Logic and Artificial Intelligence

Lecture 21

Eric Pacuit

Currently Visiting the Center for Formal Epistemology, CMU

Center for Logic and Philosophy of Science
Tilburg University

ai.stanford.edu/∼epacuit
e.j.pacuit@uvt.nl

November 16, 2011

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future
- $[\alpha]\varphi$: after doing action α, φ is true

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future
- $[\alpha]\varphi$: after doing action α, φ is true
- $[\delta\varphi]\psi$: after bringing about φ, ψ is true

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future
- $[\alpha]\varphi$: after doing action α, φ is true
- $[\delta\varphi]\psi$: after bringing about φ, ψ is true
- $[i\ stit]\varphi$: the agent can "see to it that" φ is true

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future
- $[\alpha]\varphi$: after doing action α, φ is true
- $[\delta\varphi]\psi$: after bringing about φ, ψ is true
- $[i\ stit]\varphi$: the agent can "see to it that" φ is true
- $\Diamond[i\ stit]\varphi$: the agent has the ability to bring about φ

Group/Collective actions and abilities: for $J \subseteq N$, $[J]\varphi$ means "the group can make φ true..."

Group Actions: Example

Group Actions: 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.

Group Actions: 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		

Group Actions: 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		$q_0 \Rightarrow q_1,\ q_1 \Rightarrow q_1$

Many Agents

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

Strategy Logics

- *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).

Strategy Logics

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

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

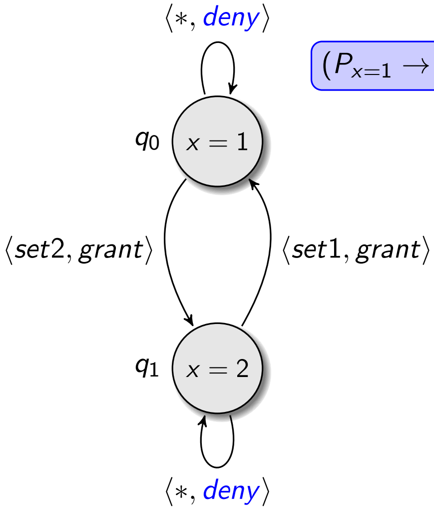
M. Pauly. *A Modal Logic for Coalition Powers in Games*. *Journal of Logic and Computation* **12** (2002).

- *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.

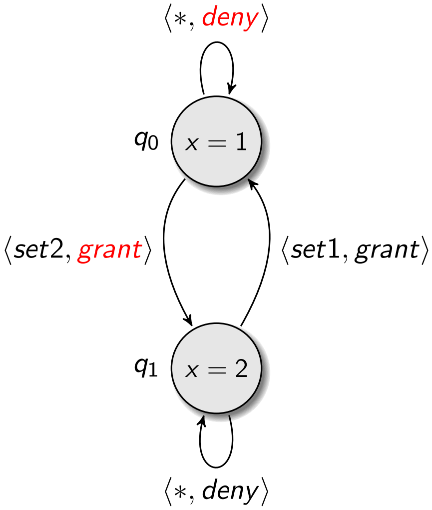
R. Alur, T. Henzinger and O. Kupferman. *Alternating-time Temproal Logic*. *Jouranl of the ACM* (2002).

Multi-agent Transition Systems



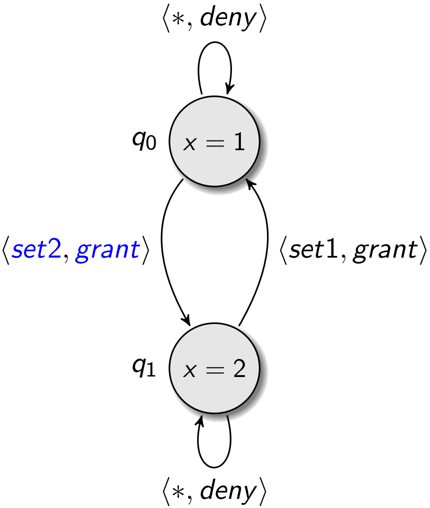
$(P_{x=1} \to [s]P_{x=1}) \wedge (P_{x=2} \to [s]P_{x=2})$

Multi-agent Transition Systems



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

Multi-agent Transition Systems



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

Logics of Abilities

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

A. Herzig and F. Schwarzentruber. *Properties of logics of individual and group agency*. C. Areces and R. Golblatt (eds.) Advances in Modal Logic, pgs. 133 - 149, 2008.

A. Herzig and E. Lorini. *A Dynamic Logic of Agency I: STIT, Capabilities and Powers*. Journal of Logic, Language and Information, 19, pgs. 89 - 121, 2010.

Logics of Abilities

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

A. Herzig and F. Schwarzentruber. *Properties of logics of individual and group agency*. C. Areces and R. Golblatt (eds.) Advances in Modal Logic, pgs. 133 - 149, 2008.

A. Herzig and E. Lorini. *A Dynamic Logic of Agency I: STIT, Capabilities and Powers*. Journal of Logic, Language and Information, 19, pgs. 89 - 121, 2010.

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

Coalitional Logic

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

Effectivity Functions

Let N be a (finite) set of agents and W a set of worlds.

Effectivity Functions

Let N be a (finite) set of agents and W a set of worlds.

For each $J \subseteq N$, the **effectivity function** for J is
$e : \wp(N) \rightarrow \wp(\wp(W))$

$X \in e(J)$ means X is a set of possible outcomes for which J is effective (or J can force the world to be in some state of X at the next step).

Playable Effectivity Functions

e is **playable** (for a set of states W) iff

1. $\emptyset \notin e(J)$ (Liveness)
2. $W \in e(J)$ (Safety)
3. if $W - X \notin e(\emptyset)$ then $X \in e(N)$ (N-maximality)
4. if $X \in e(J)$ and $X \subseteq Y$ then $Y \in e(J)$ (Outcome monotonicity)
5. If $J \cap I = \emptyset$, then if $X_1 \in e(J)$ and $X_2 \in e(I)$, then $X_1 \cap X_2 \in e(J \cup I)$ (Super-additivity)

Playable Effectivity Functions

e is the effectivity function of some strategic game provided $X \in e(J)$ if there is a joint strategy for J such that no matter what strategy is chosen by the agents $N - J$, the outcome of the game is in X.

Theorem (Pauly). An effectivity function e is playable iff it is the effectivity function of some strategic game.

Playable Effectivity Functions

e is the effectivity function of some strategic game provided $X \in e(J)$ if there is a joint strategy for J such that no matter what strategy is chosen by the agents $N - J$, the outcome of the game is in X.

Theorem (Pauly). An effectivity function e is playable iff it is the effectivity function of some strategic game.

See, also

V. Goranko, W. Jamroga, and P. Turrini. *Strategic Games and Truly Playable Effectivity Functions*. to appear in: Journal of Autonomous Agents and Multiagent Systems.

CL Models

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

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

Coalitional Logic: Axiomatics

- $\neg[J]\bot$
- $[J]\top$
- $(\neg[\emptyset]\neg\varphi) \to [N]\varphi$
- $[J](\varphi \wedge \psi) \to ([J]\varphi \wedge [J]\psi)$
- $([J_1]\varphi_1 \wedge [J_2]\varphi_2) \to [J_1 \cup J_2](\varphi_1 \wedge \varphi_2)$, where $J_1 \cap J_2 = \emptyset$

Logics of preference...

Preference (Modal) Logics

x, y objects

$x \succeq y$: x is at least as good as y

Preference (Modal) Logics

x, y objects

$x \succeq y$: x is at least as good as y

1. $x \succeq y$ and $y \not\succeq x$ ($x \succ y$)
2. $x \not\succeq y$ and $y \succeq x$ ($y \succ x$)
3. $x \succeq y$ and $y \succeq x$ ($x \sim y$)
4. $x \not\succeq y$ and $y \not\succeq x$ ($x \perp y$)

Preference (Modal) Logics

x, y objects

$x \succeq y$: x is at least as good as y

1. $x \succeq y$ and $y \not\succeq x$ ($x \succ y$)
2. $x \not\succeq y$ and $y \succeq x$ ($y \succ x$)
3. $x \succeq y$ and $y \succeq x$ ($x \sim y$)
4. $x \not\succeq y$ and $y \not\succeq x$ ($x \perp y$)

Properties: transitivity, connectedness, etc.

Preference (Modal) Logics

Modal betterness model $\mathcal{M} = \langle W, \succeq, V \rangle$

Preference (Modal) Logics

Modal betterness model $\mathcal{M} = \langle W, \succeq, V \rangle$

Preference Modalities $\langle \succeq \rangle \varphi$: "there is a world at least as good (as the current world) satisfying φ"

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

Preference (Modal) Logics

Modal betterness model $\mathcal{M} = \langle W, \succeq, V \rangle$

Preference Modalities $\langle \succeq \rangle \varphi$: "there is a world at least as good (as the current world) satisfying φ"

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

$\mathcal{M}, w \models \langle \succ \rangle \varphi$ iff there is $v \succeq w$ and $w \not\succeq v$ such that $\mathcal{M}, v \models \varphi$

Preference (Modal) Logics

1. $\langle \succ \rangle \varphi \to \langle \succeq \rangle \varphi$
2. $\langle \succeq \rangle \langle \succ \rangle \varphi \to \langle \succ \rangle \varphi$
3. $\varphi \wedge \langle \succeq \rangle \psi \to (\langle \succ \rangle \psi \vee \langle \succeq \rangle (\psi \wedge \langle \succeq \rangle \varphi))$
4. $\langle \succ \rangle \langle \succeq \rangle \varphi \to \langle \succ \rangle \varphi$

Theorem The above logic (with Necessitation and Modus Ponens) is sound and complete with respect to the class of preference models.

J. van Benthem, O. Roy and P. Girard. *Everything else being equal: A modal logic approach to* ceteris paribus *preferences*. JPL, 2008.

Preference Modalities

$\varphi \geq \psi$: the state of affairs φ is at least as good as ψ (ceteris paribus)

G. von Wright. *The logic of preference*. Edinburgh University Press (1963).

Preference Modalities

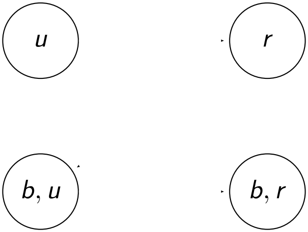
$\varphi \geq \psi$: the state of affairs φ is at least as good as ψ (ceteris paribus)

G. von Wright. *The logic of preference*. Edinburgh University Press (1963).

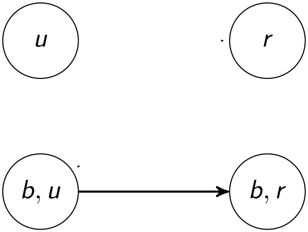
$\langle \Gamma \rangle^{\leq} \varphi$: φ is true in "better" world, *all things being equal*.

J. van Benthem, O. Roy and P. Girard. *Everything else being equal: A modal logic approach to* ceteris paribus *preferences*. JPL, 2008.

All Things Being Equal...

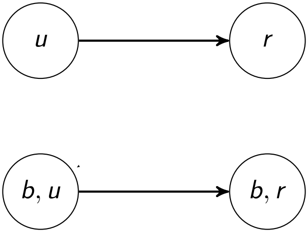


All Things Being Equal...



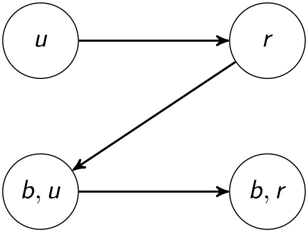
- With boots (b), I prefer my raincoat (r) over my umbrella (u)

All Things Being Equal...



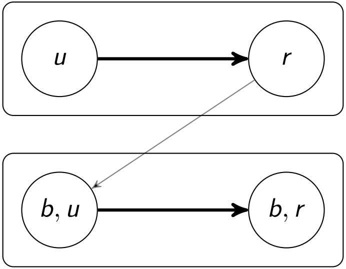
- With boots (b), I prefer my raincoat (r) over my umbrella (u)
- Without boots ($\neg b$), I also prefer my raincoat (r) over my umbrella (u)

All Things Being Equal...



- With boots (b), I prefer my raincoat (r) over my umbrella (u)
- Without boots ($\neg b$), I also prefer my raincoat (r) over my umbrella (u)
- But I do prefer an umbrella and boots over a raincoat and no boots

All Things Being Equal...



All things being equal, I prefer my raincoat over my umbrella

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:

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

Preference Lifting, I

Given a preference ordering $\preceq$ over a set of objects X, we want to **lift** this to an ordering $\hat{\preceq}$ over $\wp(X)$.

Given $\preceq$, what reasonable properties can we infer about $\hat{\preceq}$?

S. Barberá, W. Bossert, and P.K. Pattanaik. *Ranking sets of objects*. In Handbook of Utility Theory, volume 2. Kluwer Academic Publishers, 2004.

Preference Lifting, II

- You know that $x \prec y \prec z$
  Can you infer that $\{x, y\} \mathrel{\hat{\prec}} \{z\}$?

Preference Lifting, II

- You know that $x \prec y \prec z$
  Can you infer that $\{x, y\} \mathrel{\hat{\prec}} \{z\}$?

- You know that $x \prec y \prec z$
  Can you infer anything about $\{y\}$ and $\{x, z\}$?

Preference Lifting, II

- You know that $x \prec y \prec z$
  Can you infer that $\{x, y\} \hat{\prec} \{z\}$?

- You know that $x \prec y \prec z$
  Can you infer anything about $\{y\}$ and $\{x, z\}$?

- You know that $w \prec x \prec y \prec z$
  Can you infer that $\{w, x, y\} \hat{\preceq} \{w, y, z\}$?

Preference Lifting, II

- You know that $x \prec y \prec z$
  Can you infer that $\{x, y\} \mathrel{\hat{\prec}} \{z\}$?

- You know that $x \prec y \prec z$
  Can you infer anything about $\{y\}$ and $\{x, z\}$?

- You know that $w \prec x \prec y \prec z$
  Can you infer that $\{w, x, y\} \mathrel{\hat{\preceq}} \{w, y, z\}$?

- You know that $w \prec x \prec y \prec z$
  Can you infer that $\{w, x\} \mathrel{\hat{\prec}} \{y, z\}$?

Preference Lifting, III

There are different interpretations of $X \mathrel{\hat{\preceq}} Y$:

- You will get one of the elements, but cannot control which.
- You can choose one of the elements.
- You will get the full set.