Philosophy 244: Bisimulations and the Standard Translation

Guest Lecture: Cosmo Grant, October 26, 2016

The Plan

Sources: *Modal Logic for Open Minds*, by Johan van Benthem, and *Modal Logic*, by Blackburn, de Rijke and Venema. They're superb—check them out!

We'll look at some more metatheory of propositional modal logic. Our previous work on propositional modal logic focused on *frames*. Today we'll focus on *models*. The central new concept is that of a *bisimulation* between two models. We'll use bisimulations to prove some results about the *expressive power* of modal logic. And bisimulations turn out to be exactly what we need to connect modal logic with first-order logic. Here's the plan:

Bisimulations, under the name *p-relations*, were introduced by Johan van Benthem in his PhD thesis.

1. Bisimulations: definition and examples
2. Bisimulations and modal equivalence
3. Bisimulations and expressive power
4. Bisimulations and the Standard Translation

Bisimulations: definition and examples

Definition 1 (Modal equivalence). Let's say that (M, w), (M', w') are *modally equivalent* just if for every formula ϕ, $(M, w) \models \phi$ iff $(M', w') \models \phi$.

notation: $(M, w) \leftrightsquigarrow (M', w')$

When are (M, w), (M', w') modally equivalent? Well, if there is an *isomorphism* $f : W \to W'$ then (M, w), (M', w') will certainly be modally equivalent. But an isomorphism is over-kill: it guarantees modal equivalence but in a heavy-handed way. Can we do better?

A bijection $f : W \to W'$ is an *isomorphism* from M to M' just if for all w, v in M:

1. w and $f(w)$ satisfy the same proposition letters
2. Rwv iff $R'f(w)f(v)$

Definition 2 (Bisimulation). Let $M = (W, R, V)$, $M' = (W', R', V')$ be models. A non-empty binary relation $Z \subseteq W \times W'$ is a *bisimulation between M and M'* just if the following conditions are satisfied:

1. *Atomic harmony:* If wZw' then w and w' satisfy the same proposition letters.
2. *Forth condition:* If wZw' and Rwv then there exists v' such that $R'w'v'$ and vZv'.

3. *Back condition:* If wZw' and $R'w'v'$ then there exists v such that vZv' and Rwv.

If there is a bisimulation between (M, w) and (M', w'), we'll write $(M, w) \leftrightarroweq (M', w')$, and say that w, w' are *bisimilar*.

Bisimulations and modal equivalence

Proposition 1 (Bisimilarity implies modal equivalence). If $(M, w) \leftrightarroweq (M', w')$ then $(M, w) \leftrightsquigarrow (M', w')$.

Proof. We need to show that for all formulas ϕ and models (M, w), (M', w'), if $(M, w) \leftrightarroweq (M', w')$ then $(M, w) \models \phi$ iff $(M', w') \models \phi$. We proceed by strong induction on complexity. Fix a formula ϕ.

What would have gone wrong if we'd *fixed* arbitrary models, proceeded by induction and then generalized afterwards?

Induction Hypothesis: For every shorter formula ψ and all models (M, w), (M', w'), if $(M, w) \leftrightarroweq (M', w')$ then $(M, w) \models \psi$ iff $(M', w') \models \psi$.

Case 1: ϕ is atomic. The result follows from atomic harmony.

Case 2: ϕ is a Boolean combination of shorter formulas. The result is by IH and the satisfaction definitions.

Case 3: ϕ is $\diamond\psi$. Suppose $(M, w) \leftrightarroweq (M', w')$ and $(M, w) \models \diamond\psi$. So there exists v in M such that Rwv and $(M, v) \models \psi$. Since $(M, w) \leftrightarroweq (M', w')$, the forth condition tells us that there exists v' in M' such that $R'w'v'$ and $(M, v) \leftrightarroweq (M', v')$. By IH, $(M', v') \models \psi$. So $(M', w') \models \diamond\psi$. The converse is similar. □

Are worlds which satisfy the same modal formulas always bisimilar? After all, bisimilarity seems to perfectly reflect the satisfaction definitions. Perhaps we can use the modal formulas to construct a bisimulation between modally equivalent worlds. Not so!

Proposition 2 (Modal equivalence doesn't imply bisimilarity). Take a world w. For each n, add an R-branch of length n coming out of w. That's M. Now take a world w'. For each n, add an R'-branch of length n coming out of w', as before. But *also* add an infinite R'-branch coming out of w'. That's M'. Now, $(M, w) \leftrightsquigarrow (M', w')$. (Why?) But there is no bisimulation between (M, w) and (M', w'). (Why not?)

We *can*, however, get a restricted converse to Proposition 1. A preliminary definition: a model M is *image-finite* if each world in M can see only finitely many worlds.

Neither model in the previous example is image-finite, because w, w' each see infinitely many worlds.

Proposition 3 (Modal equivalence implies bisimilarity in image-finite models). Let M, M' be image-finite models. For all $w \in W$, $w' \in W'$, if $(M, w) \leftrightsquigarrow (M', w')$ then $(M, w) \leftrightarrows (M', w')$.

Proof. We will show that the relation of modal equivalence is itself a bisimulation. That is, define $Z := \{(w, w') \in W \times W' : (M, w) \leftrightsquigarrow (M', w')\}$. We need to check that Z satisfies our three conditions.

Atomic harmony is immediate. Suppose for contradiction that Rwv and wZw' but there is no v' in M' such that $Rw'v'$ and vZv'. Let S' be the set of successors of w'. Note that S' must be non-empty (why?) and finite, say $S' = \{w'_1, w'_2, w'_3, \ldots\}$. By assumption, for each $w'_i \in S'$, there exists a formula ψ_i such that $(M, v) \models \psi$ but $(M', w'_i) \not\models \psi$. Hence:

$$(M, w) \models \diamond(\psi_1 \wedge \psi_2 \wedge \ldots \wedge \psi_n) \text{ but } (M', w') \not\models \diamond(\psi_1 \wedge \psi_2 \wedge \ldots \wedge \psi_n).$$

Contradiction. That verifies the forth condition. Checking the back condition is similar. □

Bisimulations and expressive power

What properties of models can we talk about within the modal language? First, we need to make talk about 'talking about' precise.

Definition 3 (Definability). Let P be a property of (pointed) models. Let's say that a formula ϕ *defines* P just if, for any model (M, w), we have $(M, w) \models \phi$ just if (M, w) has property P. A property P is *definable* if there is some formula which defines it.

"ϕ picks out precisely the P-models"

Bisimulations give us a general strategy for showing that a property is not definable: find bisimilar models such that one has the property and the other doesn't.

The property cuts finer than the modal language!

For example, let's say that (M, w) is *reflexive* just if Rww. Is reflexivity definable (over pointed models)? No! Can you find suitable models?

Of course, reflexivity *is* definable over *frames*. Witness: $\Box p \supset p$.

From a negative point of view, such results tell us about the limits on what the modal language can say. From a positive point of view, they suggest ways of transforming models without affecting satisfiability—a handy tool.

Bisimulations and the Standard Translation

The modal language is interpreted over *graph-like structures*. We can *also* interpret a first-order language $\mathcal{L}$ over such structures. That's interesting, because then we can compare first-order logic and modal logic.

What is $\mathcal{L}$? The non-logical symbols of $\mathcal{L}$ are the unary predicates P_0, $P_1, \ldots$, corresponding to the proposition letters p_0, $p_1, \ldots$, and a binary relation symbol R.

We interpret $\mathcal{L}$ using an ordinary modal model—no adjustment needed!

The domain is W; the extension of P_i is the set of worlds at which p_i is true; the binary relation symbol R is interpreted by the accessibility relation.

Definition 4 (Standard Translation). We define the *standard translation ST_x* taking modal formulas to first order formulas with one free variable x, as follows:

1. $ST_x(p) = Px$
2. $ST_x(\neg\phi) = \neg ST_x(\phi)$
3. $ST_x(\phi \vee \psi) = ST_x(\phi) \vee ST_x(\psi)$
4. $ST_x(\diamond\phi) = \exists y(Rxy \wedge ST_y(\phi))$, where y is a fresh variable

Actually, if we're a bit cleverer, we can make sure that the translations only use *two* variables. That also removes the indeterminacy in the current definition due to the free choice of fresh variables.

Here is the fundamental fact:

Proposition 4 (Local Correspondence on Models). For all $\mathcal{M}$ and all states w of M: $(M, w) \models \phi$ iff $M \models ST_x(\phi)[w]$.

The theorem provides "a bridge between modal and first-order logic—and we can use this bridge to import results, ideas and proof techniques from one to the other."

For example, we can use it to prove the compactness over models of modal logic.

Which first-order formulas are translations of modal formulas (up to logical equivalence)? That question is answered by van Benthem's Theorem. We need a preliminary definition:

Definition 5 (Invariant for bisimulations). A first-order formula $\alpha(x)$ in $\mathcal{L}$ is *invariant for bisimulations* if for all models M, N and all states $w \in M$, $v \in N$, and all bisimulations Z between M and N such that wZv, we have $M \models \alpha(x)[w]$ iff $N \models \alpha(x)[v]$.

Theorem 1 (van Benthem Characterization Theorem). Let $\alpha(x)$ be a first order formula in $\mathcal{L}$. Then $\alpha(x)$ is equivalent to the standard translation of a modal formula just if it's invariant for bisimulations.