# Logic and Artificial Intelligence

## Lecture 22

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 23, 2011

# 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 $\varphi$"

$\mathcal{M}, w \models \langle \succeq \rangle \varphi$ 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 $\varphi$"

$\mathcal{M}, w \models \langle \succeq \rangle \varphi$ 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 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 $\{x, y\} \hat{\prec} \{z\}$?

- 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 $\{w, x, y\} \hat{\preceq} \{w, y, z\}$?

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

# Preference Lifting, III

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

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

# Preference Lifting, IV

**Kelly Principle**

(EXT) $\{x\} \hat{\precsim} \{y\}$ provided $x \prec y$

(MAX) $A \hat{\precsim} Max(A)$

(MIN) $Min(A) \hat{\precsim} A$

J.S. Kelly. *Strategy-Proofness and Social Choice Functions without Single-Valuedness*. Econometrica, 45(2), pp. 439 - 446, 1977.

# Preference Lifting, IV

**Gärdenfors Principle**

(G1) $A \hat{\precsim} A \cup \{x\}$ if $a \prec x$ for all $a \in A$

(G2) $A \cup \{x\} \hat{\precsim} A$ if $x \prec a$ for all $a \in A$

P. Gärdenfors. *Manipulation of Social Choice Functions*. Journal of Economic Theory. 13:2, 217 - 228, 1976.

# Preference Lifting, IV

**Gärdenfors Principle**

(G1) $A \hat{\precsim} A \cup \{x\}$ if $a \prec x$ for all $a \in A$

(G2) $A \cup \{x\} \hat{\precsim} A$ if $x \prec a$ for all $a \in A$

P. Gärdenfors. *Manipulation of Social Choice Functions*. Journal of Economic Theory. 13:2, 217 - 228, 1976.

**Independence**

(IND) $A \cup \{x\} \hat{\preceq} B \cup \{x\}$ if $A \hat{\preceq} B$ and $x \notin A \cup B$

# Preference Lifting, V

**Theorem** (Kannai and Peleg). If $|X| \geq 6$, then no weak order satisfies both the Gärdenfors principle and independence.

Y. Kannai and B. Peleg. *A Note on the Extension of an Order on a Set to the Power Set*. Journal of Economic Theory, 32(1), pp. 172 - 175, 1984.

# From Worlds to Sets, I

$\mathcal{M}, w \models \varphi \preceq_{\exists\exists} \psi$ iff there is $s, t$ such that $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, t \models \psi$ and $s \preceq t$

# From Worlds to Sets, I

$\mathcal{M}, w \models \varphi \preceq_{\exists\exists} \psi$ iff there is $s, t$ such that $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, t \models \psi$ and $s \preceq t$

$\mathcal{M}, w \models \varphi \prec_{\exists\exists} \psi$ iff there is $s, t$ such that $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, t \models \psi$ and $s \prec t$

# From Worlds to Sets, I

$\mathcal{M}, w \models \varphi \preceq_{\exists\exists} \psi$ iff there is $s, t$ such that $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, t \models \psi$ and $s \preceq t$

$\mathcal{M}, w \models \varphi \prec_{\exists\exists} \psi$ iff there is $s, t$ such that $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, t \models \psi$ and $s \prec t$

$\mathcal{M}, w \models \varphi \preceq_{\forall\exists} \psi$ iff for all $s$ there is a $t$ such that $\mathcal{M}, s \models \varphi$ implies $\mathcal{M}, t \models \psi$, and $s \preceq t$

# From Worlds to Sets, I

$\mathcal{M}, w \models \varphi \preceq_{\exists\exists} \psi$ iff there is $s, t$ such that $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, t \models \psi$ and $s \preceq t$

$\mathcal{M}, w \models \varphi \prec_{\exists\exists} \psi$ iff there is $s, t$ such that $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, t \models \psi$ and $s \prec t$

$\mathcal{M}, w \models \varphi \preceq_{\forall\exists} \psi$ iff for all $s$ there is a $t$ such that $\mathcal{M}, s \models \varphi$ implies $\mathcal{M}, t \models \psi$, and $s \preceq t$

$\mathcal{M}, w \models \varphi \preceq_{\forall\exists} \psi$ iff for all $s$ there is a $t$ such that $\mathcal{M}, s \models \varphi$ implies $\mathcal{M}, t \models \psi$, and $s \prec t$

# From Worlds to Sets, II

$$\varphi \preceq_{\exists\exists} \psi \;:=\; E(\varphi \wedge \Diamond^{\preceq}\psi)$$

# From Worlds to Sets, II

$$\varphi \preceq_{\exists\exists} \psi \;:=\; E(\varphi \wedge \Diamond^{\preceq}\psi)$$

$$\varphi \prec_{\exists\exists} \psi \;:=\; E(\varphi \wedge \Diamond^{\prec}\psi)$$

# From Worlds to Sets, II

$$\varphi \preceq_{\exists\exists} \psi \ := \ E(\varphi \wedge \Diamond^{\preceq}\psi)$$

$$\varphi \prec_{\exists\exists} \psi \ := \ E(\varphi \wedge \Diamond^{\prec}\psi)$$

$$\varphi \preceq_{\forall\exists} \psi \ := \ A(\varphi \rightarrow \Diamond^{\preceq}\psi)$$

# From Worlds to Sets, II

$$\varphi \preceq_{\exists\exists} \psi \;:=\; E(\varphi \wedge \Diamond^{\preceq}\psi)$$

$$\varphi \prec_{\exists\exists} \psi \;:=\; E(\varphi \wedge \Diamond^{\prec}\psi)$$

$$\varphi \preceq_{\forall\exists} \psi \;:=\; A(\varphi \rightarrow \Diamond^{\preceq}\psi)$$

$$\varphi \prec_{\forall\exists} \psi \;:=\; A(\varphi \rightarrow \Diamond^{\prec}\psi)$$

# From Worlds to Sets, III

$\mathcal{M}, w \models \varphi \preceq_{\forall\forall} \psi$ iff for all $s$, for all $t$, $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, t \models \psi$ implies $s \preceq t$

# From Worlds to Sets, III

$\mathcal{M}, w \models \varphi \preceq_{\forall\forall} \psi$ iff for all $s$, for all $t$, $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, t \models \psi$ implies $s \preceq t$

$\mathcal{M}, w \models \varphi \prec_{\forall\forall} \psi$ iff for all $s$, for all $t$, $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, t \models \psi$ implies $s \prec t$

# From Worlds to Sets,IV

$$\varphi \preceq_{\forall\forall} \psi \;:=\; A(\psi \rightarrow \Box^{\preceq}\neg\varphi)$$

# From Worlds to Sets,IV

$$\varphi \preceq_{\forall\forall} \psi \ := \ A(\psi \to \Box^{\preceq}\neg\varphi)$$

$$\varphi \prec_{\forall\forall} \psi \ := \ A(\psi \to \Box^{\prec}\neg\varphi)$$

# From Worlds to Sets,IV

$$\varphi \preceq_{\forall\forall} \psi \ := \ A(\psi \rightarrow \Box^{\preceq} \neg\varphi)$$

$$\varphi \prec_{\forall\forall} \psi \ := \ A(\psi \rightarrow \Box^{\prec} \neg\varphi)$$

*We must assume the ordering* $\preceq$ *is total*

# From Sets to Worlds

$P_1 \gg P_2 \gg P_3 \gg \cdots \gg P_n$

$x > y$ 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.

# General Issues

Once a semantics and language are fixed, then standard questions can be asked: eg. develop a proof theory, completeness, decidability, model checking.

# General Issues

How should we *compare* the different logical systems?

- Embedding one logic in another:

# General Issues

How should we *compare* the different logical systems?

- Embedding one logic in another: *coalition logic* is a fragment of ATL ($tr([C]\varphi) = \langle\langle C \rangle\rangle \bigcirc \varphi$)

# General Issues

How should we *compare* the different logical systems?

- Embedding one logic in another: *coalition logic* is a fragment of ATL ($tr([C]\varphi) = \langle\langle C \rangle\rangle \bigcirc \varphi$)
- Compare different models for a fixed language:

# General Issues

How should we *compare* the different logical systems?

- Embedding one logic in another: *coalition logic* is a fragment of ATL ($tr([C]\varphi) = \langle\langle C\rangle\rangle \bigcirc \varphi$)
- Compare different models for a fixed language:
  - Alternating-Time Temporal Logics: Three different semantics for the ATL language.

V. Goranko and W. Jamroga. *Comparing Semantics of Logics for Multiagent Systems*. KRA, 2004.

# General Issues

How should we *compare* the different logical systems?

- Embedding one logic in another: *coalition logic* is a fragment of ATL ($tr([C]\varphi) = \langle\langle C \rangle\rangle \bigcirc \varphi$)

- Compare different models for a fixed language:
  - Alternating-Time Temporal Logics: Three different semantics for the ATL language.

V. Goranko and W. Jamroga. *Comparing Semantics of Logics for Multiagent Systems*. KRA, 2004.

- Comparing different frameworks:

# General Issues

How should we *compare* the different logical systems?

- Embedding one logic in another: *coalition logic* is a fragment of ATL ($tr([C]\varphi) = \langle\langle C\rangle\rangle \bigcirc \varphi$)

- Compare different models for a fixed language:
  - Alternating-Time Temporal Logics: Three different semantics for the ATL language.

V. Goranko and W. Jamroga. *Comparing Semantics of Logics for Multiagent Systems*. KRA, 2004.

- Comparing different frameworks: eg. PDL vs. Temporal Logic, PDL vs. STIT, STIT vs. ATL, etc.

# General Issues

How should we *merge* the different logical systems?

# General Issues

How should we *merge* the different logical systems?

- Combining logics is hard! ▸ Explain

D. Gabbay, A. Kurucz, F. Wolter and M. Zakharyaschev. *Many Dimensional Modal Logics: Theory and Applications*. 2003.

# General Issues

How should we *merge* the different logical systems?

- Combining logics is hard! ▸ Explain

D. Gabbay, A. Kurucz, F. Wolter and M. Zakharyaschev. *Many Dimensional Modal Logics: Theory and Applications*. 2003.

**Theorem** $\Box\varphi \leftrightarrow \varphi$ is provable in combinations of Epistemic Logics and PDL with certain "cross axioms" ($\Box[a]\varphi \leftrightarrow [a]\Box\varphi$) (and full substitution).

R. Schmidt and D. Tishkovsky. *On combinations of propositional dynamic logic and doxastic modal logics*. JOLLI, 2008.

# Merging Logics of Rational Agency

- Entangling Knowledge/Beliefs and Preferences
- "Epistemizing" Logics of Action and Ability
- BDI (Belief + Desires + Intentions) Logics

# Logics of Knowledge and Preference

$K(\varphi \succeq \psi)$: "Ann knows that $\varphi$ is at least as good as $\psi$"

$K\varphi \succeq K\psi$: "knowing $\varphi$ is at least as good as knowing $\psi$

# Logics of Knowledge and Preference

$K(\varphi \succeq \psi)$: "Ann knows that $\varphi$ is at least as good as $\psi$"

$K\varphi \succeq K\psi$: "knowing $\varphi$ is at least as good as knowing $\psi$

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

# Logics of Knowledge and Preference

$K(\varphi \succeq \psi)$: "Ann knows that $\varphi$ is at least as good as $\psi$"

$K\varphi \succeq K\psi$: "knowing $\varphi$ is at least as good as knowing $\psi$

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

J. van Eijck. *Yet more modal logics of preference change and belief revision*. manuscript, 2009.

F. Liu. *Changing for the Better: Preference Dynamics and Agent Diversity*. PhD thesis, ILLC, 2008.

$$A(\psi \rightarrow \langle \succeq \rangle \varphi) \quad \textit{vs.} \quad K(\psi \rightarrow \langle \succeq \rangle \varphi)$$

$$\mathsf{A}(\psi \rightarrow \langle\succeq\rangle\varphi) \quad \textit{vs.} \quad \mathsf{K}(\psi \rightarrow \langle\succeq\rangle\varphi)$$

*Should preferences be restricted to information sets?*

$$A(\psi \rightarrow \langle\succeq\rangle\varphi) \;\; \textit{vs.} \;\; K(\psi \rightarrow \langle\succeq\rangle\varphi)$$

*Should preferences be restricted to information sets?*

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

$K(\psi \rightarrow \langle\succeq \cap \sim\rangle\varphi)$

# Merging Logics of Rational Agency

- ✓ Entangling Knowledge/Beliefs and Preferences
- "Epistemizing" Logics of Action and Ability
- BDI (Belief + Desires + Intentions) Logics

# Knowing how to win

Consider the following game: Two cards, Ace and Joker, lie face down and the agent i must choose one. The Ace wins, the Joker loses.

# Knowing how to win

Consider the following game: Two cards, Ace and Joker, lie face down and the agent i must choose one. The Ace wins, the Joker loses.

- Does the agent i have a strategy to win the game?

## Knowing how to win

Consider the following game: Two cards, Ace and Joker, lie face down and the agent i must choose one. The Ace wins, the Joker loses.

- Does the agent i have a strategy to win the game?
- Does the agent i know that she has a strategy to win the game?

# Knowing how to win

Consider the following game: Two cards, Ace and Joker, lie face down and the agent i must choose one. The Ace wins, the Joker loses.

- Does the agent i have a strategy to win the game?
- Does the agent i know that she has a strategy to win the game?
- Does the agent i know a strategy to win the game?

J. Fantl. *Knowing-how and knowing-that*. Philosophy Compass, 3 (2008), 451 470.

M.P. Singh. *Know-how*. In Foundations of Rational Agency (1999), M. Wooldridge and A. Rao, Eds., pp. 105 132.