499H Modal logic

Ian Hodkinson

Department of Computing
Imperial College London

499 home page:

`www.doc.ic.ac.uk/~imh/teaching/499_mtl/499.html`

Modal logic . . .

- is a common logical way of handling the notions of *necessity, possibility, knowledge, belief, change, time,* etc ('modalities')
- gives an alternative to first-order logic for these ends: is different in expressive power, arguably more natural to work with, and has computational advantages: better-bounded resources, complexity lower
- is usable with first-order logic anyway (it can subsume FO logic!)
- has a long, distinguished history (from Aristotle), and is interdisciplinary in application (philosophy, linguistics as well as computer science)

Being one of the basic species (rather, phyla) of logic, it is used in a gamut of applications: *program specification & verification, program semantics, concurrent programs, communication protocols, specification of rational agents, reasoning about knowledge and actions, natural language. . .*

Course outline

This half-course is an introduction (~ 9 lectures) to working with modal logics:

- syntax and semantics
- Sahlqvist correspondence: frame properties from modal axioms
- p-morphisms and bisimulations
- modal μ-calculus (if time)

Assessment

1. One assessed coursework
2. Exam in December: 1-hour paper, do 2 questions out of 3.

There will be lots of *unassessed exercises* (with solutions later). Some go beyond the course. But the more you do, the stronger you'll get.
I recommend you try a lot of them, in the tutorial and at home.

Modal logic is not a spectator sport.

Textbooks

- R. Goldblatt, *Logics of time and computation,* 2nd. edn., CSLI Publications, 1992. *Free pdf* at `http://sul-derivatives.stanford.edu/derivative?CSNID=00003782&mediaType=application/pdf`
  Terse but authoritative. Recommended. Library has copies.
- P. Blackburn, M. de Rijke, Y. Venema, *Modal logic,* Cambridge University Press, 2002. Very good modern text — $\geq$ £45 in paperback. Library has about 9 copies.
- J. van Benthem, *Modal logic for open minds.* Recent book, *free pdf* at `http://fenrong.net/teaching/mljvb.pdf`
- A. Chagrov, M. Zakharyaschev, *Modal logic,* Oxford University Press, 1997. Very good thorough text.
- G.E. Hughes, M.J. Cresswell, *A new introduction to modal logic,* Routledge, 1996. A more philosophically-oriented book.
- Survey: `http://www.mcs.vuw.ac.nz/~rob/papers/modalhist.pdf`
- Others in library — go and look.

Omissions, prerequisites

Omissions

We haven't time to cover all important topics in modal logic. (One is *complexity.*) See the books for more information on omitted areas.

Prerequisites

This is *not* intended as a first course in logic. A good grasp of propositional and first-order logic will help. Suggested reading:

- W. Hodges, *Logic,* Penguin, 1977
- E.J. Lemmon, *Beginning Logic,* Van Nostrand, 1965.

Or go to Comp Sci MSc course *518 Logic and AI Programming.*

1. Syntax

Syntax concerns things we write down: here, formulas and their formation rules.
It is not concerned with meaning — we'll get to that later.

Fix, throughout, a (countably infinite) set L of propositional atoms.
They stand for 'basic facts' (e.g., 'It is raining').
We typically write p, q, r, or $p_0, p_1, p_2, \ldots,$ for atoms.

Modal formulas are built from atoms and $\top$ ('truth') using the boolean connectives $\wedge, \neg$ ('and', 'not'), and the modal connective $\Box$ (the meaning of $\Box$ will be discussed soon).
More formally:

Modal formulas

Definition 1.1 (basic modal formulas)

- Any propositional atom is an (L-)formula.
- $\top$ is also a formula.
- If A, B are formulas, then so are:
  $$\neg A \qquad (A \wedge B) \qquad \Box A$$
  (pronounced 'not A', 'A and B', 'box A').
- Nothing else is a formula.

Example: $\Box(p \wedge \Box\neg(q \wedge \top))$ is a modal formula.
We write $A, B, \ldots$ for arbitrary L-formulas.
No quantifiers. It is propositional logic plus a new connective, $\Box$.

Temporal formulas

Temporal logic is a type (class) of modal logic, for handling time.

Temporal formulas (in basic temporal logic) use G and H (future and past) instead of $\Box$. '$\Box A$' above is replaced by:

$$GA \qquad HA$$

There are other ways of making modal/temporal logics. See later.

Abbreviations

- $\bot$ (pronounced 'falsity') abbreviates $\neg\top$
- $A \vee B$ ('A or B') abbreviates $\neg(\neg A \wedge \neg B)$
- $A \to B$ ('A implies B') abbreviates $\neg(A \wedge \neg B)$
- $A \leftrightarrow B$ ('A if & only if B') abbreviates $(A \to B) \wedge (B \to A)$

and also, crucially,

- $\Diamond A$ (pronounced 'diamond A') abbreviates $\neg\Box\neg A$
- FA abbreviates $\neg G \neg A$
- PA abbreviates $\neg H \neg A$

Occasionally we will treat $\bot, \vee, \Diamond$ as primitive symbols (not abbreviations).

Binding conventions

In decreasing order of tightness, $\neg, \Box, \Diamond, G, H, F, P$ (these are unary, so their mutual order is immaterial), then $\wedge, \vee, \to, \leftrightarrow$.

So $\Diamond p \wedge \Box\neg q \to r$ means $((\Diamond p) \wedge (\Box\neg q)) \to r$.
But always use brackets if in any doubt!

Meaning

We know that $\wedge, \vee, \neg, \rightarrow, \leftrightarrow$ mean 'and', 'or', 'not', 'if-then', 'if and only if'.

What about $\Box$ and $\Diamond$, G and H, F and P?

There are many meanings, depending on the application.

(Most of the course concerns this variation!)

Some examples

Informal readings of $\Box A$	**Informal readings of $\Diamond A$**
A is necessary	A is possible
A is always true	A is sometimes true
A is known	A *is thought possible?*
A is believed	A *is conceivable??*
A is obligatory	A is permitted
A is provable	A is consistent
A will be true whenever the program terminates	The program can terminate with A true.

Temporal examples

Informal reading of GA	**Informal reading of FA**
A will always be true in future	A is true at some future time

Informal reading of HA	**Informal reading of PA**
A has always been true in the past	A was true at some past time

F, P, G, H is Arthur Prior's notation. Mnemonic:

- GA = 'A is Going to be true'.
- HA = 'A Has always been true' (Historically true?)

Can use other connectives: eg, Tomorrow, Yesterday, Until, Since.

$\Box, \Diamond$ have different flavours:

- $\Box, G, H$ are like $\forall$ ('for all').
- $\Diamond, F, P$ are like $\exists$ ('there exists').

Examples of formulas

Some of these are 'inevitably true'. Which?

Formula	**Reading(s)**
$\Box\Box A$	A is known to be known. It is necessary that A is necessary.
$\Box\Diamond A$	It is necessary that A is possible.
$\Box A \to A$	If A is necessarily true, then it is true. If A is known, then A is true. If A is believed, then A is true.
$\Box\Box A \to \Box A$	If it is known that A is known, then A is known.
$HHA \to HA$	If A was always always true then it was always true.
$\Box A \to \Box\Box A$	If the agent knows A then it knows that it knows A.
$A \to GPA$	if A is true, then always in future, A will have been true at some past time.
$\Diamond\neg PA$	Possibly, A was never true.
$\Box A \wedge \Box B \to \Box(A \wedge B)$	if A, B are believed, then so is $A \wedge B$.

Other modal logics

'. . . the view that modal logic amounts to rather simple-minded uses of $\Box$ and $\Diamond$ . . . has been out of date for at least 30 years'

Propositional dynamic logic (PDL)

Fix a set $\mathcal{P}$ of non-deterministic, possibly non-terminating *programs.* We require that:

- Various basic programs $a, b, c, \ldots$ are in $\mathcal{P}$,

and if $\pi_1, \pi_2 \in \mathcal{P}$ then:

- $\pi_1 \cup \pi_2 \in \mathcal{P}$ (non-deterministically choose to execute π_1 or π_2)
- $\pi_1 \,;\, \pi_2 \in \mathcal{P}$ (do π_1 then π_2)
- $\pi_1^* \in \mathcal{P}$ (execute π_1 a finite number (possibly zero) of times).

So $\mathcal{P}$ contains some arbitrary basic programs and is *closed* under certain program-forming operations.

PDL-formulas

View atoms (in L) as basic statements about *states* of a machine that runs programs in $\mathcal{P}$.
For each $\pi \in \mathcal{P}$, introduce a box $[\pi]$.

- Any atom $p \in L$ is a PDL-formula.
- If A, B are PDL-formulas and $\pi \in \mathcal{P}$ then $A \wedge B$, $\neg A$, and $[\pi]A$ are PDL-formulas.

Idea: $[\pi]A$ means 'A will hold after every halting run of π'.

Adopt earlier abbreviations.
Now, $\langle\pi\rangle A$ abbreviates $\neg[\pi]\neg A$.
It means 'A will hold after some halting run of π'.

‘Truths’ in PDL

We’d expect

$$\langle \pi^* \rangle A \leftrightarrow A \vee \langle \pi \,;\, \pi^* \rangle A$$

(for any formula A and any $\pi \in \mathcal{P}$) to be always true.
How about Segerberg’s axiom,

$$[\pi^*](A \to [\pi]A) \to (A \to [\pi^*]A) \quad ?$$

More PDL-formulas

Can allow another program-formation rule:

- If A is a PDL-formula then $A? \in \mathcal{P}$.

$A?$ is a program that tests to see if A is true in the current state. If so, it halts. If not, it ‘hangs’ (doesn’t terminate). Now,

$$(A? \,;\, \pi_1) \cup ((\neg A)? \,;\, \pi_2) \in \mathcal{P}$$

is the (non-deterministic) program ‘if A then π_1 else π_2’.

Natural and contingent truth

- Truth (loosely speaking) of arbitrary formulas, like $\Box\Diamond A$, will depend on the fact expressed by A as well as the meaning of $\Box$.
- But formulas like $\Box A \to A$ *may* (or may not) be true automatically *for any A* (whatever 'true' means), depending on the meaning we give to $\Box$.
  (This is one source of the variety within modal logic.)
  They are 'contingent truths' — like 'laws of chemistry'.
- Others, like $\Box A \to \Box A$ and $\Box(A \to B) \to (\Box A \to \Box B)$, seem to be true for any A, B, *whatever* reasonable ($\forall$-style) meaning we give to $\Box$. (Can you think of an exception?)
  They are 'natural truths' — like 'laws of physics'.
- And $A \to GPA$ must be true if past and future are opposites.
- Similarly, $\langle \pi^* \rangle A \leftrightarrow A \vee \langle \pi ; \pi^* \rangle A$ must be true in PDL.

What's going on?

2. Formal semantics

Semantics is meaning. Formal semantics gives a formal (= mathematical) meaning to our modal formulas.
Why do we want formal semantics for modal logic? Why are the intuitive notions of ‘possible’, ‘known’, etc., not good enough?

- We want to distinguish between natural truths, contingent truths, and other formulas, in a bid to understand this area of human reasoning. So we need to be very *precise* about the meaning of ‘truth’.
- The notions of ‘possible’, etc., are not well understood. To model and study them with modal logic, we should use a totally *independent* definition of semantics.
- We want to *abstract* away from particular meanings of $\Diamond, \Box$, and arrive at their essentials. Then we can use modal logic in new applications.
- If we know exactly what modal logic is (and is not) talking about, we (or others) can use powerful tools to study it.

What will the semantics be like?

A good semantics is intuitive and easily understood, yet powerful enough to cover most of the variations in modal logics.
We will study only *'Kripke' semantics.* But there are many other semantics for modal logic. This is good, since:

- a particular semantics will probably force on us a particular notion of natural/contingent truths.
  Eg. in Kripke semantics, 'normality' holds: any formula $\Box(A \to B) \to (\Box A \to \Box B)$ is a 'natural truth'.
  In other semantics, it can fail.
- different applications *sometimes* need different (eg., non-normal) semantics.

Further reading: S. Kripke, *Naming and necessity,* Harvard University Press, 1980.

Kripke semantics

A revolution in modal logic

Dates from work of Saul Kripke (c. 1960) but has earlier antecedents (e.g., in work of Jónsson & Tarski, 1951).
Is now universally used — the most popular modal semantics.

Idea (Leibniz): 'possibly true' = *'true in some possible world'.*
New addition: which worlds are possible may depend on the 'current' world. E.g., in temporal logic, which worlds are in the future depends on 'now'.

So take a set W of possible worlds (or states).
At each one, list which worlds are possible from it.
Can do this using a binary relation R on W:

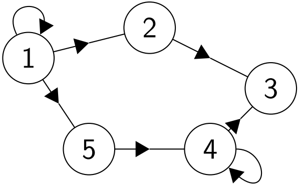
$R(t,u)$ means 'u is a possible world at t'.

Definition 2.1 (Kripke frame)

This much — (W,R), where W is a non-empty set and R a binary relation on it — is called a *(Kripke) frame.*
R is called the *accessibility relation.*

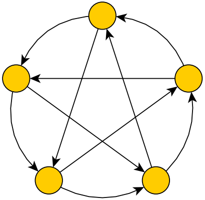
Commonly-used Kripke frames

- Simple examples, such as:



- $(\mathbb{N}, <), (\mathbb{Z}, <), (\mathbb{Q}, <), (\mathbb{R}, <)$. Here,
  $\mathbb{N} = \{0, 1, 2, \ldots\}$ (natural numbers),
  $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$ (integers, 'Zahlen'),
  $\mathbb{Q}$ is the set of rational numbers (p/q, for $p, q \in \mathbb{Z}$, $q \neq 0$),
  $\mathbb{R}$ is the set of real numbers (e.g., all numbers representable by decimal-point notation).
  The accessibility relation is $<$.
  $(\mathbb{N}, <)$ is common in temporal logic.
- $(\mathbb{N}, \leq)$, etc. (The accessibility relation is $\leq$.)

- More generally, any dense/discrete linear order with/without endpoints.
- Trees, like $\{0,1\}^*$, or finite versions. Can be finitely or infinitely branching.
- Circular frames, like:



- Equivalence relations
- Cooked-up examples, such as Makinson's 'recession frame' $(\mathbb{N}, R)$, where $R(m,n)$ iff $n \geq m-1$. Need considerable creativity here.

Note (P. Halmos): 'iff' abbreviates 'if and only if'.

Kripke models

To give semantics to modal formulas in Kripke semantics, truth or falsity of basic facts (atoms) at worlds of a frame needs to be stated too.

So we add an *assignment* or *valuation:* a map

$$h : L \to \wp(W).$$

($\wp(W)$ is the set of all sets of worlds — the set of all subsets of W. It is called the power set of W.)

An atom p is said to be *true at a world* w iff $w \in h(p)$, and false, otherwise.

To model an application, choose h so that for each atom p, $h(p)$ is the set of worlds at which p is true/you want p to be true.

Definition 2.2 (Kripke model)

A triple (W, R, h) (W, R, h as above) is called a *(Kripke) model.*

Note: a single frame can be made into many different models (differing h).

Formal Kripke semantics

We can now define whether a formula is true or false at any given world of a Kripke model.

Definition 2.3 (Kripke semantics)

Let $\mathcal{M} = (W, R, h)$ be a Kripke model. We define $\boldsymbol{\mathcal{M}, t \models A}$, for a formula A and a world $t \in W$, by induction on A:

- For an atom p, we define $\mathcal{M}, t \models p$ iff $t \in h(p)$.
- $\mathcal{M}, t \models \top$ (always).
- $\mathcal{M}, t \models \neg A$ iff $\mathcal{M}, t \not\models A$.
- $\mathcal{M}, t \models A \wedge B$ iff $\quad \mathcal{M}, t \models A$ and $\mathcal{M}, t \models B$.
- $\mathcal{M}, t \models \Box A$ iff $\quad \mathcal{M}, u \models A$ for all $u \in W$ with $R(t, u)$.

Read $\mathcal{M}, t \models A$ as: *'A is true at world t in model $\mathcal{M}$'.*
Other common notations:
$\mathcal{M}, t \Vdash A$ (horrible!), $\quad \mathcal{M} \models_t A$, $\quad ||A||_t^{\mathcal{M}} = 1$, $\quad \ldots$

Abbreviations

Exercise 2.4

Check that our abbreviations 'work':

- $\mathcal{M}, t \models A \vee B$ *iff* $\mathcal{M}, t \models A$ *or* $\mathcal{M}, t \models B$ *(or both).*
- $\mathcal{M}, t \models A \to B$ *iff (if* $\mathcal{M}, t \models A$ *then* $\mathcal{M}, t \models B$*).*
- $\mathcal{M}, t \models A \leftrightarrow B$ *iff (*$\mathcal{M}, t \models A$ *iff* $\mathcal{M}, t \models B$*).*
- $\mathcal{M}, t \models \Diamond A$ *iff* $\mathcal{M}, u \models A$ *for some* $u \in W$ *with* $R(t, u)$*.*

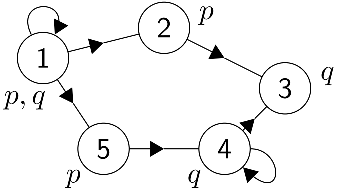
Also check the following (expected) equivalences:

- $\mathcal{M}, t \models \neg\Box A$ *iff* $\mathcal{M}, t \models \Diamond\neg A$*.*
- $\mathcal{M}, t \models \neg\Diamond A$ *iff* $\mathcal{M}, t \models \Box\neg A$*.*

Example model

Example 2.5

In the following model $\mathcal{M}$ (for atoms p, q only, true as shown):



- $\Box p$ is true at 1 and at 3, and false at 4.
- $\Box \neg p$ is also true at 3! So is $\Box \bot$!!
- $\Diamond q \wedge \Diamond \neg q$ is true at 1. $\Box q$ is false at 1.
- $\Diamond q, \Box q$ are both true at 2 (as only 3 is accessible from 2).
- $\Diamond \top \to \Diamond q$ is true at every world. We say it is *valid* in $\mathcal{M}$.

Semantics of temporal logic

Here, the notions of frame and model are exactly the same as before, but we use G and H instead of $\Box$.

$R(t,u)$ is taken to mean (is read as) 'u is in the future of t', or 't is in the past of u'.

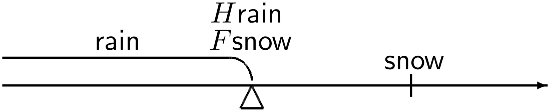
So R will usually be *transitive:* $R(t,u)$ and $R(u,v)$ imply $R(t,v)$.

(Exception: circular time.)

The clause for $\Box A$ above is replaced by the two clauses:

- $\mathcal{M}, t \models GA$ iff $\mathcal{M}, u \models A$ for all $u \in W$ with $R(t,u)$
  (G is just like $\Box$)
- $\mathcal{M}, t \models HA$ iff $\mathcal{M}, u \models A$ for all $u \in W$ with $R(u,t)$.

An example in linear time:



(Souped-up) Kripke semantics for PDL

Take W to be the set of all *states* (of a machine that executes $\mathcal{P}$-programs).

As each $\pi \in \mathcal{P}$ has its own box $[\pi]$, we need an accessibility relation R_π on W for each $\pi \in \mathcal{P}$.
So a frame has the form $(W, (R_\pi : \pi \in \mathcal{P}))$ (sometimes called a 'transition system').

Idea: $R_\pi(x, y)$ holds if in state x, the machine can execute π and terminate in state y.

So the different R_π ($\pi \in \mathcal{P}$) should be related according to the way π is constructed.

Relations between the R_π

We require that for all $\pi, \pi_1, \pi_2 \in \mathcal{P}$ and all $x, y \in W$:

- $R_{\pi_1 \cup \pi_2}(x, y)$ iff $R_{\pi_1}(x, y)$ or $R_{\pi_2}(x, y)$
- $R_{\pi_1 ; \pi_2}(x, y)$ iff for some $z \in W$, we have $R_{\pi_1}(x, z)$ and $R_{\pi_2}(z, y)$
- R_{π^*} is the *reflexive transitive closure* of R_π
- $R_{A?}(x, y)$ iff $x = y$ and $\mathcal{M}, x \models A$
  (model-dependent! Need an assignment first!)

3. Validity

Note: we do validity in Kripke semantics. Validity in other semantics is different.

Validity is an important idea in logic.
'Valid' is not the same as 'true'.
Valid = true *for all* $\langle$*something*$\rangle$.
E.g., all worlds/models/assignments/...
(We must always say what.)

The most extreme possibility is 'universal' validity:

Definition 3.1 (valid formula)

A modal formula A is said to be *valid* if $\mathcal{M}, t \models A$ for *every* model $\mathcal{M}$ and *every* world t of $\mathcal{M}$.

$$\boxed{\text{Valid} = \text{true at all worlds of all models} = \text{always true} = \text{unfalsifiable.}}$$

A valid formula is like a law of physics of Kripke's universe.
It's always true in Kripke semantics.

Examples of valid formulas

Instances of propositional tautologies:
e.g., $\Box p \vee \neg\Box p$, or $p \to (\Box q \to p)$.
Genuinely modal validities: e.g., $\Box\top$, and more coming up soon.

Which are valid?

- p (an atom)
- $\Diamond p$
- $\neg\Diamond p$
- $p \vee \neg p$
- $\Diamond p \vee \neg\Diamond p$
- $\Diamond p \vee \Diamond\neg p$

Related definitions: satisfiable, equivalent

Definition 3.2 (satisfiable formula)

A formula A is *satisfiable* if $\mathcal{M}, t \models A$ for some model $\mathcal{M}$ and some world t of $\mathcal{M}$.

Definition 3.3 (equivalent formulas)

Formulas A, B are *equivalent* if for every model $\mathcal{M}$ and every world t of $\mathcal{M}$, we have $\mathcal{M}, t \models A$ iff $\mathcal{M}, t \models B$.

Simple facts (check them!)

- A is valid iff $\neg A$ is not satisfiable, iff A is equivalent to $\top$.
- A is satisfiable iff $\neg A$ is not valid, iff A is not equivalent to $\bot$.
- A and B are equivalent iff $A \leftrightarrow B$ is valid.

Useful equivalences

For modal formulas A, B, write $A \equiv B$ if A, B are equivalent.

- Modal logic inherits propositional equivalences (tautologies). Eg, $A \to B \equiv \neg A \vee B$, for any *modal* formulas A, B.
- Replacing a subformula by an equivalent formula preserves $\equiv$: $\Box(A \to B) \equiv \Box(\neg A \vee B)$, $\Diamond\Box\neg\neg A \equiv \Diamond\Box A$, etc.
- $\neg\Box A \equiv \neg\Box\neg\neg A = \Diamond\neg A$
- $\neg\Diamond A = \neg\neg\Box\neg A \equiv \Box\neg A$
  Take the $\neg$ through, and swap $\Box, \Diamond$.
  Compare: $\neg\forall x A \equiv \exists x \neg A$.
- $\neg\Diamond\neg A = \neg\neg\Box\neg\neg A \equiv \Box A$
- $\neg\Box\neg A = \Diamond A$.

These equivalences can be verified by calculations in Kripke models.

Normality

So can validities. The ‘normality axiom’ $\Box(A \to B) \to (\Box A \to \Box B)$ is one of the most powerful validities. When combined with propositional tautologies, it allows all validities to be derived (details beyond scope of course).

Proposition 3.4

$\Box(A \to B) \to (\Box A \to \Box B)$ *is valid, for any modal formulas* A, B*.*

‘If in every possible world, $A \to B$ holds, then if also in every possible world A holds, then B must hold in every possible world.’ It doesn’t matter what the formulas A, B are.

For a formal proof of proposition 3.4 we must use the Kripke semantics.

Proof of proposition 3.4

Proof.

Take any world t of any model $\mathcal{M}$. We show that $\mathcal{M}, t \models \Box(A \to B) \to (\Box A \to \Box B)$.

So suppose that $\mathcal{M}, t \models \Box(A \to B)$; we have to show $\mathcal{M}, t \models \Box A \to \Box B$.

So suppose further that $\mathcal{M}, t \models \Box A$. We show that $\mathcal{M}, t \models \Box B$.

To do this, let u be an arbitrary world of $\mathcal{M}$ such that $R(t, u)$. We need to show that $\mathcal{M}, u \models B$.

Well, $\mathcal{M}, t \models \Box A$, so because $R(t, u)$, we have $\mathcal{M}, u \models A$. And $\mathcal{M}, t \models \Box(A \to B)$, so as $R(t, u)$, we have $\mathcal{M}, u \models A \to B$.

So we must have $\mathcal{M}, u \models B$, as required. □

Contingent validity

To capture formulas that are ‘contingent truths’, we work relative to a model, a frame, or a class (‘collection’) of frames.

Definition 3.5 (contingent validity)

A formula A is:

- *valid in a model $\mathcal{M}$* if it’s true at every world of $\mathcal{M}$.
- *valid in a frame $\mathcal{F}$* if it’s valid in every model based on $\mathcal{F}$.
- *valid in a class $\mathcal{C}$ of frames* if it’s valid in every frame in $\mathcal{C}$.

Exercise: show that A is valid (as per definition 3.1) iff it’s valid in the class of all frames (as per definition 3.5).

Can generalise ‘satisfiable’, ‘equivalent’ in the same way.

Example: reflexivity

The formula $\Box p \to p$ is an interesting example of a contingent validity.

Proposition 3.6

For an atom p, $\Box p \to p$ is valid in a frame $\mathcal{F} = (W, R)$ iff R is reflexive (that is, $R(t,t)$ holds for every world $t \in W$).

Conclude that *reflexive frames are good for modeling knowledge, and bad for belief.*
$\Box p \to p$ is a 'law of chemistry' of reflexive frames.

Proof. Suppose R is reflexive. We show that $\Box p \to p$ is valid in $\mathcal{F}$.
So let $\mathcal{M}$ be any model with frame $\mathcal{F}$, and let t be any world of $\mathcal{F}$.
We show that $\mathcal{M}, t \models \Box p \to p$.
If $\mathcal{M}, t \models \Box p$, then $\mathcal{M}, u \models p$ for all worlds $u \in W$ with $R(t,u)$.
But by reflexivity, $R(t,t)$. So $\mathcal{M}, t \models p$, as required.
As $\mathcal{M}, t$ were arbitrary, $\Box p \to p$ is valid in $\mathcal{F}$.

Proof of converse: $\Box p \to p$ valid $\Rightarrow R$ reflexive

Conversely, assume that $\Box p \to p$ is valid in $\mathcal{F}$.
Let $t \in W$ be arbitrary. We show that $R(t,t)$.
Let g be an assignment into $\mathcal{F}$ satisfying

$$g(p) = \{u \in W : R(t,u)\}.$$

(So p is true at just the worlds that are possible at t. We don't care about $g(q)$ for other atoms q.)
Let $\mathcal{M} = (\mathcal{F}, g)$.
As $\Box p \to p$ is valid in $\mathcal{F}$, it is true at t in the model $(\mathcal{F}, h)$ for *every* assignment h into $\mathcal{F}$.
Therefore, it is true at t in $(\mathcal{F}, g) = \mathcal{M}$. So $\mathcal{M}, t \models \Box p \to p$.
But by definition of g, we have $\mathcal{M}, t \models \Box p$.
We deduce that $\mathcal{M}, t \models p$. So $t \in g(p)$. So by choice of $g(p)$ we have $R(t,t)$.
(Our consideration of a *particular* assignment g does not damage the *conclusion* $R(t,t)$, which does not involve assignments.) □

We'll develop this trick later.

$\Box p \to p$ valid $\Rightarrow R$ reflexive: natural deduction proof

If you are not familiar with natural deduction, please ignore this slide. It is not examinable.

If you are, this may help. In line 5, we choose an assignment h satisfying $h(p) = \{z \in W : R(t,z)\}$.

1	$\Box p \to p$ valid in $\mathcal{F}$	given
2	$\forall x \forall h(\forall y(R(x,y) \to y \in h(p)) \to x \in h(p))$	what 1 means
3	t	$\forall I$ const
4	$\forall h(\forall y(R(t,y) \to y \in h(p)) \to t \in h(p))$	$\forall E(2)$
5	$\forall y(R(t,y) \to y \in \{z \in W : R(t,z)\}) \to t \in \{z \in W : R(t,z)\}$	$\forall E(4)$
6	$\forall y(R(t,y) \to R(t,y)) \to R(t,t)$	5 in other notation
7	d	$\forall I$ const
8	$R(t,d)$	assume
9	$R(t,d)$	✓(8)
10	$R(t,d) \to R(t,d)$	$\to I(8,9)$
11	$\forall y(R(t,y) \to R(t,y))$	$\forall I(7,10)$
12	$R(t,t)$	$\to E(11,6)$
13	$\forall x R(x,x)$	$\forall I(3,12)$

Conclusion

We asked why $\Box A \to A$ *could* always be true, depending on meaning of $\Box$ ('contingently true'), while $\Box(A \to B) \to (\Box A \to \Box B)$ is 'naturally' true.

Now we see: latter is valid ('law of physics'), while former is only valid over reflexive frames ('law of chemistry of reflexivity').
So, *accepting the Kripke semantics,* our answer is:

- 'naturally true' $=$ valid
- 'contingently true' $=$ valid in the class of all frames satisfying certain frame conditions (arising from the application)

We will want to find a frame condition to make a given modal formula valid (coming right up).
One can also determine which modal formulas are valid under varying frame conditions from different applications (beyond scope of course).

4. Frame conditions — Sahlqvist's theorem

We already saw (proposition 3.6) that $\Box q \to q$ is valid in a frame iff the frame's accessibility relation is reflexive.
Many other formulas also have associated frame conditions.
Eg., for a frame $\mathcal{F} = (W, R)$,

- $q \to \Box\Diamond q$ is valid in $\mathcal{F}$ iff R is symmetric.
- $\Box q \to \Box\Box q$ is valid in $\mathcal{F}$ iff R is transitive.
  (Cf. $\Diamond\Diamond p \to \Diamond p$ in tutorial 1.)
- $\Box q \to \Diamond q$ is valid in $\mathcal{F}$ iff $\mathcal{F}$ is serial — it satisfies $\forall t \exists u\, R(t, u)$.

Is there a general way of finding a frame condition associated with a given formula (if there is such a condition)?

More generally?

Probably the best way is by a technique of *Henrik Sahlqvist.*
Also involves work of van Benthem, Sambin & Vaccaro.
For this, we temporarily (until section 5) take $\top, \bot, \wedge, \vee, \neg, \Box, \Diamond$ all to be primitive (not abbreviations).
But $A \to B$ still abbreviates $\neg(A \wedge \neg B)$.

Definition 4.1 (Sahlqvist formula)

- A *boxed atom* is a formula of the form $\Box\Box\cdots\Box p$, for an atom p. *Note:* the chain of $\Box$s can be empty. So p, q, etc. are boxed atoms.
- A *positive* formula is one using no $\neg$.
- A *negative* formula is one of the form $\neg A$ where A is positive.
- An *untied formula* is one made from negative formulas and boxed atoms using only $\wedge$ and $\Diamond$. (Don't blame me for the name.)
- A *Sahlqvist formula* is a negated untied one.

Examples

Can often find one or more Sahlqvist formulas equivalent to a given formula. Recall (slide 10): $A \to B$ abbreviates $\neg(A \wedge \neg B)$.

- $\Box p \to p$ *is* a Sahlqvist formula: $\neg([\Box p] \wedge [\neg p])$.
- $\Box p \to \Box\Box p$ is the Sahlqvist formula $\neg([\Box p] \wedge [\neg\Box\Box p])$, and is equivalent to the Sahlqvist formula $\neg([\Box p] \wedge \Diamond\Diamond[\neg p])$.
- $\Diamond p \to \Box\Diamond p$ is the Sahlqvist formula $\neg(\Diamond[p] \wedge [\neg\Box\Diamond p])$, and is equivalent to the Sahlqvist formula $\neg(\Diamond[p] \wedge \Diamond[\neg\Diamond p])$.
- $\Diamond\Box p \to \Box\Diamond p$ is the Sahlqvist formula $\neg(\Diamond[\Box p] \wedge [\neg\Box\Diamond p])$ — and is equivalent to what?

But Löb's formula, $\Box(\Box p \to p) \to \Box p$, is not equivalent to any Sahlqvist formula (see Goldblatt's book, p51).
Nor is McKinsey's formula, $\Box\Diamond p \to \Diamond\Box p$ (see Goldblatt, p53).

Other definitions of Sahlqvist formula

There are many definitions of Sahlqvist formula in the literature.
Mostly, they are equivalent.
Sahlqvist himself defined his formulas differently: see Goldblatt p.51.
Our approach is a bit simpler, and effectively just as good: any 'real' Sahlqvist formula is equivalent to a conjunction ($\wedge$) of our Sahlqvist formulas.

You are *not* getting a watered-down version of his work.

Sahlqvist's theorem (1973, published 1975)

We will show that any Sahlqvist formula has an associated frame condition, which we can find with an algorithm. (Full statement in theorem 4.8.)

$\Diamond$ 'A real jewel of modal logic.' $\Diamond$

To prove this, we need two preliminaries.

Preliminary 1: Positive $\Rightarrow$ monotonic

First, we consider increasing (or decreasing) an assignment.

Definition 4.2 (ordering of assignments)

Let $\mathcal{F} = (W, R)$ be a frame and let $h, h' : L \to \wp(W)$ be assignments. We write $h \leq h'$ if $h(p) \subseteq h'(p)$ for all $p \in L$.

So h' is 'bigger' than h. Or, h is got by 'shrinking' h'.

Positive formulas preserved when assignment increases

Lemma 4.3

Let A be a positive formula, let $\mathcal{F} = (W, R)$ be a frame and let $h, h' : L \to \wp(W)$ be assignments with $h \leq h'$. Then for all $t \in W$,

if $(W, R, h), t \models A$ *then* $(W, R, h'), t \models A$*.*

This says 'A is monotonic'. A stays true if you 'increase' h.

Proof.

By induction on A. For an atom, p,
$(W, R, h), t \models p \;\Rightarrow\; t \in h(p) \;\Rightarrow\; t \in h'(p) \;\Rightarrow\; (W, R, h'), t \models p$.
If $A = \top$ or $\bot$, it is obvious. The cases $A \wedge B$ and $A \vee B$ are easy.
The case $\neg A$ does not arise (A is positive!).
Now we do the case $\Box A$ ($\Diamond A$ is similar — exercise).
Assume the result for A inductively.
Suppose $(W, R, h), t \models \Box A$. We want $(W, R, h'), t \models \Box A$, too.
Pick arbitrary $u \in W$ with $R(t, u)$. Then $(W, R, h), u \models A$.
By inductive hypothesis, $(W, R, h'), u \models A$.
So $(W, R, h'), t \models \Box A$, as required. □

Preliminary 2: The standard translation

We can translate modal formulas into formulas of first-order predicate logic. The translation reflects the Kripke semantics.

Definition 4.4 (standard translation)

For any first-order variable x, we'll translate A to a first-order formula A^x, with at most 1 free variable, x. (The translation depends on the choice of x.) A^x is defined by induction on A:

- An atom p translates to $P(x)$, where P is a unary relation symbol associated with p. So $p^x = P(x)$.
- $\top$ translates to itself: $\top^x = \top$. And $\bot^x = \bot$.
- $(\neg A)^x$ is $\neg(A^x)$.
- $(A \wedge B)^x$ is $A^x \wedge B^x$, and $(A \vee B)^x$ is $A^x \vee B^x$.
- $(\Box A)^x$ is $\forall y(R(x,y) \to A^y)$, where R is a binary relation symbol for the accessibility relation, and y is a new variable.
- $(\Diamond A)^x$ is $\exists y(R(x,y) \wedge A^y)$.

Examples of standard translation

Example 4.5 (doing standard translations)

- $$\begin{aligned} (\neg\Box\neg p)^x &= \neg(\Box\neg p)^x \\ &= \neg\forall y(R(x,y) \to (\neg p)^y) \\ &= \neg\forall y(R(x,y) \to \neg(p^y)) \\ &= \neg\forall y(R(x,y) \to \neg P(y)) \\ &\equiv \exists y\neg(R(x,y) \to \neg P(y)) \\ &\equiv \exists y(R(x,y) \wedge P(y)) \quad \text{— as expected.} \end{aligned}$$
- $(\Box p \to p)^x$ is $(\forall y(R(x,y) \to P(y))) \to P(x)$
- $(\Box(p \vee \Diamond q))^x$ is $\forall y(R(x,y) \to (P(y) \vee \exists z(R(y,z) \wedge Q(z))))$

Sometimes better to *partially* translate:
$(\Diamond A \to B)^x = \exists y(R(x,y) \wedge A^y) \to B^x$.
We haven't worked through the translations of A and B here.

Standard translation reflects the Kripke semantics.

Definition 4.6 (first-order structure ($\mathcal{M}\dagger$) from Kripke model)

Let $\mathcal{M} = (W, R, h)$ be a Kripke model. Define a first-order structure $\mathcal{M}\dagger$ for the signature $\{R\} \cup \{P : p \in L\}$, with domain W, by interpreting R as R, and P as $h(p)$ for each atom $p \in L$.

Lemma 4.7

Let $\mathcal{M} = (W, R, h)$ be any Kripke model. Then for any modal formula A and any $t \in W$,

$$\mathcal{M}, t \models A \iff \mathcal{M}\dagger \models A^x[t].$$

The second $\models$ is standard first-order evaluation with x assigned to t.

Proof.

Exercise — a simple induction on A. □

Sahlqvist's correspondence theorem (1975)

Theorem 4.8 (Sahlqvist correspondence)

For any Sahlqvist formula A, there is a corresponding first-order sentence in the language of Kripke frames (a 'frame property') that holds of a frame iff A is valid in the frame.
This sentence can be obtained from A by a simple algorithm.
It is called the Sahlqvist correspondent of A.

This theorem is a modal classic, and is extremely useful. Its proof method is probably the sharpest known for this kind of result. It generalises easily to multi-modal and temporal logics, and to some extent to the μ-calculus (see later for this topic).
Generalised by van Benthem, Goranko, Kikot, Vakarelov, . . .

Proof of Sahlqvist's correspondence theorem

We prove it by going through a (fairly general) example.
We will use approximately van Benthem's method.
Suppose A is the following 'generic' untied formula:

$$A \quad = \quad \Diamond\big(\Diamond[q] \wedge \Diamond([\neg C] \wedge \Diamond[\Box\Box p])\big) \ \wedge\ \Diamond\Diamond[\Box p],$$

where C is positive (e.g., $C = \Diamond p \vee \Box q$). A is built from the boxed atoms $\Box\Box p, \Box p, q$, and the negative $\neg C$ using only $\wedge, \Diamond$. So it's untied.
Let $\mathcal{F} = (W, R)$ be any frame.

Suppose the Sahlqvist formula $\neg A$ is *not* valid in $\mathcal{F}$. What does this mean?
It means that there is a model $\mathcal{M} = (\mathcal{F}, h)$ on frame $\mathcal{F}$, and a world $t \in W$, with $\mathcal{M}, t \models A$.
In the example, we get

$$\mathcal{M}, t \models \Diamond\big(\Diamond[q] \wedge \Diamond([\neg C] \wedge \Diamond[\Box\Box p])\big) \ \wedge\ \Diamond\Diamond[\Box p].$$

What does this say?

$$\mathcal{M}, t \models \Diamond(\Diamond[q] \wedge \Diamond([\neg C] \wedge \Diamond[\Box\Box p])) \;\wedge\; \Diamond\Diamond[\Box p]$$

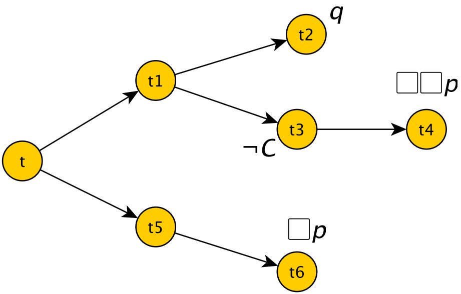
means that there's a 'pattern' of 6 worlds in $\mathcal{F}$ (for 6 $\Diamond$s), related to t by R in a certain way, and at each of them some specified formula(s) holds:

- either a boxed atom ($\Box p$, $\Box\Box p$, and q)
- or a negative one such as $\neg C$.

To understand the situation, use (i) a diagram, and (ii) the standard translation A^t, to get names for the worlds in the pattern. Use a different variable for each world in the pattern.
You only need translate the 'untied' part yet. Leave the boxed atoms and negative subformulas alone.

In our example, the standard translation *says* that there exists the following pattern of worlds $t_1, \ldots, t_6$, not necessarily distinct, with

$$\begin{array}{llll} R(t,t_1), & R(t_1,t_2), & \mathcal{M},t_2 \models q, & \\ & R(t_1,t_3), & \mathcal{M},t_3 \models \neg C, & \\ & & R(t_3,t_4), & \mathcal{M},t_4 \models \Box\Box p, \\ R(t,t_5), & R(t_5,t_6), & \mathcal{M},t_6 \models \Box p. & \end{array}$$



Critical step: shrinking the assignment

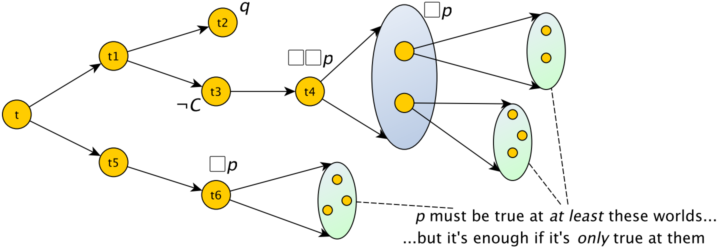
Now by lemma 4.3, if we 'shrink' h (make the atoms true at fewer worlds) then the negative formulas (here, just $\neg C$) all stay true at their own worlds in the pattern.
Eg, $\neg(\Diamond p \vee \Box q)$ would stay true if p, q shrink.

So what's the *smallest* assignment (with respect to $\leq$) that we could get away with, that still keeps all the formulas above true at their worlds (and so keeps A true at t)?

Answer: it is the smallest assignment that keeps all the boxed atoms ($\Box\Box\cdots\Box p$ etc) true at their worlds.
Boxed atoms have the property that there is indeed such a smallest assignment — as we'll now see.

What is this ‘smallest’ assignment?

$\Box\Box p$ being true at world t_4 says that ‘every world accessible from t_4 in two R-steps satisfies p.’
So $\Box\Box p$ will continue to hold at t_4 even if p is shrunk right down to just those worlds x accessible from t_4 in 2 steps: those x satisfying $\exists y(R(t_4, y) \wedge R(y, x))$.



p must be true at *at least* these worlds...
...but it's enough if it's *only* true at them

We have to keep $\Box p$ true at t_6 too. So (same idea) make p true *additionally* at all worlds x such that $R(t_6, x)$.
And we have to keep q true at t_2.
So (same idea) make q true at all worlds x satisfying $x = t_2$ (!)

Lazy assignment

So if we make p, q true at *only* the worlds where we *have to* in order to keep the boxed atoms true at their worlds in the pattern, then the original untied formula A remains true at t.
Call this 'lazy' assignment h°.

This new h° is definable from the pattern by first-order formulas:

- we made p true at world x iff
  $\exists y(R(t_4, y) \wedge R(y, x)) \vee R(t_6, x)$ holds in the frame,
- we made q true at x iff $x = t_2$ holds in the frame.

These are *first-order* conditions, with *no atoms.*

So for this lazy assignment h°, we can simplify the standard translation A^t of A:
We find the (standard translations of the) atoms inside A^t, and replace them by these 'lazy' first-order expressions for them.

Simplifying the translation A^t

- Compute standard translation A^t of A. Don't bother to translate boxed atoms (but you have to do the negative formulas).
- Move all $\exists$s *for named worlds* $t_1, \ldots, t_6$ to the front. (Change the names of variables if necessary to preserve equivalence.)
- Replace every $P(x)$ by our lazy expression for it: here, it is
  $$\exists y(R(t_4, y) \wedge R(y, x)) \vee R(t_6, x).$$
- Similarly, replace $P(y)$ by $\exists z(R(t_4, z) \wedge R(z, y)) \vee R(t_6, y)$ (I changed $\exists y$ to $\exists z$ to avoid clash of bound variables).
- Similarly for any subformulas $P(z)$, $P(t)$, $P(t_4)$, etc.
- Replace any $Q(x)$ by $x = t_2$, $Q(t_5)$ by $t_5 = t_2$, etc.
- May as well substitute $\top$ for (the translations of) the boxed atoms $\Box\Box\cdots\Box p$ (we chose h° precisely to make these all true, so this step is legitimate, and it shortens the formula).

Call the result $\alpha(t)$.

Finishing off

Now, the following are equivalent:

- $\neg A$ is not valid in $\mathcal{F}$.
- $(\mathcal{F}, h), t \models A$ for some assignment h and some $t \in W$.
- $(\mathcal{F}, h^\circ), t \models A$ for some $t \in W$ and 'pattern' $t_1, \ldots, t_6$ (of course, h° depends on these).
- $\mathcal{F} \models \alpha[t]$ for some $t \in W$ ("$\models$" as in classical logic).

Therefore, *$\neg A$ is valid in $\mathcal{F}$ iff $\mathcal{F} \models \forall t \neg \alpha(t)$.*

So the original Sahlqvist formula $\neg A$ 'corresponds' to the first-order frame condition $\forall t \neg \alpha(t)$.

Note that the construction of α is algorithmic.

Summary of Sahlqvist's algorithm

It finds a frame condition equivalent to the validity of a Sahlqvist formula. (Can often find a Sahlqvist equivalent of a given formula, but you have to hack this bit.)

Eg: frame condition corresponding to validity of

$\Box p \to p \equiv \neg([\Box p] \wedge [\neg p])$ is reflexivity, $\forall x R(x, x)$.

Given: a Sahlqvist formula $\neg A$, where A is untied.

To find frame condition corresponding to $\neg A$:

1. Identify negative formulas and boxed atoms of A.
2. Draw a diagram showing what it means for A to be true at a world t. Include names for worlds (eg. t_1, t_2, etc), R-relations between them, and which boxed atoms and negative formulas are true at which worlds.
3. Work out the lazy assignment (that just makes the boxed atoms true at their worlds). Get a first-order expression for it, in terms of the named worlds.

4. Work out the standard translation A^t of A. Use the names of the named worlds for variables. Don't bother to translate boxed atoms. Then manipulate it as follows:
5. Move all ∃s for named worlds (those ∃s from ♢s outside the negative formulas) to the front of A^t.
   DO THIS NOW! Easily forgotten, but VITAL.
6. Find the bits that come from the boxed atoms. Replace them with $\top$.
7. Replace all remaining subformulas $P(x), Q(y)$, etc. (they come from atoms in negative formulas) by the corresponding first-order expressions got from the lazy assignment.
8. Simplify, if possible.

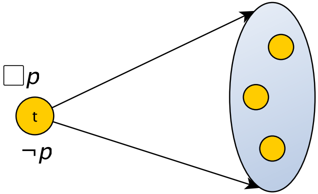
What you get is a first-order formula $\alpha(t)$ expressing that A is true in $\mathcal{F}$ at some world t under some assignment.
So $\neg A$ is valid in $\mathcal{F}$ iff $\forall t \neg\alpha(t)$ holds in $\mathcal{F}$.
This $\forall t \neg\alpha(t)$ is the frame condition you want.

Example 1: reflexivity: $\Box p \to p$

The algorithm is messy to explain, but it's easy to use it on particular examples.
I suggest you use diagrams.

- Try $\Box p \to p$.

This is the Sahlqvist formula $\neg([\Box p] \wedge [\neg p])$. It isn't valid in a frame $\mathcal{F}$ iff there are h, t with $(\mathcal{F}, h), t \models [\Box p] \wedge [\neg p]$.
The 'pattern' here is just t (as no $\Diamond$s around).
For $\Box p$ to be true at t, the 'lazy' assignment makes p true at a world x just when $R(t, x)$.



Use this as a definition of p. Take the standard translation $(\Box p)^t \wedge \neg P(t)$ of $\Box p \wedge \neg p$. Then

- replace the $\Box p$-part by $\top$ (the lazy assignment always makes this true)
- in the negative part $\neg P(t)$, replace $P(t)$ by its lazy definition $R(t, x)$ (with $x = t$).

We get $\top \wedge \neg R(t, t)$. Simplify this to the equivalent $\neg R(t, t)$.
So $\Box p \wedge \neg p$ holds at t under some assignment iff $\mathcal{F} \models \neg R(t, t)$.
So $\Box p \to p$ is valid in $\mathcal{F}$ iff $\mathcal{F} \models \forall t \neg\neg R(t, t)$.
That is,

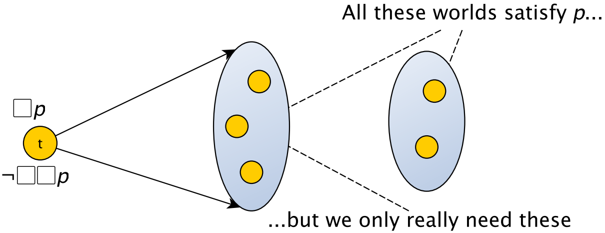
$$\mathcal{F} \models \forall t\, R(t, t).$$

Reflexivity! □
We knew this. But Sahlqvist's algorithm is far more general.

Example 2: transitivity: $\Box p \to \Box\Box p$

- Let's try Sahlqvist's algorithm on $\Box p \to \Box\Box p$.

If this is *not* valid in frame $\mathcal{F}$, then there are h, t making $[\Box p] \wedge [\neg\Box\Box p]$ true at t.
The pattern here is again just t, as no diamonds around.


All these worlds satisfy *p*...
□*p*
t
¬□□*p*
...but we only really need these

To make $\Box p$ true at t, we need only that p is true at any x with $R(t, x)$.
So the lazy assignment makes p true at x iff $R(t, x)$ holds.
Use $R(t, x)$ as the definition of p in the standard translation $([\Box p] \wedge [\neg\Box\Box p])^t$.

The standard translation $([\Box p] \wedge [\neg\Box\Box p])^t$ is $(\Box p)^t \wedge \neg\forall u(R(t,u) \to \forall v(R(u,v) \to P(v)))$.

- May as well replace the $\Box p$ part by $\top$ (the lazy assignment forces it to be true)
- Replace $P(v)$ by the lazy definition of P at v: namely, $R(t,v)$.

We get $\top \wedge \neg\forall u(R(t,u) \to \forall v(R(u,v) \to R(t,v)))$.
It simplifies to $\neg\forall u(R(t,u) \to \forall v(R(u,v) \to R(t,v)))$.
Call this $\alpha(t)$.
So $\Box p \wedge \neg\Box\Box p$ is true at t under some assignment iff $\mathcal{F} \models \alpha(t)$.
So $\Box p \to \Box\Box p$ is valid in $\mathcal{F}$ iff $\mathcal{F} \models \forall t\neg\alpha(t)$. That is,

$$\mathcal{F} \models \forall t\neg\neg\forall u(R(t,u) \to \forall v(R(u,v) \to R(t,v))).$$

Using first-order equivalences, we can clean this up to:

$$\mathcal{F} \models \forall tuv(R(t,u) \wedge R(u,v) \to R(t,v)).$$

This just says that $\mathcal{F}$ is transitive! □

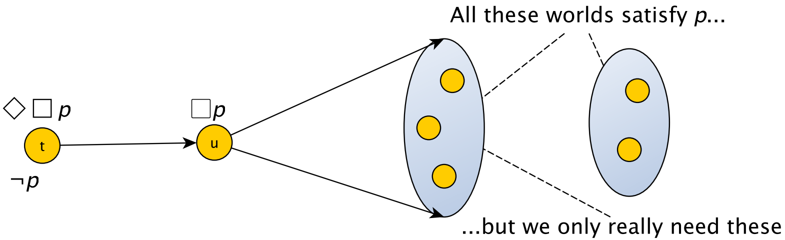
Alternative Sahlqvist formula for transitivity

We can write $\Box p \to \Box\Box p$ as $\neg([\Box p] \land \Diamond\Diamond[\neg p])$ too.
It's a Sahlqvist formula — try the algorithm on it. You should get the same answer as above.

Example 3: symmetry (variant): $\Diamond\Box p \to p$

- Let's try $\Diamond\Box p \to p$.

It's a Sahlqvist formula: it is $\neg(\Diamond[\Box p] \wedge [\neg p])$.
Let $A = \Diamond[\Box p] \wedge [\neg p]$ (untied).
A is true at t in a model whose frame is $\mathcal{F}$ iff the following pattern of worlds t, u exists:



All these worlds satisfy *p*...

...but we only really need these

The lazy assignment makes p true at a world x iff $R(u, x)$.

The standard translation A^t of $A = \Diamond[\Box p] \wedge [\neg p]$ is

$$\exists u(R(t,u) \wedge (\Box p)^u) \wedge \neg P(t).$$

u is a named world in the diagram, so *move $\exists u$ to the front* (preserving logical equivalence): $\exists u(R(t,u) \wedge (\Box p)^u \wedge \neg P(t))$.
Do this now — before the steps below.

Now replace $P(t)$ by the lazy $R(u,t)$, and $(\Box p)^u$ by $\top$:

$$\exists u(R(t,u) \wedge \top \wedge \neg R(u,t)).$$

A is true at t under some assignment iff this holds.

So $\neg A$ is valid in $\mathcal{F}$ iff
$\mathcal{F} \models \forall t \neg \exists u(R(t,u) \wedge \top \wedge \neg R(u,t))$ — that is,

$$\mathcal{F} \models \forall tu(R(t,u) \to R(u,t)).$$

This says R is symmetric! □

What if we don't move $\exists u$ to the front?

We'd get $\neg \exists t(\exists u(R(t,u) \wedge \top) \wedge \neg R(u,t))$, which has u free!
This is not a sentence (because it has free variables).
So as a frame condition, it makes no sense whatever.

5. p-morphisms and bisimulations

These are two (related) kinds of similarity between Kripke models/frames. Can generalise to multi-modal, temporal logic.

- p-morphisms are certain maps (or functions) that preserve (validity of) modal formulas. We can use them to show that certain frame properties are *not* modally expressible.
- Bisimulations are more general, using relations instead of maps.

  Bisimulations also preserve modal formulas — in a sense, modal formulas are *precisely* the ones so preserved. This gives us a measure of the strength of modal logic.

  Bisimulations connect modal logic with concurrency.

Warning: from now on, $\vee, \Diamond, \bot$ are *abbreviations* again.

5.1 Frame p-morphisms

Let $\mathcal{F} = (W, R)$, $\mathcal{F}' = (W', R')$ be Kripke frames and $f : W \to W'$ a map.

Definition 5.1

We say f is a *frame p-morphism* from $\mathcal{F}$ to $\mathcal{F}'$ if:

1. **forth:** If $t, u \in W$ and $R(t, u)$, then $R'(f(t), f(u))$.
2. **back:** If $t \in W$, $x \in W'$, and $R'(f(t), x)$, then there is some $u \in W$ with $R(t, u)$ and $f(u) = x$.

The name apparently comes from ‘pseudo-epimorphism’.
Some people write ‘bounded morphism’ instead.
Goldblatt is a good reference for p-morphisms.

Alternative definition of frame p-morphism

Here's an equivalent and slightly simpler definition of frame p-morphism.
Recall that for any sets X, Y, a map $f : X \to Y$ is *onto*, or *surjective*, if for every $y \in Y$ there is some $x \in X$ such that $f(x) = y$.

Proposition 5.2

$f : W \to W'$ is a frame p-morphism iff:

(pM): *for each world t of $\mathcal{F}$, the map f maps the worlds accessible from t in $\mathcal{F}$ onto the worlds accessible from $f(t)$ in $\mathcal{F}'$.*

That is, for all $t \in W$,

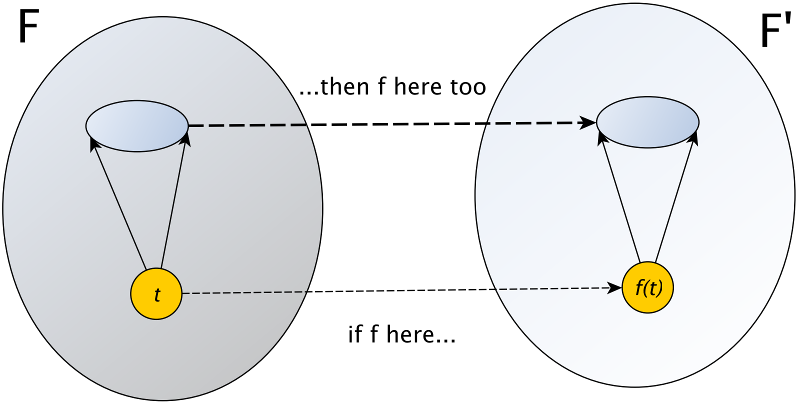
$$\{f(u) : u \in W,\ R(t, u)\} = \{v \in W' : R'(f(t), v)\}.$$

Proof.

Exercise. □

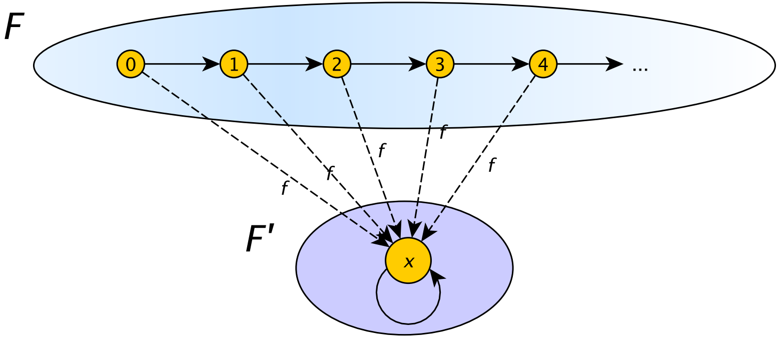
Illustration of property (pM)

Pictures made by yed



Example of frame p-morphism

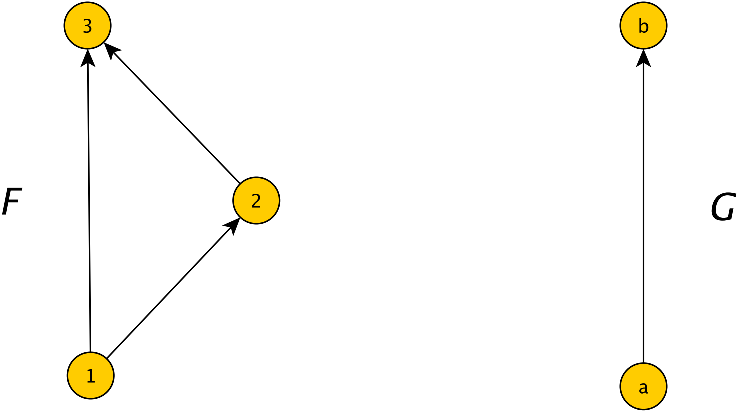
Let $\mathcal{F} = (\mathbb{N}, <)$.
Let $\mathcal{F}'$ be a one-world reflexive frame ($W' = \{x\}$, say, with $R'(x, x)$).
Here's a frame p-morphism from $\mathcal{F}$ to $\mathcal{F}'$:



Exercise

Exercise 5.3

Are there any p-morphisms between these two frames?



p-morphic images

Definition 5.4

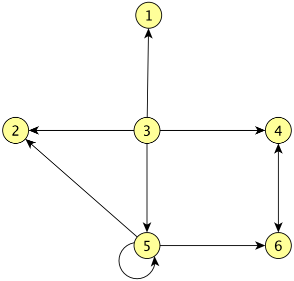
Let $\mathcal{F}, \mathcal{F}'$ be Kripke frames.
We say that $\mathcal{F}'$ is a *p-morphic image* of $\mathcal{F}$ if there exists a frame p-morphism from $\mathcal{F}$ *onto* $\mathcal{F}'$.

Example 5.5

The 1-world frame $\mathcal{F}'$ (slide 76) is a p-morphic image of $(\mathbb{N}, <)$.

Exercise 5.6

Find a frame with at most 3 worlds that's a p-morphic image of:



Remark: taking p-morphic images preserves symmetry

Many frame properties are preserved by p-morphic images.
Eg: a frame is *symmetric* if it satisfies $\forall xy(R(x,y) \to R(y,x))$.

Proposition 5.7

Any p-morphic image of a symmetric frame is also symmetric.

Proof.

Let $\mathcal{F} = (W, R)$ be symmetric. Let $\mathcal{F}' = (W', R')$ be any frame.
Suppose that $f : \mathcal{F} \to \mathcal{F}'$ is a surjective (onto) frame p-morphism.
To show $\mathcal{F}'$ is symmetric, let $x', y' \in W'$ be arbitrary.
Suppose that $R'(x', y')$ holds. We need to show that $R'(y', x')$.
As f is onto, there is $x \in W$ with $f(x) = x'$.
By the **back** property of frame p-morphisms, there is $y \in W$ with $R(x, y)$ and $f(y) = y'$.
As $\mathcal{F}$ is symmetric, we deduce that $R(y, x)$.
By the **forth** property of frame p-morphisms, we have $R'(f(y), f(x))$ — that is, $R'(y', x')$. As required. □

5.2 Model p-morphisms

These are souped-up frame p-morphisms, defined on Kripke *models.*
Let $\mathcal{F}, \mathcal{F}'$ be Kripke frames.
Let $\mathcal{M} = (\mathcal{F}, h)$ and $\mathcal{M}' = (\mathcal{F}', h')$ be Kripke models.

Definition 5.8

We say that f is a *(model) p-morphism* from $\mathcal{M}$ to $\mathcal{M}'$ if:

1. f is a frame p-morphism : $\mathcal{F} \to \mathcal{F}'$.
2. If $t \in W$, then for each atom p we have
   $\mathcal{M}, t \models p$ iff $\mathcal{M}', f(t) \models p$.

'Atoms are preserved'

Example of model p-morphism

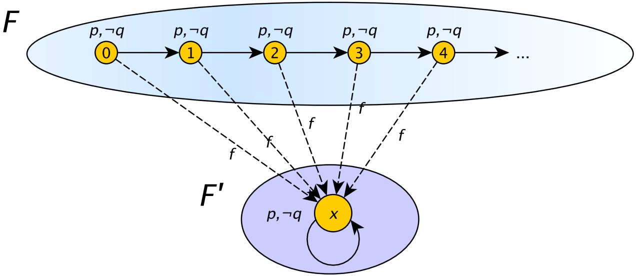
Let $\mathcal{M}'$ be any model over the one-world reflexive frame $\mathcal{F}'$ on slide 76.

Let $\mathcal{M}$ be the model $(\mathbb{N}, <, h)$ defined by: for each atom $p \in L$,

$$h(p) = \begin{cases} \mathbb{N}, & \text{if } \mathcal{M}', x \models p \text{ (where } x \text{ is the sole world of } \mathcal{F}'\text{)}, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Then the map f of slide 76 is a model p-morphism : $\mathcal{M} \to \mathcal{M}'$.

Example when $\mathcal{M}', x \models p$ and $\mathcal{M}', x \not\models q$:



Model p-morphisms preserve truth of modal formulas

Theorem 5.9

Let $\mathcal{M} = (W, R, h)$ and $\mathcal{M}' = (W', R', h')$ be models, and let f be a model p-morphism from $\mathcal{M}$ to $\mathcal{M}'$. [f does not have to be onto.] Then for any world t of $\mathcal{M}$ and any modal formula A,

$$\mathcal{M}, t \models A \textit{ iff } \mathcal{M}', f(t) \models A.$$

Proof.

By induction on A. The theorem holds for atomic A because p-morphisms preserve atoms (definition 5.8(2)).

The boolean cases $(\top, \wedge, \neg)$ are easy.

The main case is $\Box A$. Assume (inductively) the result for A.

Then $\mathcal{M}, t \models \Box A$

iff $\mathcal{M}, u \models A$ for all $u \in W$ with $R(t, u)$ (by semantics of $\Box$),

iff $\mathcal{M}', f(u) \models A$ for all $u \in W$ with $R(t, u)$ (by ind. hyp.),

iff $\mathcal{M}', x \models A$ for all $x \in W'$ with $R'(f(t), x)$ (because by property (pM), $\{f(u) : u \in W, R(t, u)\} = \{x \in W' : R'(f(t), x)\}$),

iff $\mathcal{M}', f(t) \models \Box A$ (by semantics of $\Box$). □

p-morphic images 'preserve' validity

— important consequence of theorem 5.9

Theorem 5.10

Let $\mathcal{F} = (W, R)$ and $\mathcal{F}' = (W', R')$ be Kripke frames.
Suppose that $\mathcal{F}'$ is a p-morphic image of $\mathcal{F}$.
Then any modal formula that's valid in $\mathcal{F}$ is valid in $\mathcal{F}'$.

Proof.

Let A be a formula valid in $\mathcal{F}$. We show it's valid in $\mathcal{F}'$.
So let $h' : L \to \wp(W')$ be an arbitrary assignment, and take any $t' \in W'$. We show $(\mathcal{F}', h'), t' \models A$.
Let $f : W \to W'$ be a *surjective* frame p-morphism from $\mathcal{F}$ to $\mathcal{F}'$.
Define $h : L \to \wp(W)$ by

$$h(p) = \{w \in W : f(w) \in h'(p)\} \quad \text{for each } p \in L.$$

Then f is a model p-morphism from $(\mathcal{F}, h)$ to $(\mathcal{F}', h')$.
Pick $t \in W$ such that $f(t) = t'$ (use 'f onto').
We assumed that A is valid in $\mathcal{F}$. So $(\mathcal{F}, h), t \models A$.
By theorem 5.9, $(\mathcal{F}', h'), t' \models A$. □

Consequences

Theorem 5.10 says that

Taking p-morphic images preserves modal validity (forwards).

It is very important. It has many consequences.
We can use it to show that

1. various frame properties are preserved by taking p-morphic images,
2. various frame properties, such as irreflexivity, are not definable by the validity of any modal formula,
3. any modal formula valid in all irreflexive frames is valid (in all frames).

Modally definable frame properties and p-morphic images

Definition 5.11

Let $\mathcal{P}$ be a property of frames (e.g, reflexivity, $\forall x R(x,x)$). We say that $\mathcal{P}$ is *modally definable* if there is a modal formula A such that for any frame $\mathcal{F}$, $\mathcal{F}$ has property $\mathcal{P}$ iff A is *valid* in $\mathcal{F}$.

Example: reflexivity is modally definable (by $\Box p \to p$).

Proposition 5.12

Any modally definable frame property $\mathcal{P}$ is preserved (forwards) by taking p-morphic images.

That is, if $\mathcal{F}$ is a frame with property $\mathcal{P}$, and $\mathcal{F}'$ is a p-morphic image of $\mathcal{F}$, then $\mathcal{F}'$ has property $\mathcal{P}$ as well.

Proof. As $\mathcal{P}$ is modally definable, there is a modal formula A that is valid in precisely the frames having property $\mathcal{P}$. Then

$\mathcal{F}$ has $\mathcal{P}$ iff A is valid in $\mathcal{F}$ $\underset{\text{thm. 5.10}}{\Rightarrow}$ A is valid in $\mathcal{F}'$ iff $\mathcal{F}'$ has $\mathcal{P}$. $\square$

p-morphic images preserve Sahlqvist correspondents

Corollary 5.13

The correspondent of any Sahlqvist formula is preserved (forwards) by taking p-morphic images.

Proof.

By proposition 5.12, because the correspondent of a Sahlqvist formula A is modally definable (by A). □

Example 5.14

$\forall xy(R(x,y) \to R(y,x))$ (symmetry) is preserved by p-morphic images, as it is the correspondent of $\Diamond\Box p \to p$ (slide 70).
This gives an easier proof of proposition 5.7.

But for non-modally-definable frame properties, this method fails and you need the direct approach of proposition 5.7 to show preservation by p-morphic images.

No modal formula characterises irreflexivity

Recall $\Box p \to p$ is valid in precisely the reflexive frames. But. . .

Proposition 5.15

There's no modal formula that's valid in precisely the irreflexive frames (satisfying $(\forall x \neg R(x,x))$).
That is, irreflexivity is not modally definable.

Proof.

By proposition 5.12, it is enough to show that irreflexivity is not always preserved when we move to a p-morphic image.
To show this, let $\mathcal{F}$ be the irreflexive frame $(\mathbb{N}, <)$ on slide 76.
As we saw, the one-point reflexive frame $\mathcal{F}'$ is a p-morphic image of $\mathcal{F}$.
But $\mathcal{F}'$ is not irreflexive. □

Exercise 5.16

Show that the frame property 'for every world w of $\mathcal{F}$, at least two worlds are accessible from w' is not modally definable.

5.3 Bulldozing

We can strengthen proposition 5.15 to show that any modal formula that's valid in all irreflexive frames is valid (in *all* frames). So 'irreflexive frames have no special laws of chemistry'.

We will use *bulldozing* (a trick of Segerberg).
We bulldoze a frame by replacing groups of mutually related worlds by larger groups of worlds strung out in a line.
E.g., we can bulldoze the 1-point reflexive frame (above) to get $(\mathbb{N}, <)$.

It's a rather flexible method. It can be used in many ways: 'frame surgery'.
But in simple cases, we can get the same effects by taking the *product* of the frame with another frame, such as $(\mathbb{N}, <)$.

Products of frames; irreflexivity

Recall: if X, Y are sets, then $X \times Y = \{(x, y) : x \in X,\ y \in Y\}$.

Definition 5.17

Let $\mathcal{F}_1 = (W_1, R_1)$ and $\mathcal{F}_2 = (W_2, R_2)$ be frames.
The *product* $\mathcal{F}_1 \times \mathcal{F}_2$ is the frame $(W_1 \times W_2,\ R^\times)$, where, for any $(x_1, x_2), (y_1, y_2) \in W_1 \times W_2$, we define

$$R^\times((x_1, x_2),\ (y_1, y_2)) \text{ iff } R_1(x_1, y_1) \text{ and } R_2(x_2, y_2).$$

Lemma 5.18

Let $\mathcal{F}_1 = (W_1, R_1)$ and $\mathcal{F}_2 = (W_2, R_2)$ be frames. Suppose that at least one of them is irreflexive. Then $\mathcal{F}_1 \times \mathcal{F}_2$ is irreflexive.

Proof.

Assume that $\mathcal{F}_1$ is irreflexive. Pick any $(w_1, w_2) \in W_1 \times W_2$. Then $\neg R_1(w_1, w_1)$. So by definition 5.17, $\neg R^\times((w_1, w_2), (w_1, w_2))$.
Hence $\mathcal{F}_1 \times \mathcal{F}_2$ is irreflexive.
The other case (when $\mathcal{F}_2$ is irreflexive) is left as an exercise. □

Products and p-morphisms

Lemma 5.19

Let $\mathcal{F}_1 = (W_1, R_1)$ be any frame, and let $\mathcal{F}_2 = (W_2, R_2)$ be any serial frame (slide 43). The 'projection' map $\pi : \mathcal{F}_1 \times \mathcal{F}_2 \to \mathcal{F}_1$ given by $\pi(x_1, x_2) = x_1$ is a surjective p-morphism.

Proof.

1. **Forth:** Take any $(x_1, x_2), (y_1, y_2) \in W_1 \times W_2$. Assume that $R^\times((x_1, x_2), (y_1, y_2))$. Then $R_1(x_1, y_1)$ (and $R_2(x_2, y_2)$).
   But $\pi(x_1, x_2) = x_1$, and $\pi(y_1, y_2) = y_1$.
   So $R_1(\pi(x_1, x_2), \pi(y_1, y_2))$ holds.
2. **Back:** Take any $(x_1, x_2) \in W_1 \times W_2$ and $y_1 \in W_1$, and assume that $R_1(\pi(x_1, x_2), y_1)$. That is, $R_1(x_1, y_1)$.
   As $\mathcal{F}_2$ is serial, we can choose $y_2 \in W_2$ with $R_2(x_2, y_2)$.
   Then $(y_1, y_2) \in W_1 \times W_2$, $R^\times((x_1, x_2), (y_1, y_2))$, and $\pi(y_1, y_2) = y_1$.
3. π is surjective, since taking any $x_2 \in W_2$, we have $\pi(x_1, x_2) = x_1$ for every $x_1 \in W_1$.

So $\pi : \mathcal{F}_1 \times \mathcal{F}_2 \to \mathcal{F}_1$ is a surjective p-morphism, as required. □

Bulldozing (products) and irreflexivity

Theorem 5.20

Let A be any modal formula. Then A is valid in all irreflexive frames iff A is valid (in all frames).

Proof.

$\Leftarrow$ is trivial. We prove $\Rightarrow$.
Let $\mathcal{F}$ be any frame. We show A is valid in $\mathcal{F}$.
By lemma 5.18, $\mathcal{F} \times (\mathbb{N}, <)$ is irreflexive. So A is valid in it.
But $(\mathbb{N}, <)$ is serial. So by lemma 5.19, $\mathcal{F}$ is a p-morphic image of $\mathcal{F} \times (\mathbb{N}, <)$.
By theorem 5.10, p-morphic images preserve modal validity.
So A is valid in $\mathcal{F}$. □

So any satisfiable formula can be satisfied in (made true at a world of a model on) an *irreflexive* frame.
We conclude that the 'modal logic of irreflexive frames' is the same as the 'modal logic of all frames' (same validities, same 'laws of chemistry').

5.4 Bisimulations — generalised p-morphisms

Definition 5.21

Let $\mathcal{M} = (W, R, h)$ and $\mathcal{M}' = (W', R', h')$ be Kripke models. Let $t \in W$, $t' \in W'$. A *bisimulation* between $(\mathcal{M}, t)$ and $(\mathcal{M}', t')$ is a relation $B \subseteq W \times W'$ satisfying:

1. $B(t, t')$,

and for every $u \in W$ and $u' \in W'$ such that $B(u, u')$:

2. For all atoms p: $\mathcal{M}, u \models p$ iff $\mathcal{M}', u' \models p$.
3. **forth:** If $v \in W$ and $R(u, v)$, then there is $v' \in W'$ with $R'(u', v')$ and $B(v, v')$.
4. **back:** If $v' \in W'$ and $R'(u', v')$, then there is $v \in W$ with $R(u, v)$ and $B(v, v')$.

Definition 5.22

We say $(\mathcal{M}, t)$ and $(\mathcal{M}', t')$ are *bisimilar* if there exists a bisimulation between $\mathcal{M}, t$ and $\mathcal{M}', t'$.

Bisimulation invariance of modal formulas

Theorem 5.23

Let $(\mathcal{M}, t)$ and $(\mathcal{M}', t')$ be bisimilar and let A be any modal formula. Then $\mathcal{M}, t \models A$ iff $\mathcal{M}', t' \models A$.

Proof.

Induction on A (like theorem 5.9). □

So modal formulas are *invariant under bisimulation.*

It turns out that (slogan):

> *Modal logic 'is' the bisimulation-invariant fragment of first-order logic!*

We'll end with a quick look at this.

Bisimulation invariance of first-order formulas?

Definition 5.24

Let $\alpha(x, P_1, \ldots, P_n)$ be a first-order formula written with the relation symbol R, equality, and unary relation symbols $P_1, \ldots, P_n$ corresponding to atoms $p_1, \ldots, p_n \in L$.
We say that α is *bisimulation-invariant* if whenever $(\mathcal{M}, t)$ and $(\mathcal{M}', t')$ are bisimilar then $\mathcal{M}\dagger \models \alpha(t)$ iff $\mathcal{M}'\dagger \models \alpha(t')$.

See definition 4.6 for $\mathcal{M}\dagger$. (Each P_k $(k = 1, \ldots, n)$ is interpreted in $\mathcal{M}\dagger$ in the same way as p_k in $\mathcal{M}$.)

Recall that every modal formula A has a *standard translation* A^x into first-order logic. A^x has the same 'meaning' as A (lemma 4.7). So by theorem 5.23, A^x is always bisimulation-invariant. Therefore:

Lemma 5.25

Any first-order formula $\alpha(x, P_1, \ldots, P_n)$ *that is equivalent to (the standard translation of) a modal formula is bisimulation-invariant.*

Is there a converse to this? It would be very surprising/interesting.

Characterisation of the strength of modal logic

Theorem 5.26 (van Benthem, 1976)

A first-order formula $\alpha(x, P_1, \ldots, P_n)$ is logically equivalent to the standard translation of a modal formula iff α is bisimulation-invariant.

Proof. '$\Rightarrow$' is lemma 5.25.

Idea of $\Leftarrow$ [not examinable]: if $\alpha(x, P_1, \ldots, P_n)$ is *not* equivalent to A^x for any modal formula A, van Benthem used mathematical logic to find models $\mathcal{M} = (W, R, h)$, $\mathcal{M}' = (W', R', h')$ and worlds $w \in W$, $w' \in W'$, such that

- $\{(x, x') \in W \times W' : \mathcal{M}, x \models A \iff \mathcal{M}', x' \models A$, for every modal fmla $A\}$ is a bisimulation between $(\mathcal{M}, w)$ & $(\mathcal{M}', w')$
- $\mathcal{M}\dagger \models \alpha[w]$ but $\mathcal{M}'\dagger \models \neg\alpha[w']$.

So α is not bisimulation-invariant. □

So modal logic 'is' the bisimulation-invariant fragment of first-order logic. But whether a first-order formula is bisimulation-invariant is undecidable!

Bisimulations and concurrency

Bisimulations have been used elsewhere in computer science.

They were introduced in concurrency by Park (1981).

In concurrency theory, two processes are defined to be bisimilar in a similar way to definition 5.21. It 'means' they can't reasonably be distinguished.

Hennessy and Milner gave a modal logic (H–M logic) like PDL, to describe processes.

Any two bisimilar processes satisfy the same Hennessy–Milner formulas.

6. Modal μ-calculus

Thanks to Clemens Kupke for much of this material

Recall

Modal logic: decidable fragment of first-order logic.

Question

Is this fragment large (expressive) enough?

We will see

- there are properties we want to express that are not expressible in modal logic
- we can increase the expressivity of modal logic greatly and still obtain a decidable logic
- we do this by adding *fixed point operators* to the language
- we get a system called the *modal μ-calculus*

Forefathers of the modal μ-calculus include de Bakker and Scott. The father of the μ-calculus in its present form is Dexter Kozen.

The limited expressivity of basic modal logic

Let A be a formula. Suppose we want to express the following property that we call "$\textsc{AtSomePoint}(A)$":

> *"from the current world we can reach a world at which A is true, in a finite number of steps"*

- This is obviously not expressible in the basic modal language (prove this as an exercise!).
- It is expressible in temporal logic interpreted in $(\mathbb{N}, <)$, but not in an arbitrary Kripke model.
- The property can be nicely expressed as a fixed point.

Fixed points

Definition 6.1

Let S be a set, and $f : S \to S$ be a function.
We call an element $x \in S$ a *fixed point of f* if $f(x) = x$.

Intuitively, we can see the property $\textsc{AtSomePoint}(A)$ as a kind of fixed point of the 'function' $X \mapsto A \vee \Diamond X$:

$$\textsc{AtSomePoint}(A) \equiv A \vee \Diamond\textsc{AtSomePoint}(A)$$

We will

- see and understand that the property $\textsc{AtSomePoint}(A)$ can be defined as a *least* fixed point
- enrich the modal language with so-called least and greatest fixed point operators, denoted by μ ('mu') and ν ('nu')
- introduce the fixed point semantics

Monotonic functions

The language of the modal μ-calculus will contain formulas of the form $\mu p A$ and $\nu p A$. These formulas will be interpreted as *least* and *greatest* fixed points of certain *monotonic* functions.

Definition 6.2

Let S be a set and let $f : \wp(S) \to \wp(S)$ be a function (recall $\wp(S) = \{U : U \subseteq S\}$).

1. We say that f is *monotonic* if for all sets $U, V \subseteq S$,
$$\text{if} \quad U \subseteq V \quad \text{then} \quad f(U) \subseteq f(V).$$
2. Suppose that U is a fixed point of f: that is, $f(U) = U$. We say that U is the *least fixed point* of f if for all fixed points V of f we have $U \subseteq V$. We call U the *greatest fixed point* of f if for all fixed points V of f we have $V \subseteq U$.
3. If they exist, we write $\mathrm{LFP}(f)$ and $\mathrm{GFP}(f)$ for the least and greatest fixed point of f, respectively.
(Obviously, they are unique.)

Fixed points on finite sets

We first give a concrete representation of fixed points on *finite* sets.

Proposition 6.3

Let S be a finite set, and let $f : \wp(S) \to \wp(S)$ be a monotonic function. Define, for all $U \subseteq S$ and all $n \in \mathbb{N}$,

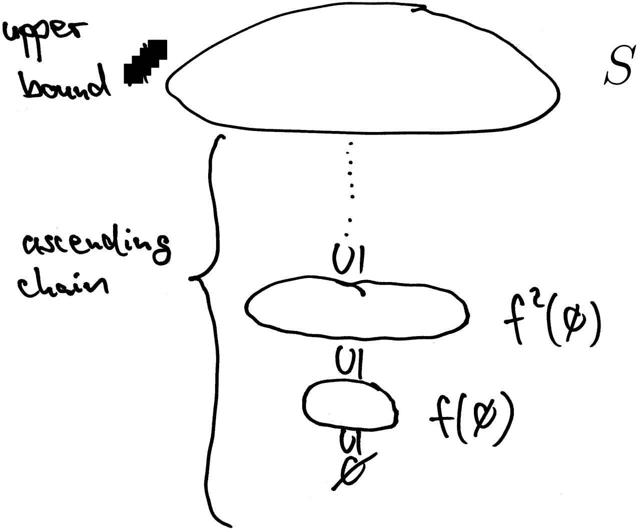
- $f^0(U) = U$
- $f^{n+1}(U) = f(f^n(U))$

Then the least fixed point $\mathrm{LFP}(f)$ and the greatest fixed point $\mathrm{GFP}(f)$ of f exist, and can be computed as follows:

$$\mathrm{LFP}(f) = \bigcup_{n\in\mathbb{N}} f^n(\emptyset) \qquad \mathrm{GFP}(f) = \bigcap_{n\in\mathbb{N}} f^n(S)$$

Proof. We will now prove that this characterisation of the least fixed point of f is correct.
The details for the greatest fixed point are left as an exercise.

Approximation of the least fixed point



$\bigcup_{n\in\mathbb{N}} f^n(\emptyset)$ is a fixed point of f

We have $f^0(\emptyset) = \emptyset \subseteq f^1(\emptyset)$, as $\emptyset \subseteq U$ for all U.

Since $f^0(\emptyset) \subseteq f^1(\emptyset)$, by monotonicity of f we have $f(f^0(\emptyset)) \subseteq f(f^1(\emptyset))$ — that is, $f^1(\emptyset) \subseteq f^2(\emptyset)$.
Since $f^1(\emptyset) \subseteq f^2(\emptyset)$, by monotonicity of f we get $f^2(\emptyset) \subseteq f^3(\emptyset)$.

And so on. We get a sequence

$$\emptyset \subseteq f^1(\emptyset) \subseteq f^2(\emptyset) \subseteq \cdots \subseteq f^n(\emptyset) \subseteq f^{n+1}(\emptyset) \subseteq \cdots \subseteq S$$

As S is finite, this sequence has to stabilise at some point: there must be some $m \in \mathbb{N}$ such that $f^m(\emptyset) = f(f^m(\emptyset))$.

So $f^m(\emptyset)$ is a fixed point of f, and $\bigcup_{n\in\mathbb{N}} f^n(\emptyset) = f^m(\emptyset)$.

$\bigcup_{n\in\mathbb{N}} f^n(\emptyset)$ is the least fixed point of f

To show that $f^m(\emptyset)$ is the *least* fixed point of f, let U be any fixed point of f. We show that $f^m(\emptyset) \subseteq U$.

Claim. $f^n(\emptyset) \subseteq U$ for all $n \in \mathbb{N}$.

Proof of claim. The proof is by induction on n.
We have $f^0(\emptyset) = \emptyset \subseteq U$ obviously.

Assume inductively that $f^n(\emptyset) \subseteq U$.
Monotonicity of f gives $f(f^n(\emptyset)) \subseteq f(U)$.
But $f(U) = U$. So this boils down to $f^{n+1}(\emptyset) \subseteq U$.
This completes the induction. **Claim** ✓

In particular, $f^m(\emptyset) \subseteq U$.

As U was an arbitrary fixed point of f, this shows that $f^m(\emptyset)$ is the least fixed point of f. □

Example

Exercise 6.4

- *Let* $S = \{1, 2, 3, \dots, 10\}$.
- *Let* $f : \wp(S) \to \wp(S)$ *be the function given by*

$$f(U) = U \cup \{2\} \cup \{n + 2 : n \in U,\ n \leq 8\},$$

for $U \subseteq S$.

1. *Show that the function* f *is monotonic.*
2. *What is the least fixed point of* f*?*
3. *What is the greatest fixed point of* f*?*

What about infinite sets?

We saw that the least and greatest fixed point of a monotonic function $f : \wp(S) \to \wp(S)$ always exist if S is finite.

In fact, *these fixed points always exist* — even if S is infinite. (Follows from 'Knaster–Tarski theorem'.)

For infinite sets S, proposition 6.3 can fail, but we do have the following. (Proof is an exercise.)

Lemma 6.5

For any monotonic function $f : \wp(S) \to \wp(S)$ we have

$$\bigcup_{n\in\mathbb{N}} f^n(\emptyset) \subseteq \mathrm{LFP}(f), \qquad \mathrm{GFP}(f) \subseteq \bigcap_{n\in\mathbb{N}} f^n(S).$$

In particular, if $\bigcup_{n\in\mathbb{N}} f^n(\emptyset)$ and $\bigcap_{n\in\mathbb{N}} f^n(S)$ are fixed points of f, then they are the least and greatest fixed points of f, respectively.

Modal μ-calculus: syntax

We will define μ-calculus formulas in “negation normal form”: negations only occur directly before atoms.
To get decent expressivity, we treat $\vee$, $\Diamond$, and $\bot$ as primitive connectives, not abbreviations.
Recall that L is our fixed set of atoms.

Definition 6.6 (the set L_μ of modal μ-calculus formulas)

- any propositional atom $p \in L$ is an L_μ-formula
- $\top$ and $\bot$ are L_μ-formulas
- if p is a propositional atom, then $\neg p$ is an L_μ-formula
- if A, B are L_μ-formulas, then so are
  $$(A \wedge B) \qquad (A \vee B) \qquad \Box A \qquad \Diamond A$$
- if A is a L_μ-formula and $\neg p$ does not occur in A, then
  $$\mu p A \qquad \text{and} \qquad \nu p A$$
  are L_μ-formulas.

Semantics of L_μ-formulas — modal cases

Let (W, R) be a Kripke frame. For a set $U \subseteq W$, write

$$\begin{array}{rcl} \Box U & = & \{w \in W : \forall w'(R(w,w') \Rightarrow w' \in U)\} \\ \Diamond U & = & \{w \in W : \exists w'(R(w,w') \text{ and } w' \in U)\} \end{array}$$

For each assignment h into W, we define the *semantics*

$$[\![A]\!]_h = \{w \in W : (W, R, h), w \models A\}$$

of an L_μ-formula A by induction on the structure of A:

$$\begin{array}{rcl} [\![p]\!]_h & = & h(p), \quad \text{for an atom } p \\ [\![\top]\!]_h & = & W \text{ and } [\![\bot]\!]_h = \emptyset \\ [\![\neg p]\!]_h & = & W \setminus h(p) = \{w \in W : w \notin h(p)\} \\ [\![A \wedge B]\!]_h & = & [\![A]\!]_h \cap [\![B]\!]_h \\ [\![A \vee B]\!]_h & = & [\![A]\!]_h \cup [\![B]\!]_h \\ [\![\Box A]\!]_h & = & \Box[\![A]\!]_h \\ [\![\Diamond A]\!]_h & = & \Diamond[\![A]\!]_h \end{array}$$

Semantics of L_μ-formulas — μ- and ν-formulas

Assume $[\![A]\!]_h$ has been defined for all h, and $\neg p$ doesn't occur in A. For each h, define a function $A_p^h : \wp(W) \to \wp(W)$ by

$$A_p^h(U) = [\![A]\!]_{h[p \mapsto U]} \quad \text{for each } U \subseteq W,$$

where the assignment $h[p \mapsto U] : L \to \wp(W)$ is defined by

$$h[p \mapsto U](q) = \begin{cases} U & \text{if } q = p \\ h(q) & \text{otherwise} \end{cases}$$

As $\neg p$ doesn't occur in A, this function turns out to be monotonic. (Cf. lemma 4.3.)

We interpret $\mu p A$ and $\nu p A$ as its least and greatest fixed points:

$$\begin{array}{rcl} [\![\mu p A]\!]_h & = & \mathrm{LFP}(A_p^h) \\ [\![\nu p A]\!]_h & = & \mathrm{GFP}(A_p^h) \end{array}$$

By the Knaster–Tarski theorem, this is well defined in all Kripke models.

Summary: fixed points are just repeated evaluation

So you should be able to do it...

Let p be an atom, A a L_μ-formula with no $\neg p$, and $\mathcal{M} = (W, R, h)$ a *finite* model. By definition of $[\![\mu p A]\!]_h$ and proposition 6.3,

$$[\![\mu p A]\!]_h = \mathrm{LFP}(A_p^h) = \bigcup_{n \in \mathbb{N}} (A_p^h)^n(\emptyset).$$

To compute this,

1. calculate $[\![A]\!]_h$ with p set to $\emptyset$ — say the result is $U_1 \subseteq W$
2. calculate $[\![A]\!]_h$ with p set to U_1 — say the result is $U_2 \subseteq W$
3. calculate $[\![A]\!]_h$ with p set to U_2 — say the result is $U_3 \subseteq W$

and so on, until the U_n stabilise. The stable value is $[\![\mu p A]\!]_h$.
In *infinite* models, *if* $\bigcup_{n \in \mathbb{N}} U_n$ is a fixed point of A_p^h, it is $[\![\mu p A]\!]_h$.

For $\nu p A$:

- use $\bigcap$ instead of $\bigcup$
- start with p set to W instead of $\emptyset$

Examples: formulas as operators

Example 6.7

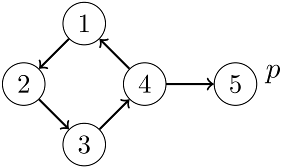
Let $\mathcal{M} = (W, R, h)$ be a Kripke model. For a set $U \subseteq W$, recall

$$\begin{array}{rcl} \Box U & = & \{w \in W : \forall w'(R(w,w') \Rightarrow w' \in U)\} \\ \Diamond U & = & \{w \in W : \exists w'(R(w,w') \text{ and } w' \in U)\} \end{array}$$

1. Let $A = p \vee \Diamond q$. Then
$$A_q^h(U) = h(p) \cup \Diamond U.$$
2. Let $B = \Box\Diamond(p \vee q)$. Then
$$B_q^h(U) = \Box\Diamond(h(p) \cup U).$$
3. Let $C = p \vee (r \wedge \Box q)$. Then
$$C_q^h(U) = h(p) \cup (h(r) \cap \Box U).$$

Example: Semantics of a formula

Example 6.8



Let $B = \Box\Diamond(p \vee q)$ and let's compute $[\![\mu q B]\!]_h$ and $[\![\nu q B]\!]_h$.
By definition, $[\![\mu q B]\!]_h = \mathsf{LFP}(f)$ with $f = B_q^h$. We have:

$$\begin{array}{rcl} f(\emptyset) & = & \Box\Diamond(h(p) \cup \emptyset) = \Box\Diamond(\{5\}) = \{3,5\} \\ f^2(\emptyset) & = & \Box\Diamond(h(p) \cup f(\emptyset)) = \Box\Diamond(\{3,5\}) = \{1,3,5\} \\ f^3(\emptyset) & = & \Box\Diamond(h(p) \cup f^2(\emptyset)) = \Box\Diamond(\{1,3,5\}) = \{1,3,5\} \end{array}$$

This implies that $\bigcup_{n\in\mathbb{N}} f^n(\emptyset) = f^2(\emptyset)$.
By proposition 6.3 (slide 105), $[\![\mu q B]\!]_h = \mathsf{LFP}(f) = \{1,3,5\}$.

Exercise: Compute $[\![\nu q B]\!]_h$ in example 6.8!

Semantics of $\mu q(A \vee \Diamond q)$ — reachability

Example 6.9

Let A be any formula, and q an atom not occurring in A.
Let $\mathcal{M} = (W, R, h)$ be a *finite* Kripke model.
Let $f = (A \vee \Diamond q)^h_q$. For any $U \subseteq W$ we have

$$f(U) = [\![A]\!]_h \cup \Diamond U.$$

So $[\![\mu q(A \vee \Diamond q)]\!]_h = \mathsf{LFP}(f) = \bigcup_{n \in \mathbb{N}} f^n(\emptyset)$ (by proposition 6.3).
Let's compute it:

$$\begin{array}{rcl}
f(\emptyset) & = & [\![A]\!]_h \\
f^2(\emptyset) & = & [\![A]\!]_h \cup \Diamond [\![A]\!]_h \\
f^3(\emptyset) & = & [\![A]\!]_h \cup \Diamond([\![A]\!]_h \cup \Diamond [\![A]\!]_h) = \text{‘}\Diamond^{\leq 2}[\![A]\!]_h\text{’} \\
\vdots & & \\
f^n(\emptyset) & = & \{w : \text{‘}[\![A]\!]_h \text{ is } R\text{-reachable from } w \\
 & & \qquad\qquad \text{in fewer than } n \text{ steps’}\},
\end{array}$$

where the last line can be proved by induction on n.

$\mu q(A \vee \Diamond q)$ expresses $\textsc{AtSomePoint}(A)$

So

$$\begin{aligned}
[\![\mu q(A \vee \Diamond q)]\!]_h &= \mathrm{LFP}(f) = \textstyle\bigcup_{n\in\mathbb{N}} f^n(\emptyset) \\
&= \{w \in W : \text{`}[\![A]\!]_h \text{ is } R\text{-reachable from } w \\
&\qquad\qquad\qquad \text{in a finite number of steps'.}\}
\end{aligned}$$

One can check that this is also true on an *arbitrary* (possibly *infinite*) model $\mathcal{M}$! (exercise!)

So we have expressed $\textsc{AtSomePoint}(A)$ in the μ-calculus using the *least fixed point* operator, by $\mu q(A \vee \Diamond q)$, where q does not occur in A.

What about $\nu q(A \vee \Diamond q)$?

Example 6.10

Let $\mathcal{M} = (W, R, h)$ be a model. We claim that

$$\begin{array}{rcl} [\![\nu q(A \vee \Diamond q)]\!]_h & = & X := \{w \in W : \textit{“either from } w \textit{ we can reach} \\ & & \quad \textit{a world in } [\![A]\!]_h \textit{ in a finite number of steps,} \\ & & \quad \textit{or there is an infinite path starting at } w\textit{”}\ \} \end{array}$$

To see this, let $f = (A \vee \Diamond q)^h_q$ again, and compute $f^n(W)$:

$$\begin{array}{rcl} f(W) & = & [\![A]\!]_h \cup \Diamond W \\ f^2(W) & = & [\![A]\!]_h \cup \Diamond([\![A]\!]_h \cup \Diamond W) = [\![A]\!]_h \cup \Diamond[\![A]\!]_h \cup \Diamond^2 W \\ f^n(W) & = & \{w \in W : \text{“either } [\![A]\!]_h \text{ is } R\text{-reachable from } w \text{ in} \\ & & \quad < n \text{ steps, or a path of length } n \text{ starts at } w\text{”}\} \end{array}$$

Our claim follows, for finite models $\mathcal{M}$ (take $n > |W|$).
The claim also holds for infinite models: show X is a fixed point of f, and (tricky) any fixed point Y is contained in X.

Another example: $\nu q\big(p \vee (r \wedge \Box q)\big)$

Example 6.11

Let $\mathcal{M} = (W, R, h)$ be a model. We claim that

$$\begin{array}{rcl} [\![\nu q\big(p \vee (r \wedge \Box q)\big)]\!]_h & = & \{w \in W : \text{"on all paths starting from } w, \\ & & \qquad \text{while } \neg p \text{ is true, } r \text{ has to be true."}\} \end{array}$$

Let $f = (p \vee (r \wedge \Box q))^h_q$. We calculate

$$\begin{array}{rcl} f(W) & = & h(p) \cup (h(r) \cap \Box W) = h(p) \cup h(r) \\ f^2(W) & = & h(p) \cup (h(r) \cap \Box f(W)) \\ & = & h(p) \cup \big(h(r) \cap \Box(h(p) \cup h(r))\big) \\ f^3(W) & = & h(p) \cup (h(r) \cap \Box f^2(W)) \\ \vdots & & \\ f^n(W) & = & \{w \in W : \text{"on all paths of length} < n \text{ starting at } w, \\ & & \qquad \text{while } \neg p \text{ is true, } r \text{ is true."}\} \end{array}$$

In order to compute $[\![\nu q(p \vee (r \wedge \Box q))]\!]_h$ on a finite model we now compute the intersection

$$\begin{aligned} [\![\nu q(p \vee (r \wedge \Box q))]\!]_h &= \bigcap_{n \in \mathbb{N}} f^n(W) \\ &= \{w \in W : \text{ `on all paths starting at } w\text{,} \\ &\qquad\qquad \text{while } \neg p \text{ is true, } r \text{ is true.'}\} \end{aligned}$$

Can check that the above equality holds on infinite models as well. So $\nu q(p \vee (r \wedge \Box q))$ is like While, or weak Until.

Some more examples for you to try

Exercise 6.12

Show that

1. $\nu q(r \wedge \Box q)$ *is true at* w *in a model* $\mathcal{M}$ *iff "*r *holds at all worlds reachable from* w *(including* w *itself)"*
2. $\mu p\big((\nu q(r \wedge \Box q)) \vee \Diamond p\big)$ *is true at* w *in a model* $\mathcal{M}$ *iff "from* w *we can reach a world* v *in a finite number of steps such that* r *is true at all worlds reachable from* v*"*

Important properties of the modal μ-calculus

Theorem 6.13 (Kozen, 1982)

The modal μ-calculus has the finite model property: that is, a formula $A \in L_\mu$ is satisfiable iff A is satisfiable in a finite model.

Proofs use 'better-quasi-orderings', automata, or tableaux.

Theorem 6.14 (Kozen & Parikh, 1984, Emerson & Jutla, 1988)

The modal μ-calculus is decidable (and EXPTIME*-complete).*

K–P: reduction to decidable logic 'SnS'. E–J: automata.

Theorem 6.15 (Janin & Walukiewicz, 1996)

Over Kripke frames, the modal μ-calculus corresponds to the bisimulation-invariant fragment of monadic second-order logic.

Compare van Benthem's theorem 5.26. Proof uses automata.

Modal μ-calculus: conclusion

- Very powerful but still decidable extension of basic modal logic: 'essentially the "ultimate" program logic' (Vardi 1998).
- Can express many lesser logics (such as PDL, 'CTL*').
  Increasingly used for theoretical purposes.
  But few people can read or write complex μ-formulas!
- We didn't show the semantics is always well defined (no time).
- Big omission: didn't look at the game-theoretic semantics of the modal μ-calculus. This is particularly useful for more complicated formulas with 'nesting' of fixed point operators:

$$\mu p\,\nu q((s \wedge \Box p) \vee (\neg s \wedge \Box q))$$

- Sahlqvist theory for μ-calculus is still being developed.

Summary of the course

We introduced basic modal logic with $\Box$. (We mentioned other modal logics: eg. PDL.)

We saw how certain modal formulas correspond to frame properties (reflexivity, transitivity, etc). General form: Sahlqvist's theorem. So we can begin to tailor modal logic to applications.

We saw how p-morphisms preserve modal formulas, and bisimulations exactly capture their strength. So we see their limited expressivity.

We looked at the mu-calculus, a powerful extension of basic modal logic of increasing popularity in applications.

Limitations of the course

Through lack of time, we did not cover:

- Canonical model & canonicity, filtration.
- Model checking (done in 3rd year verification course).
- Decidability. Complexity (you need course 438 for this).
- Many-dimensional modal logics. Formulas are evaluated at *sequences* of possible worlds. Eg intervals of time, multi-agent/distributed systems. The going gets tough.
- Related topic: first-order modal and temporal logic.
- Hybrid logic. Spatial logic. Coalgebras. Description logic.
- Algebraic logic. By regarding models as algebras, and using different techniques, this allows a finer analysis.
- A lot more.

See the books (eg. Blackburn–de Rijke–Venema) for more information on these areas.

Enjoy your holidays!