (Abl_i\varphi \vee Abl_i\psi)$
5. $Abl_i(\varphi \wedge \psi) \rightarrow (Abl_i\varphi \wedge Abl_i\psi)$
6. $Abl_iAbl_j\varphi \rightarrow Abl_i\varphi$

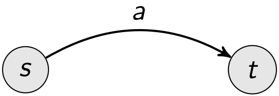
G. Governatori and A. Rotolo. *On the Axiomatisation of Elgesem's Logic of Agency and Ability*. Journal of Philosophical Logic, 34, pgs. 403 - 431 (2005).

Actions

1. Actions as *transitions between states, or situations*:

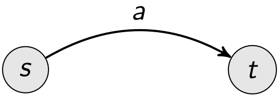
Actions

1. Actions as *transitions between states, or situations*:



Actions

1. Actions as *transitions between states, or situations*:



2. Actions *restrict* the set of possible future histories.

Propositional Dynamic Logic

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 $V : \mathsf{At} \rightarrow \wp(W)$

- $R_{\alpha\cup\beta} := R_\alpha \cup R_\beta$
- $R_{\alpha;\beta} := R_\alpha \circ R_\beta$
- $R_{\alpha^*} := \cup_{n\geq 0} R_\alpha^n$
- $R_{\varphi?} = \{(w, w) \mid \mathcal{M}, w \models \varphi\}$

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

Background: Propositional Dynamic Logic

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

Background: Propositional Dynamic Logic

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

Propositional Dynamic Logic

Theorem PDL is sound and weakly complete with respect to the Segerberg Axioms.

Theorem The satisfiability problem for **PDL** is decidable ($\mathrm{EXPTIME}$-Complete).

D. Kozen and R. Parikh. *A Completeness proof for Propositional Dynamic Logic*.
.

D. Harel, D. Kozen and Tiuryn. *Dynamic Logic*. 2001.

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*. JPL, 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 α, and
2. α 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 α, and
2. α 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 α, and
2. α only terminates at states in which it is true that A

Interestingly, Segerberg also briefly considers a third condition:

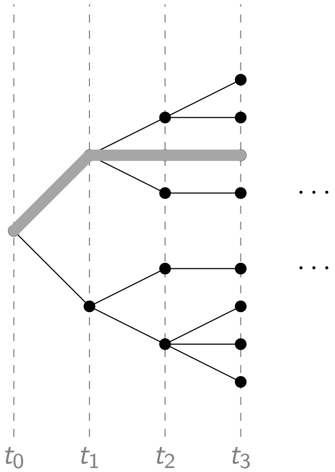
3. p is optimal (in some sense: shortest, maximally efficient, most convenient, etc.) in the set of computations satisfying conditions (1) and (2).

The axioms:

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

Actions and Agency in Branching Time

Alternative accounts of agency do not include explicit description of the actions:



STIT

- Each node represents a choice point for the agent.
- A **history** is a maximal branch in the above tree.
- Formulas are interpreted at history moment pairs.
- At each moment there is a choice available to the agent (partition of the histories through that moment)
- The key modality is $[stit]\varphi$ which is intended to mean that the agent i can "see to it that φ is true".
    - $[stit]\varphi$ is true at a history moment pair provided the agent can choose a (set of) branch(es) such that every future history-moment pair satisfies φ

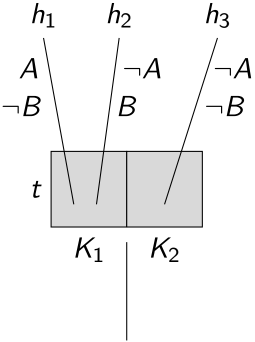
STIT

We use the modality '$\Diamond$' to mean historic possibility.

$\Diamond[\mathit{stit}]\varphi$: "the agent has the ability to bring about φ".

STIT

The following model will falsifies: $\varphi \rightarrow \Diamond[stit]\varphi$ and $\Diamond[stit](\varphi \vee \psi) \rightarrow \Diamond[stit]\varphi \vee \Diamond[stit]\psi$.



J. Horty. *Agency and Deontic Logic*. 2001.

to be continued...