# Advanced Modal Logic IX

Yanjing Wang

Department of Philosophy, Peking University

Mar. 17th, 2020

Advanced Modal Logic (2020 Spring)

# Expressive power and distinguishing power

Logical languages can express properties of mathematical structures (models). The standard notions for comparing how much logical languages can say about models are *distinguishing power* (can a language tell the difference between two models?), and *expressive power* (which classes of models can be defined by a formula of the language?).

# Distinguishing power vs. Expressive power

Let $\mathcal{L}_1 = \langle L_1, \mathcal{C}, \vDash \rangle$, $\mathcal{L}_2 = \langle L_2, \mathcal{C}, \Vdash \rangle$ be two logics defined on the *same* class of models $\mathcal{C}$.
We say $\mathcal{L}_2$ is at least as *distinguishing* as $\mathcal{L}_1$ ($\mathcal{L}_1 \preccurlyeq_d \mathcal{L}_2$) iff

$$\forall \mathcal{M}, w, \mathcal{N}, v \in \mathcal{C} : \mathcal{M}, w \parallel_{L_1} \mathcal{N}, v \textit{ implies } \mathcal{M}, w \parallel_{L_2} \mathcal{N}, v.$$

where $\mathcal{M}, w \parallel_L \mathcal{N}, v$ denotes that $\mathcal{M}, w \not\equiv_L \mathcal{N}, v$. Note that $\mathcal{L}_1 \preccurlyeq_d \mathcal{L}_2$ iff $\equiv_{\mathcal{L}_2} \subseteq \equiv_{\mathcal{L}_1}$.

We say $\mathcal{L}_2$ is at least as *expressive* (in defining properties i.e. classes of models) as logic $\mathcal{L}_1$ on $\mathcal{M}$ ($\mathcal{L}_1 \preccurlyeq_e \mathcal{L}_2$) iff

$$\forall \varphi_1 \in L_1 \exists \varphi_2 \in L_2 : \forall \mathcal{M}, w \in \mathcal{C} : \mathcal{M}, w \vDash \varphi_1 \textit{ iff } \mathcal{M}, w \Vdash \varphi_2$$

# Distinguishing power vs. Expressive power

We say $\mathcal{L}_1$ and $\mathcal{L}_2$ are equally distinguishing ($\mathcal{L}_1 \approx_d \mathcal{L}_2$) if $\mathcal{L}_1 \preccurlyeq_d \mathcal{L}_2$ and $\mathcal{L}_2 \preccurlyeq_d \mathcal{L}_1$. Similarly, $\mathcal{L}_1$ and $\mathcal{L}_2$ are equally expressive ($\mathcal{L}_1 \approx_e \mathcal{L}_2$) if $\mathcal{L}_1 \preccurlyeq_e \mathcal{L}_2$ and $\mathcal{L}_2 \preccurlyeq_e \mathcal{L}_1$.
It is not hard to prove that $\mathcal{L}_1 \preccurlyeq_e \mathcal{L}_2$ implies $\mathcal{L}_1 \preccurlyeq_d \mathcal{L}_2$, but the converse may fail (e.g., compare the propositional logic and its syntactic fragment with proposition letters only: they have the same distinguishing power but different expressive power).

# Distinguishing power vs. Expressive power

Therefore showing that there is a pair of models that one logic can distinguish but the other one cannot is sufficient to demonstrate that these two logics have *different* expressive power. Actually, we have seen such ideas before: if a property can “distinguish” two models that are modally-equivalent (or bisimilar) then this property cannot be expressed by a modal formula. On the other hand, in general, by showing that there is no such a pair, we cannot prove immediately that the two logics have the same expressive power. However, there are cases where the comparison of expressive power can be reduced to the comparison of distinguishing power (exercise). Modal Logic is also such an example when compared to FOL.

# ML as a fragment of FOL

Kripke models $\langle W, \{R_\nabla\}_{\nabla \in \mathbf{O}}, V\rangle$ can be also viewed as first-order structures with interpretation of relation symbols and predicates: $W$ is the domain, $R_\nabla$ is the interpretation for a relation symbol and $V(p)$ is the interpretation of a predicate.
Thus we may use first-order formulas (with one free variable, why?) to express the meaning of modal formulas.
For example, $\Diamond\Box p$ corresponds to the following first-order formula $\alpha(x)$:

$$\exists y(xRy \land \forall z(yRz \to Pz))$$

# Standard translation

**Definition (First-order language of $\mathtt{FOL}_\tau(\mathbf{P})$)**

Given $\mathbf{P}$, and a modal similarity type $\tau = \{\mathbf{O}, \rho\}$, the first order language (with equality) $\mathtt{FOL}_\tau(\mathbf{P})$ has unary predicates $P$ (corresponding to $p$ in $\mathbf{P}$) and $n + 1$-ary relation symbols $R_\nabla$ for each $\nabla \in \mathbf{O}$ such that $\rho(\nabla) = n > 0$.

$$\phi \quad ::= \quad Px \mid x \approx x \mid R_\nabla \underbrace{x \ldots x}_{\rho(\nabla)+1} \mid (\phi \wedge \phi) \mid \neg\phi \mid \forall x \phi$$

We write $\alpha(x)$ for a first-order formula with one free variable $x$.

# Standard translation

## Definition (Standard translation)

$ST : \mathtt{ML}_\tau(\mathbf{P}) \to \mathtt{FOL}_\tau(\mathbf{P})$:

$$
\begin{aligned}
ST_x(p) &= Px \\
ST_x(\top) &= x \approx x \\
ST_x(\neg\phi) &= \neg ST_x(\phi) \\
ST_x(\phi \wedge \psi) &= ST_x(\phi) \wedge ST_x(\psi) \\
ST_x(\nabla(\phi_1,\ldots,\phi_k)) &= \forall y_1 \forall y_2 \ldots \forall y_k (R_\nabla x y_1 y_2 \ldots y_k \to \\
&\quad ST_{y_1}(\phi_1) \vee \cdots \vee ST_{y_k}(\phi_k))
\end{aligned}
$$

## Example

$ST_x(\Diamond p) = ST_x(\neg\Box\neg p) = \neg ST_x(\Box\neg p)$
$= \neg(\forall y(xRy \to ST_y(\neg p)) = \neg(\forall y(xRy \to \neg Py)) = \exists y(xRy \wedge Py)$

# Standard translation

We use $\Vdash$ for the satisfaction relation for first-order formulas.

**Theorem (Local and global correspondence on models)**

(1) $\mathcal{M}, w \vDash \phi \iff \mathcal{M} \Vdash ST_x(\phi)[w]$
(2) $\mathcal{M} \vDash \phi \iff \mathcal{M} \Vdash \forall x ST_x(\phi)$

**By induction on $\phi$.**

Given a binary similarity type, we only show (1).
Boolean cases are trivial. Consider $\phi = \Box_a \psi$:

$$\mathcal{M}, w \vDash \phi \iff \text{for all } v : w \xrightarrow{a} v \text{ implies } \mathcal{M}, v \vDash \psi$$
$$\iff \text{for all } v : w \xrightarrow{a} v \text{ implies } \mathcal{M} \Vdash ST_y(\psi)[v]$$
$$\iff \mathcal{M} \Vdash \forall y (x R_a y \to ST_y(\psi))[w] \qquad \Box$$

# Standard translation

**Definition (Standard translation: $n+1$-variable version, $n$ is the maximal arity of operators in $\tau$)**

Given a unary similarity type $\tau$, $ST : \mathtt{ML}_\tau(\mathbf{P}) \rightarrow \mathtt{FOL}_\tau(\mathbf{P})$ with only two variables $x, y$:

$$
\begin{array}{lclllcl}
ST_x(p) & = & Px & & ST_y(p) & = & Py \\
ST_x(\top) & = & x = x & & ST_y(\top) & = & y = y \\
ST_x(\neg\phi) & = & \neg ST_x(\phi) & & ST_y(\neg\phi) & = & \neg ST_y(\phi) \\
ST_x(\phi \wedge \psi) & = & ST_x(\phi) \wedge ST_x(\psi) & & ST_y(\phi \wedge \psi) & = & ST_y(\phi) \wedge ST_y(\psi) \\
ST_x(\Box_a\phi) & = & \forall y(xR_a y \rightarrow ST_y(\phi)) & & ST_y(\Box_a\phi) & = & \forall x(yR_a x \rightarrow ST_x(\phi))
\end{array}
$$

Can we also translate a FO formula to a modal formula?

# Heritages from FOL

**Theorem (Compactness Theorem)**

*Every finitely satisfiable set of modal formulas is satisfiable.*

**Theorem (Löwenheim-Skolem Theorem)**

*if a set of modal formulas is satisfiable in at least one infinite model, then it is satisfiable in models of every infinite cardinality (assuming the modal language is countable).*

# ML as a proper fragment of FOL

By Standard Translation, every formula in $\mathtt{ML}_\tau(\mathbf{P})$ is equivalent to a formula $\alpha(x)$ in the corresponding first-order language $\mathtt{FOL}_\tau(\mathbf{P})$ with one free variable. Clearly, the converse is not true (e.g., consider $xRx$).Thus $\mathtt{ML}_\tau(\mathbf{P})$ can be seen as a proper fragment of $\mathtt{FOL}_\tau(\mathbf{P})$. Now we may ask: which $\alpha(x)$ is equivalent to a modal formula?

In other words, we would like to “characterize” the fragment of $\mathtt{FOL}_\tau(\mathbf{P})$ which corresponds to $\mathtt{ML}_\tau(\mathbf{P})$.

# A simple characterization

Clearly such $\alpha(x)$ needs to be invariant under modal equivalence $\equiv_{\mathtt{ML}}$, namely: for any $\mathcal{M}, w, \mathcal{N}, v$:
if $\mathcal{M}, w \equiv_{\mathtt{ML}} \mathcal{N}, v$ then $\mathcal{M} \Vdash \alpha(x)[w] \iff \mathcal{N} \Vdash \alpha(x)[v]$.

**Theorem (A characterization via $\equiv_{\mathtt{ML}}$)**

*Let $\alpha(x)$ be a first-order formula with one free variable in $\mathtt{FOL}_\tau(\mathbf{P})$. $\alpha(x)$ is invariant under modal equivalence iff it is equivalent to (the standard translation of) a modal formula.*

**Corollary**

*Let $\alpha(x)$ be a first-order formula with one free variable in $\mathtt{FOL}_\tau(\mathbf{P})$. $\alpha(x)$ is invariant under $\omega$-bisimilarity iff it is equivalent to (the standard translation of) a modal formula.*

Do we need to restrict the language in the above corollary?

# A simple characterization

**Proof.**

Let $MOC(\alpha(x)) = \{ST_x(\phi) \mid \alpha(x) \Vdash ST_x(\phi) \text{ and } \phi \in \mathtt{ML}\}$.
Claim 1: If $MOC(\alpha(x)) \Vdash \alpha(x)$ then there is a modal formula $\phi$ such that $ST_x(\phi)$ is equivalent to $\alpha(x)$.
Claim 2: $MOC(\alpha(x)) \Vdash \alpha(x)$ is indeed true if $\alpha(x)$ is invariant for modal equivalence.
The first claim can be proved by an argument based on compactness.
For the second claim: suppose $\mathcal{M}, w \Vdash MOC(\alpha(x))$ then we collect all the modal formulas that are true on $\mathcal{M}, w$ and then show the set of their first-order correspondences together with $\alpha(x)$ is satisfiable on some model $\mathcal{N}, v$ (again by a compactness argument). Since $\alpha(x)$ can not distinguish modally equivalent models then $\alpha(x)$ holds on $\mathcal{M}, w$. □

# ML as a proper fragment of FOL

The previous result characterizes ML within FOL by using $\equiv_{\mathtt{ML}}$ (or $\underline{\leftrightarrow}_{\omega}$), which is not as elegant as the following van Benthem characterization.

**Theorem (van Benthem Characterization Theorem)**

*Let $\alpha(x)$ be a first-order formula in $\mathtt{FOL}_{\tau}$. $\alpha(x)$ is invariant under bisimilarity iff it is equivalent to the standard translation of a modal formula.*

To prove this theorem based on the previous result, we only need to show that:
$\alpha(x)$ is invariant under bisimilarity iff $\alpha(x)$ is invariant for $\equiv_{\mathtt{ML}}$ (it is not trivial since $\equiv_{\mathtt{ML}} \neq \underline{\leftrightarrow}$ in general).

# A “detour” strategy

Since $\underline{\leftrightarrow} \subseteq \equiv_{\mathtt{ML}}$, we only need to prove that if $\alpha(x)$ is invariant under bisimilarity then it is invariant under modal equivalence. Now assume $\mathcal{M}, w \equiv_{\mathtt{ML}} \mathcal{N}, v$, $\alpha(x)$ is invariant under bisimilarity and $\mathcal{M} \Vdash \alpha(x)[w]$ we need to show $\mathcal{N} \Vdash \alpha(x)[v]$. The strategy is as follows:

$$
\begin{array}{ccccc}
1.\alpha(x) & \mathcal{M}, w & \equiv_{\mathtt{ML}} & \mathcal{N}, v & 4.\alpha(x) \\
 & \downarrow \equiv_{\mathtt{FOL}} & & \downarrow \equiv_{\mathtt{FOL}} & \\
2.\alpha(x) & \mathcal{M}^*, w^* & \equiv_{\mathtt{ML}} \; = \; \underline{\leftrightarrow} & \mathcal{N}^*, v^* & 3.\alpha(x)
\end{array}
$$

Based on $\mathcal{M}, w$ and $\mathcal{N}, v$ we construct m-saturated models $\mathcal{M}^*, w^*$ and $\mathcal{N}^*, v^*$ such that FOL formulas are preserved (thus modal formulas are preserved too). Since for m-saturated models $\underline{\leftrightarrow}$ coincides with $\equiv_{\mathtt{ML}}$, $\mathcal{M}^*, w^* \; \underline{\leftrightarrow} \; \mathcal{N}^*, v^*$.