2

Models

In Section 1.3 we defined what it means for a formula to be *satisfied* at a state in a model — but as yet we know virtually nothing about this fundamental semantic notion. What exactly can we say about models when we use modal languages to describe them? Which properties of models can modal languages express, and which lie beyond their reach?

In this chapter we examine such questions in detail. We introduce *disjoint unions, generated submodels, bounded morphisms*, and *ultrafilter extensions*, the 'big four' operations on models that leave modal satisfaction unaffected. We discuss two ways to obtain finite models and show that modal languages have the *finite model property*. Moreover, we define the *standard translation* of modal logic into first-order logic, thus opening the door to *correspondence theory*, the systematic study of the relationship between modal and classical logic. All this material plays a fundamental role in later work; indeed, the basic track sections in this chapter are among the most important in the book.

But the central concept of the chapter is that of a *bisimulation* between two models. Bisimulations reflect, in a particularly simple and direct way, the locality of the modal satisfaction definition. We introduce them early on, and they gradually come to dominate our discussion. By the end of the chapter we will have a good understanding of modal expressivity over models, and the most interesting results all hinge on bisimulations.

Chapter guide

Section 2.1: Invariance Results (Basic track). We introduce three classic ways of constructing new models from old ones that do not affect modal satisfaction: disjoint unions, generated submodels, and bounded morphisms. We also meet isomorphisms and embeddings.

Section 2.2: Bisimulations (Basic track). We introduce bisimulations and show that modal satisfaction is invariant under bisimulation. We will see that the model constructions introduced in the first section are all special cases of bisimulation, learn that modal equivalence does not always imply bisimilarity, and examine an important special case in which it does.

Section 2.3: Finite Models (Basic track). Here we show that modal languages enjoy the finite model property. We do so in two distinct ways: by the selection method (finitely approximating a bisimulation), and by filtration (collapsing a model into a finite number of equivalence classes).

Section 2.4: The Standard Translation (Basic track). We start our study of correspondence theory. By defining the standard translation, we link modal languages to first-order (and other classical) languages and raise the two central questions that dominate later sections: What part of first-order logic does modal logic correspond to? And which properties of models are definable by modal means?

Section 2.5: Modal Saturation via Ultrafilter Extensions (Basic track). The first step towards obtaining some answers is to introduce ultrafilter extensions, the last of the big four modal model constructions. We then show that although modal equivalence does not imply bisimilarity, it does imply bisimilarity somewhere else, namely in the ultrafilter extensions of the models concerned.

Section 2.6: Characterization and Definability (Advanced track). We prove the two main results of this chapter. First, we prove van Benthem's theorem stating that modal languages are the bisimulation invariant fragments of first-order languages. Second, we show that modally definable classes of (pointed) models are those that are closed under bisimulations and ultraproducts and whose complements are closed under ultrapowers.

Section 2.7: Simulations and Safety (Advanced track). We prove two results that give the reader a glimpse of recent work in modal model theory. The first describes the properties that are preserved under simulations (a one-way version of bisimulation), the second characterizes the first-order definable operations on binary relations which respect bisimilarity.

2.1 Invariance Results

Mathematicians rarely study structures in isolation. They are usually interested in the relations *between* different structures, and in *operations* that build new structures from old. Questions that naturally arise in such a context concern the structural properties that are invariant under or preserved by such relations and operations. We'll not give precise definitions of these notions, but roughly speaking, a property is *preserved* by a certain relation or operation if, whenever two structures are linked by the relation or operation, then the second structure has the property if the first one has it. We speak of *invariance* if the property is preserved in both directions.

When it comes to this research topic, logic is no exception to the rule — indeed, logicians add a descriptive twist to it. For instance, modal logicians want to know when two structures, or perhaps two points in distinct structures, are indistinguishable by modal languages in the sense of satisfying the same modal formulas.

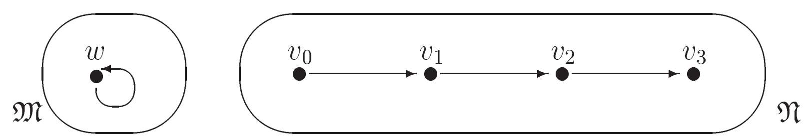
Definition 2.1 Let $\mathfrak{M}$ and $\mathfrak{M}'$ be models of the same modal similarity type τ, and let w and w' be states in $\mathfrak{M}$ and $\mathfrak{M}'$ respectively. The *τ-theory* (or *τ-type*) *of* w is the set of all τ-formulas satisfied at w: that is, $\{\phi \mid \mathfrak{M}, w \Vdash \phi\}$. We say that w and w' are *(modally) equivalent* (notation: $w \leftrightsquigarrow w'$) if they have the same τ-theories.

The *τ-theory* of the model $\mathfrak{M}$ is the set of all τ-formulas satisfied by all states in $\mathfrak{M}$: that is, $\{\phi \mid \mathfrak{M} \Vdash \phi\}$. Models $\mathfrak{M}$ and $\mathfrak{M}'$ are called *(modally) equivalent* (notation: $\mathfrak{M} \leftrightsquigarrow \mathfrak{M}'$) if their theories are identical. $\dashv$

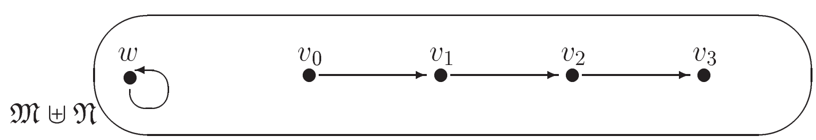
We now introduce three important ways of constructing new models from old ones which leave the theories associated with states unchanged: *disjoint unions*, *generated submodels*, and *bounded morphisms*. These constructions (together with *ultrafilter extensions*, which we introduce in Section 2.5) play an important role throughout the book. For example, in the following chapter we will see that they lift to the level of frames (where they preserve validity), we will use them repeatedly in our work on completeness and complexity, and in Chapter 5 we will see that they have important algebraic analogs.

Disjoint Unions

Suppose we have the following two models:



Don't worry that we haven't specified the valuations — they're irrelevant here. All that matters is that $\mathfrak{M}$ and $\mathfrak{N}$ have disjoint domains, for we are now going to lump them together to form the model $\mathfrak{M} \uplus \mathfrak{N}$:



The model $\mathfrak{M} \uplus \mathfrak{N}$ is called the *disjoint union* of $\mathfrak{M}$ and $\mathfrak{N}$. It gathers together all the information in the two smaller models unchanged: we have not altered the way the points are related, nor the way atomic information is distributed. Suppose we're working in the basic modal language, and suppose that a formula ϕ is true at (say) v_1 in $\mathfrak{N}$: is ϕ still true at v_1 in $\mathfrak{M} \uplus \mathfrak{N}$? More generally, is modal satisfaction *preserved* from points in the original models to the points in the disjoint union? And what about the reverse direction: if a modal formula is true at some state in $\mathfrak{M} \uplus \mathfrak{N}$, is it also true at that same state in the smaller model it came from?

The answer to these questions is clearly *yes*: modal satisfaction must be *invariant* (that is, preserved in both directions) under the formation of disjoint unions. Modal satisfaction is intrinsically local: only the points accessible from the current state are relevant to truth or falsity. If we evaluate a formula ϕ at (say) w, it is completely irrelevant whether we perform the evaluation in $\mathfrak{M}$ or $\mathfrak{M} \uplus \mathfrak{N}$; ϕ simply cannot detect the presence or absence of states in other islands.

Definition 2.2 (Disjoint Unions) We first define disjoint unions for the basic modal language. We say that two models are *disjoint* if their domains contain no common elements. For disjoint models $\mathfrak{M}_i = (W_i, R_i, V_i)$ ($i \in I$), their *disjoint union* is the structure $\biguplus_i \mathfrak{M}_i = (W, R, V)$, where W is the union of the sets W_i, R is the union of the relations R_i, and for each proposition letter p, $V(p) = \bigcup_{i \in I} V_i(p)$.

Now for the general case. For disjoint τ-structures $\mathfrak{M}_i = (W_i, R_{\triangle i}, V_i)_{\triangle \in \tau}$ ($i \in I$) of the same modal similarity type τ, their *disjoint union* is the structure $\biguplus_i \mathfrak{M}_i = (W, R_\triangle, V)_{\triangle \in \tau}$ such that W is the union of the sets W_i; for each $\triangle \in \tau$, $R_\triangle$ is the union $\bigcup_{i \in I} R_{\triangle i}$; and V is defined as in the basic modal case.

If we want to put together a collection of models that are *not* disjoint, we first have to make them disjoint (say by indexing the domains of these models). To use the terminology introduced shortly, we simply take mutually disjoint isomorphic copies of the models we wish to combine, and combine the copies instead. $\dashv$

Proposition 2.3 *Let τ be a modal similarity type and, for all $i \in I$, let $\mathfrak{M}_i$ be a τ-model. Then, for each modal formula ϕ, for each $i \in I$, and each element w of $\mathfrak{M}_i$, we have $\mathfrak{M}_i, w \Vdash \phi$ iff $\biguplus_{i \in I} \mathfrak{M}_i, w \Vdash \phi$. In words: modal satisfaction is invariant under disjoint unions.*

Proof. We will prove the result for the basic similarity type. The proof is by induction on ϕ (we explained this concept in Exercise 1.3.1). Let i be some index; we will prove, for each basic modal formula ϕ, and each element w of $\mathfrak{M}_i$, that $\mathfrak{M}_i, w \Vdash \phi$ iff $\mathfrak{M}, w \Vdash \phi$, where $\mathfrak{M}$ is the disjoint union $\biguplus_{i \in I} \mathfrak{M}_i$.

First suppose that ϕ contains no connectives. Now, if ϕ is a proposition letter p, then we have $\mathfrak{M}_i, w \Vdash \phi$ iff $w \in V_i(p)$ iff (by definition of V) $w \in V(p)$ iff $\mathfrak{M}, w \Vdash \phi$. On the other hand, ϕ could be $\bot$ (for the purposes of inductive proofs it is convenient to regard $\bot$ as a propositional letter rather than as a logical connective). But trivially $\bot$ is false at w in both models, so we have the desired equivalence here too.

Our inductive hypothesis is that the desired equivalence holds for all formulas containing at most n connectives (where $n \geq 0$). We must now show that the equivalence holds for all formulas ϕ containing $n+1$ connectives. Now, if ϕ is of the form $\neg\psi$ or $\psi \vee \theta$ this is easily done — we will leave this to the reader — so as we are working with the basic similarity type, it only remains to establish the equivalence for formulas of the form $\Diamond\psi$. So assume that $\mathfrak{M}_i, w \Vdash \Diamond\psi$. Then there is a state v in $\mathfrak{M}_i$ with $R_i wv$ and $\mathfrak{M}_i, v \Vdash \psi$. By the inductive hypothesis, $\mathfrak{M}, v \Vdash \psi$. But by definition of $\mathfrak{M}$, we have Rwv, so $\mathfrak{M}, w \Vdash \Diamond\psi$.

For the other direction, assume that $\mathfrak{M}, w \Vdash \Diamond\psi$ holds for some w in $\mathfrak{M}_i$. Then there is a v with Rwv and $\mathfrak{M}, v \Vdash v$. It follows by the definition of R that $R_j wv$ for some j, and by the disjointness of the universes we must have that $j = i$. But then we find that v belongs to $\mathfrak{M}_i$ as well, so we may apply the inductive hypothesis; this yields $\mathfrak{M}_i, v \Vdash \psi$, so we find that $\mathfrak{M}_i, w \Vdash \Diamond\psi$. $\dashv$

We will use Proposition 2.3 all through the book — here is a simple application which hints at the ideas we will explore in Chapter 7.

Example 2.4 Defined modalities are a convenient shorthand for concepts we find useful. We have already seen some examples. In this book $\Box$, the 'true at all accessible states modality', is shorthand for $\neg\Diamond\neg\phi$, and we have inductively defined a 'true somewhere n-steps from here' modality $\Diamond^n$ for each natural number n (see Example 1.22). But while it is usually easy to show that some modality *is* definable (we need simply write down its definition), how do we show that some proposed operator is *not* definable? Via invariance results! As an example, consider the *global modality*. The global diamond E has as its (intended) accessibility relation the relation $W \times W$ implicitly present in any model. That is:

$$\mathfrak{M}, w \Vdash \mathrm{E}\phi \text{ iff } \mathfrak{M}, v \Vdash \phi \text{ for } \textit{some} \text{ state } v \text{ in } \mathfrak{M}.$$

Its dual, A, the global box, thus has the following interpretation:

$$\mathfrak{M}, w \Vdash \mathrm{A}\phi \text{ iff } \mathfrak{M}, v \Vdash \phi \text{ for } \textit{all} \text{ states } v \text{ in } \mathfrak{M}.$$

Thus the global modality brings a genuinely global dimension to modal logic. But is it definable in the basic modal language? Intuitively, *no*: as $\Diamond$ and $\Box$ work locally, it seems unlikely that they can define a truly global modality over arbitrary structures. Fine — but how do we *prove* this?

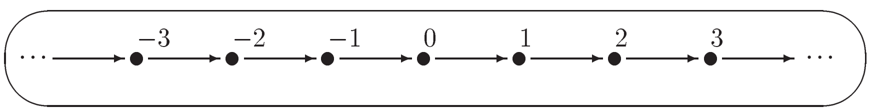
With the help of the previous proposition. Suppose we could define A. Then we could write down an expression $\alpha(p)$ containing only symbols from the basic modal language such that for every model $\mathfrak{M}$, $\mathfrak{M}, w \Vdash \alpha(p)$ iff $\mathfrak{M} \Vdash p$. We now derive a contradiction from this supposition. Consider a model $\mathfrak{M}_1$ where p holds everywhere, and a model $\mathfrak{M}_2$ where p holds nowhere. Let w be some point in $\mathfrak{M}_1$. It follows that $\mathfrak{M}_1, w \Vdash \alpha(p)$, so as (by assumption) $\alpha(p)$ contains only symbols from the basic modal language, by Proposition 2.3 we have that $\mathfrak{M}_1 \uplus \mathfrak{M}_2, w \Vdash \alpha(p)$. But this implies that $\mathfrak{M}_1 \uplus \mathfrak{M}_2, v \Vdash p$ for every v in $\mathfrak{M}_2$, which, again by Proposition 2.3, in turn implies that $\mathfrak{M}_2 \Vdash p$: contradiction. We conclude that the global box (and hence the global diamond) is *not* definable in the basic modal language.

So, if we want the global modality, then we either have to introduce it as a primitive (we will do this in Section 7.1), or we have to work with restricted classes of models on which it *is* definable (in Exercise 1.3.3 we worked with a class of models in which we could define A in the basic temporal language). $\dashv$

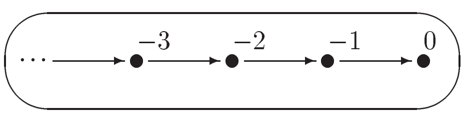
Generated submodels

Disjoint unions are a useful way of making bigger models from smaller ones — but we also want methods for doing the reverse. That is, we would like to know when it is safe to throw points away from a satisfying model without affecting satisfiability. Disjoint unions tell us a little about this (if a model is a disjoint union of smaller models, we are free to work with the component models), but this is not useful in practice. We need something sharper, namely *generated submodels*.

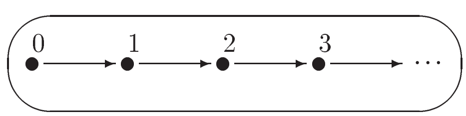
Suppose we are using the basic modal language to talk about a model $\mathfrak{M}$ based on the frame $(\mathbb{Z}, <)$, the integers with their usual order. It does not matter what the valuation is — all that's important is that $\mathfrak{M}$ looks something like this:



First suppose that we form a *submodel* $\mathfrak{M}^-$ of $\mathfrak{M}$ by throwing away all the positive numbers, and restricting the original valuation (whatever it was) to the remaining numbers. So $\mathfrak{M}^-$ looks something like this:



The basic modal language certainly *can* see that $\mathfrak{M}$ and $\mathfrak{M}^-$ are different. For example, it sees that 0 has successors in $\mathfrak{M}$ (note that $\mathfrak{M}, 0 \Vdash \Diamond\top$) but is a dead end in $\mathfrak{M}^-$ (note that $\mathfrak{M}^-, 0 \nVdash \Diamond\top$). So there's no invariance result for *arbitrary* submodels. But now consider the submodel $\mathfrak{M}^+$ of $\mathfrak{M}$ that is formed by omitting the negative numbers, and restricting the original valuation to the numbers that remain:



Suppose a basic modal formula ϕ is satisfied at some point n in $\mathfrak{M}$. Is ϕ also satisfied at the same point n in $\mathfrak{M}^+$? The answer must be *yes*. The only points that are relevant to ϕ's satisfiability are the points greater than n — *and all such points belong to* $\mathfrak{M}^+$. Similarly, it is clear that if $\mathfrak{M}^+$ satisfies a basic modal formula ϕ at m, then $\mathfrak{M}$ must too.

In short, it seems plausible that modal invariance holds for submodels which are closed under the accessibility relation of the original model. Such models are called *generated submodels*, and they do indeed give rise to the invariance result we are looking for.

Definition 2.5 (Generated Submodels) We first define generated submodels for the basic modal language. Let $\mathfrak{M} = (W, R, V)$ and $\mathfrak{M}' = (W', R', V')$ be two models; we say that $\mathfrak{M}'$ is a *submodel* of $\mathfrak{M}$ if $W' \subseteq W$, R' is the restriction of R to W' (that is: $R' = R \cap (W' \times W')$), and V' is the restriction of V to $\mathfrak{M}'$ (that is: for each p, $V'(p) = V(p) \cap W'$). We say that $\mathfrak{M}'$ is a *generated submodel* of $\mathfrak{M}$ (notation: $\mathfrak{M}' \rightarrowtail \mathfrak{M}$) if $\mathfrak{M}'$ is a submodel of $\mathfrak{M}$ and for all points w the following closure condition holds:

$$\text{if } w \text{ is in } \mathfrak{M}' \text{ and } Rwv, \text{ then } v \text{ is in } \mathfrak{M}'.$$

For the general case, we say that a model $\mathfrak{M}' = (W', R'_\triangle, V')_{\triangle \in \tau}$ is a *generated submodel* of the model $\mathfrak{M} = (W, R_\triangle, V)_{\triangle \in \tau}$ (notation: $\mathfrak{M}' \rightarrowtail \mathfrak{M}$) whenever $\mathfrak{M}'$ is a submodel of $\mathfrak{M}$ (with respect to $R_\triangle$ for all $\triangle \in \tau$), and the following closure condition is fulfilled for all $\triangle \in \tau$

$$\text{if } u \in W' \text{ and } R_\triangle u u_1 \ldots u_n, \text{ then } u_1, \ldots, u_n \in W'.$$

Let $\mathfrak{M}$ be a model, and X a subset of the domain of $\mathfrak{M}$; the *submodel generated by* X is the smallest generated submodel of $\mathfrak{M}$ whose domain contains X (such a model always exists: why?). Finally, a *rooted* or *point generated* model is a model that is generated by a singleton set, the element of which is called the *root* of the frame. ⊣

Proposition 2.6 *Let τ be a modal similarity type and let $\mathfrak{M}$ and $\mathfrak{M}'$ be τ-models such that $\mathfrak{M}'$ is a generated submodel of $\mathfrak{M}$. Then, for each modal formula ϕ and each element w of $\mathfrak{M}'$ we have that $\mathfrak{M}, w \Vdash \phi$ iff $\mathfrak{M}', w \Vdash \phi$. In words: modal satisfaction is invariant under generated submodels.*

Proof. By induction on ϕ. The reader unused to such proofs should write out the proof in full. In Proposition 2.19 we provide an alternative proof based on the observation that generated submodels induce a bisimulation. ⊣

Four remarks. First, note that the invariance result for disjoint unions (Proposition 2.3) is a special case of the result for generated submodels: any component of a disjoint union is a generated submodel of the disjoint union. Second, using an argument analogous to that used in Example 2.4 to show that the global box can't be defined in the basic modal language, we can use Proposition 2.6 to show that we cannot define a backward looking modality in terms of $\diamond$; see Exercise 2.1.2. Thus if we want such a modality we have to add it as a primitive — which is exactly what we did, of course, when defining the basic temporal language. Third, although we have not explicitly discussed generated submodels for the basic temporal language, PDL, or arrow logic, the required concepts are all special cases of Definition 2.5, and thus the respective invariance results are special cases of Proposition 2.6. But it is worth making a brief comment about the basic temporal language. When we think explicitly in terms of bidirectional frames (see Example 1.25) it is obvious that we are interested in submodels closed under both R_F and R_P. But when working with the basic temporal language we usually leave R_P implicit: we work with ordinary models (W, R, V), and use $R^{\breve{}}$, the converse of R, as R_P. Thus a *temporal* generated submodel of (W, R, V) is a submodel (W', R', V') that is closed under both R and $R^{\breve{}}$. Finally, generated submodels are heavily used throughout the book: given a model $\mathfrak{M}$ that satisfies a formula ϕ at a state w, very often the first thing we will do is form the submodel of $\mathfrak{M}$ generated by w, thus trimming what may be a very unwieldy satisfying model down to a more manageable one.

Morphisms for modalities

In mathematics the idea of *morphisms* or *structure preserving maps* is of fundamental importance. What notions of morphism are appropriate for modal logic? That is, what kinds of morphism give rise to invariance results? We will approach the answer bit by bit, introducing a number of important concepts on the way. We will start by considering the general notion of *homomorphism* (this is too weak to yield invariance, but it is the starting point for better attempts), then we will define *strong homomorphisms*, *embeddings*, and *isomorphisms* (these do give us invariance, but are not particularly modal), and finally we will zero in on the answer: *bounded morphisms*.

Definition 2.7 (Homomorphisms) Let τ be a modal similarity type and let $\mathfrak{M}$ and $\mathfrak{M}'$ be τ-models. By a *homomorphism* f from $\mathfrak{M}$ to $\mathfrak{M}'$ (notation: $f : \mathfrak{M} \to \mathfrak{M}'$) we mean a function f from W to W' with the following properties.

(i) For each proposition letter p and each element w from $\mathfrak{M}$, if $w \in V(p)$, then $f(w) \in V'(p)$.

(ii) For each $n > 0$ and each n-ary $\triangle \in \tau$, and $(n+1)$-tuple $\overline{w}$ from $\mathfrak{M}$, if $(w_0, \ldots, w_n) \in R_\triangle$ then $(f(w_0), \ldots, f(w_n)) \in R'_\triangle$ (the *homomorphic condition*).

We call $\mathfrak{M}$ the *source* and $\mathfrak{M}'$ the *target* of the homomorphism. $\dashv$

Note that for the basic modal language, item (ii) is just this:

if Rwu then $R'f(w)f(u)$.

Thus item (ii) simply says that homomorphisms preserve relational links.

Are modal formulas invariant under homomorphisms? No: although homomorphisms reflect the structure of the source in the structure of the target, they do not reflect the structure of the target back in the source. It is easy to turn this observation into a counterexample, and we will leave this task to the reader as Exercise 2.1.3.

So let us try and strengthen the definition. There is an obvious way of doing so: turn the conditionals into equivalences. This leads to a number of important concepts.

Definition 2.8 (Strong Homomorphisms, Embeddings and Isomorphisms) Let τ be a modal similarity type and let $\mathfrak{M}$ and $\mathfrak{M}'$ be τ-models. By a *strong homomorphism* of $\mathfrak{M}$ into $\mathfrak{M}'$ we mean a homomorphism $f : \mathfrak{M} \to \mathfrak{M}'$ which satisfies the following stronger version of the above items (i) and (ii):

(i) For each proposition letter p and element w from $\mathfrak{M}$, $w \in V(p)$ iff $f(w) \in V'(p)$.

(ii) For each $n \geq 0$ and each n-ary $\triangle$ in τ and $(n+1)$-tuple $\overline{w}$ from $\mathfrak{M}$, $(w_0, \ldots, w_n) \in R_\triangle$ iff $(f(w_0), \ldots, f(w_n)) \in R'_\triangle$ (the *strong homomorphic condition*).

An *embedding* of $\mathfrak{M}$ into $\mathfrak{M}'$ is a strong homomorphism $f : \mathfrak{M} \to \mathfrak{M}'$ which is injective. An *isomorphism* is a bijective strong homomorphism. We say that $\mathfrak{M}$ is *isomorphic* to $\mathfrak{M}'$, in symbols $\mathfrak{M} \cong \mathfrak{M}'$, if there is an isomorphism from $\mathfrak{M}$ to $\mathfrak{M}'$. $\dashv$

Note that for the basic modal language, item (ii) is just:

Rwu iff $R'f(w)f(u)$.

That is, item (ii) says that relational links are preserved from the source model to the target *and back again*. So it is not particularly surprising that we have a number of invariance results.

Proposition 2.9 *Let τ be a modal similarity type and let $\mathfrak{M}$ and $\mathfrak{M}'$ be τ-models. Then the following holds:*

(i) *For all elements w and w' of $\mathfrak{M}$ and $\mathfrak{M}'$, respectively, if there exists a surjective strong homomorphism $f : \mathfrak{M} \to \mathfrak{M}'$ with $f(w) = w'$, then w and w' are modally equivalent.*

(ii) *If* $\mathfrak{M} \cong \mathfrak{M}'$, *then* $\mathfrak{M} \leftrightsquigarrow \mathfrak{M}'$.

Proof. The first item follows by induction on ϕ; the second one is an immediate consequence. $\dashv$

None of the above results is particularly modal. For a start, as in all branches of mathematics, 'isomorphic' basically means 'mathematically identical'. Thus, we do not want to be able to distinguish isomorphic structures in modal (or indeed, any other) logic. Quite the contrary: we want to be free to work with structures 'up to isomorphism' — as we did, for example, in our discussion of disjoint union, when we talked of taking isomorphic copies. Item (ii) tells us that we can do this, but it isn't a surprising result.

But why is item (i), the invariance result for strong homomorphisms, not 'genuinely modal'? Quite simply, because there are many morphisms which do give rise to invariance, but which fail to qualify as strong homomorphisms. To ensure modal invariance we need to ensure that some target structure is reflected back in the source, but strong morphisms do this in a much too heavy-handed way. The crucial concept is more subtle.

Definition 2.10 (Bounded Morphisms — the Basic Case) We first define bounded morphisms for the basic modal language. Let $\mathfrak{M}$ and $\mathfrak{M}'$ be models for the basic modal language. A mapping $f : \mathfrak{M} = (W, R, V) \to \mathfrak{M}' = (W', R', V')$ is a *bounded morphism* if it satisfies the following conditions:

(i) w and $f(w)$ satisfy the same proposition letters.
(ii) f is a homomorphism with respect to the relation R (that is, if Rwv then $R'f(w)f(v)$).
(iii) If $R'f(w)v'$ then there exists v such that Rwv and $f(v) = v'$ (the *back condition*).

If there is a *surjective* bounded morphism from $\mathfrak{M}$ to $\mathfrak{M}'$, then we say that $\mathfrak{M}'$ is a *bounded morphic image* of $\mathfrak{M}$, and write $\mathfrak{M} \twoheadrightarrow \mathfrak{M}'$. $\dashv$

The idea embodied in the back condition is utterly fundamental to modal logic — in fact, it is the idea that underlies the notion of bisimulation — so we need to get a good grasp of what it involves right away. Here's a useful example.

Example 2.11 Consider the models $\mathfrak{M} = (W, R, V)$ and $\mathfrak{M}' = (W', R', V')$, where

- $W = \mathbb{N}$ (the natural numbers), Rmn iff $n = m + 1$, and $V(p) = \{n \in \mathbb{N} \mid n \text{ is even}\}$
- $W' = \{e, o\}$, $R' = \{(e, o), (o, e)\}$, and $V'(p) = \{e\}$.

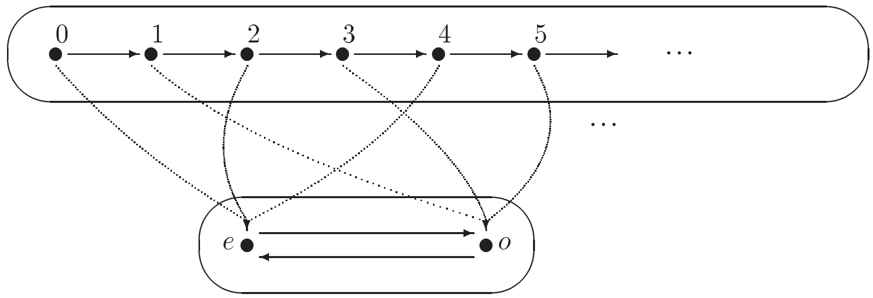


Fig. 2.1. A bounded morphism

Now, let $f : W \to W'$ be the following map:

$$f(n) = \begin{cases} e & \text{if } n \text{ is even} \\ o & \text{if } n \text{ is odd} \end{cases}$$

Figure 2.1 sums this all up in a simple picture.

Now, f is *not* a strong homomorphism (why not?), but it *is* a (surjective) bounded morphism from $\mathfrak{M}$ to $\mathfrak{M}'$. Let's see why. Trivially f satisfies item (i) of the definition. As for the homomorphic condition consider an arbitrary pair $(n, n+1)$ in R. There are two possibilities: n is either even or odd. Suppose n is even. Then $n+1$ is odd, so $f(n) = e$ and $f(n+1) = o$. But then we have $R'f(n)f(n+1)$, as required. The argument for n odd is analogous.

And now for the interesting part: the back condition. Take an arbitrary element n of W and assume that $R'f(n)w'$. We have to find an $m \in W$ such that Rnm and $f(m) = w'$. Let's assume that n is odd (the case for even n is similar). As n is odd, $f(n) = o$, so by definition of R', we must have that $w' = e$. But then $f(n+1) = w'$ since $n+1$ is even, and by the definition of R we have that $n+1$ is a successor of n. Hence, $n+1$ is the m that we were looking for. $\dashv$

Definition 2.12 (Bounded Morphisms — the General Case) The definition of a bounded morphism for general modal languages is obtained from the above by adapting the homomorphic and back conditions of Definition 2.10 as follows:

(ii)$'$ For all $\triangle \in \tau$, $R_\triangle w v_1 \dots v_n$ implies $R'_\triangle f(w) f(v_1) \dots f(v_n)$.

(iii)$'$ If $R'_\triangle f(w) v'_1 \dots v'_n$ then there exist $v_1 \dots v_n$ such that $R_\triangle w v_1 \dots v_n$ and $f(v_i) = v'_i$ (for $1 \leq i \leq n$). $\dashv$

Example 2.13 Suppose we are working in the modal similarity type of arrow logic; see Example 1.16 and 1.27. Recall that the language has a modal constant 1', a unary operator $\otimes$ and a single dyadic operator $\circ$. Semantically, to these operators correspond a unary relation I, a binary R and a ternary C. We will define a bounded morphism from a square model to a model based on the addition of the integer numbers. We will use the following notation: if x is an element of $\mathbb{Z} \times \mathbb{Z}$, then x_0 denotes its first component, and x_1 its second component.

Consider the two models $\mathfrak{M} = (W, C, R, I, V)$ and $\mathfrak{M}' = (W', C', R', I', V')$ where

- $W = \mathbb{Z} \times \mathbb{Z}$, $Cxyz$ iff $x_0 = y_0$, $y_1 = z_0$ and $z_1 = x_1$, Rxy if $x_0 = y_1$ and $x_1 = y_0$, Ix iff $x_0 = x_1$, and finally, the valuation V is given by $V(p) = \{(x_0, x_1) \mid x_1 - x_0 \text{ is even }\}$,
- $W' = \mathbb{Z}$, $C'stu$ iff $s = t + u$, $R'st$ iff $s = -t$, $I's$ iff $s = 0$, and the valuation V' is given by $V'(p) = \{s \in \mathbb{Z} \mid s \text{ is even }\}$.

This example is best understood by looking at Figure 2.2. The left picture shows a fragment of the model $\mathfrak{M}$; the points of $\mathbb{Z} \times \mathbb{Z}$ are represented as disks or circles, depending on whether p is true or not. The diagonal is indicated by the dashed diagonal line. The picture on the right-hand side shows the image under f of the points in $\mathbb{Z} \times \mathbb{Z}$.

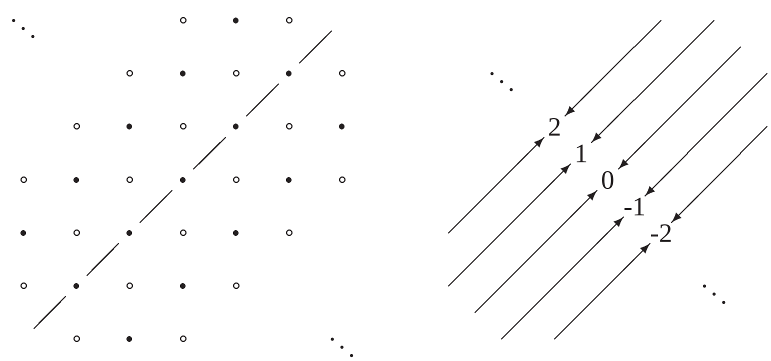


Fig. 2.2. Another bounded morphism.

We claim that the function $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ given by

$$f(z) = z_1 - z_0$$

is a bounded morphism for this similarity type. The clause for the propositional variables is trivial. For the unary relation I we only have to check that for any z in $\mathbb{Z} \times \mathbb{Z}$, $z_0 = z_1$ iff $z_1 - z_0 = 0$. This is obviously true. We leave the case of the binary relation R to the reader.

So let's turn to the clauses for the ternary relation C. To check item (ii)$'$ (the homomorphic condition), assume that $Cxyz$ holds for x, y and z in W. That is, we have that $x_0 = y_0$, $y_1 = z_0$ and $z_1 = x_1$. But then we find that

$$f(x) = x_1 - x_0 = z_1 - y_0 = z_1 - z_0 + y_1 - y_0 = f(z) + f(y),$$

so by definition of C' we do indeed find that $C'f(x)f(y)f(z)$.

For item (iii)′ (the back condition) assume that we have $C'f(x)tu$ for some $x \in \mathbb{Z} \times \mathbb{Z}$ and $t, u \in \mathbb{Z}$. In other words, we have that $x_1 - x_0 = t + u$. Consider the pairs $y := (x_0, x_0 + t)$ and $z := (x_0 + t, x_1)$. It is obvious that $Cxyz$; we also find that $f(y) = t$ and $f(z) = x_1 - (x_0 + t) = (x_1 - x_0) - t = u$. Hence y and z are the elements of W that we need to satisfy item (iii)′. $\dashv$

Definition 2.12 covers the basic temporal language, PDL, and arrow logic, as special cases — but once more it is worth issuing a warning concerning the basic temporal language. Although R_P is usually presented implicitly (as the converse of the relation R in some model (W, R, V)) we certainly cannot ignore it. Thus a *temporal* bounded morphism from (W_1, R_1, V_1) to (W_2, R_2, V_2) is a bounded morphism from $(W_1, R_1, \breve{R_1}, V_1)$ to $(W_2, R_2, \breve{R_2}, V_2)$.

Proposition 2.14 *Let τ be a modal similarity type and let $\mathfrak{M}$ and $\mathfrak{M}'$ be τ-models such that $f : \mathfrak{M} \to \mathfrak{M}'$. Then, for each modal formula ϕ, and each element w of $\mathfrak{M}$ we have $\mathfrak{M}, w \Vdash \phi$ iff $\mathfrak{M}', f(w) \Vdash \phi$. In words: modal satisfaction is invariant under bounded morphisms.*

Proof. Let $\mathfrak{M}$, $\mathfrak{M}'$ and f be as in the statement of the proposition. We will prove that for each formula ϕ and state w, $\mathfrak{M}, w \Vdash \phi$ iff $\mathfrak{M}', f(w) \Vdash \phi$. The proof is by induction on ϕ. We will assume that τ is the basic similarity type, leaving the general case to the reader.

The base step and the boolean cases are routine, so let's turn to the case where ϕ is of the form $\Diamond\psi$. Assume first that $\mathfrak{M}, w \Vdash \Diamond\psi$. This means there is a state v with Rwv and $\mathfrak{M}, v \Vdash \psi$. By the inductive hypothesis, $\mathfrak{M}', f(v) \Vdash \psi$. By the homomorphic condition, $R'f(w)f(v)$, so $\mathfrak{M}', f(w) \Vdash \Diamond\psi$.

For the other direction, assume that $\mathfrak{M}', f(w) \Vdash \Diamond\psi$. Thus there is a successor of $f(w)$ in $\mathfrak{M}'$, say v', such that $\mathfrak{M}', v' \Vdash \psi$. Now we use the back condition (of Definition 2.10). This yields a point v in $\mathfrak{M}$ such that Rwv and $f(v) = v'$. Applying the inductive hypothesis, we obtain $\mathfrak{M}, v \Vdash \psi$, so $\mathfrak{M}, w \Vdash \Diamond\psi$. $\dashv$

Here is a simple application: we will now show that any satisfiable formula can be satisfied in a *tree-like* model. To put it another way: modal logic has the *tree model property*.

Let τ be a modal similarity type containing only diamonds (thus if $\mathfrak{M}$ is a τ-model, it has the form $(W, R_1, R_2, \ldots, V)$, where each R_i is a binary relation on W). In this context we will call a τ-model $\mathfrak{M}$ *tree-like* if the structure $(W, \bigcup_i R_i, V)$ is a tree in the sense of Example 1.5.

Proposition 2.15 *Assume that τ is a modal similarity type containing only diamonds. Then, for any rooted τ-model $\mathfrak{M}$ there exists a tree-like τ-model $\mathfrak{M}'$ such that $\mathfrak{M}' \twoheadrightarrow \mathfrak{M}$. Hence any satisfiable τ-formula is satisfiable in a tree-like model.*

Proof. Let w be the root of $\mathfrak{M}$. Define the model $\mathfrak{M}'$ as follows. Its domain W' consists of all finite sequences $(w, u_1, \ldots, u_n)$ such that $n \geq 0$ and for some modal operators $\langle a_1 \rangle, \ldots, \langle a_n \rangle \in \tau$ there is a path $wR_{a_1}u_1 \cdots R a_n u_n$ in $\mathfrak{M}$. Define $(w, u_1, \ldots, u_n)R'_a(w, v_1, \ldots, v_m)$ to hold if $m = n + 1$, $u_i = v_i$ for $i = 1, \ldots, n$, and $R_a u_n v_m$ holds in $\mathfrak{M}$. That is, R'_a relates two sequences iff the second is an extension of the first with a state from $\mathfrak{M}$ that is a successor of the last element of the first sequence. Finally, V' is defined by putting $(w, u_1, \ldots, u_n) \in V'(p)$ iff $u_n \in V(p)$. As the reader is asked to check in Exercise 2.1.4, the mapping $f : (w, u_1 \ldots, u_n) \mapsto u_n$ defines a surjective bounded morphism from $\mathfrak{M}'$ to $\mathfrak{M}$, thus $\mathfrak{M}'$ and $\mathfrak{M}$ are equivalent.

But then it follows that any satisfiable τ-formula is satisfiable in a tree-like model. For suppose ϕ is satisfiable in some τ-model at a point w. Let $\mathfrak{M}$ be the submodel generated by w. By Proposition 2.3, $\mathfrak{M}, w \Vdash \phi$, and as $\mathfrak{M}$ is rooted we can form an equivalent tree-like model $\mathfrak{M}'$ as just described. ⊣

The method used to construct $\mathfrak{M}'$ from $\mathfrak{M}$ is well known in both modal logic and computer science: it is called *unravelling* (or *unwinding*, or *unfolding*). In essence, we built $\mathfrak{M}'$ by treating the paths through $\mathfrak{M}$ as first class citizens: this untangles the (possibly very complex) way information is stored in $\mathfrak{M}$, and makes it possible to present it as a tree. We will make use of unravelling several times in later work; in the meantime, Exercise 2.1.7 asks the reader to extend the notion of 'tree-likeness' to arbitrary modal similarity types, and generalize Proposition 2.15.

Exercises for Section 2.1

2.1.1 Suppose we wanted an operator D with the following satisfaction definition: for any model $\mathfrak{M}$ and any formula ϕ, $\mathfrak{M}, w \Vdash \mathrm{D}\phi$ iff there is a $u \neq w$ such that $\mathfrak{M}, u \Vdash \phi$. This operator is called the *difference operator* and we will discuss it further in Section 7.1. Is the difference operator definable in the basic modal language?

2.1.2 Use generated submodels to show that the backward looking modality (that is, the P of the basic temporal language) cannot be defined in terms of the forward looking operator $\Diamond$.

2.1.3 Give the simplest possible example which shows that the truth of modal formulas is *not* invariant under homomorphisms, even if condition 1 is strengthened to an equivalence. Is modal truth preserved under homomorphisms?

2.1.4 Show that the mapping f defined in the proof of Proposition 2.15 is indeed a surjective bounded morphism.

2.1.5 Let $\mathfrak{B} = (B, R)$ be the transitive binary tree; that is, B is the set of finite strings of 0s and 1s, and $R\sigma\tau$ holds if σ is a proper initial segment of τ. Let $\mathfrak{N} = (\mathbb{N}, <)$ be the frame of the natural numbers with the usual ordering.

(a) Let V_0 be the valuation on $\mathfrak{N}$ given by $V_0(p) = \{2n \mid n \in \mathbb{N}\}$ for each proposition letter p. Define a valuation U_0 on $\mathfrak{B}$ and a bounded morphism from $(\mathfrak{B}, U_0)$ to $(\mathfrak{N}, V_0)$.
(b) Let U_1 be the valuation on $\mathfrak{B}$ given by $U_1(p) = \{1\sigma \mid \sigma \in B\}$ for each proposition letter p. Give a valuation V_1 on $\mathfrak{N}$ and a bounded morphism from $(\mathfrak{B}, U_0)$ to $(\mathfrak{N}, V_0)$.
(c) Can you also find *surjective* bounded morphisms?

2.1.6 Show that every model is the bounded morphic image of the disjoint union of point-generated (that is: rooted) models. This exercise may look rather technical, but in fact it is very straightforward — think about it!

2.1.7 This exercise generalizes Proposition 2.15 to arbitrary modal similarity types.

(a) Define a suitable notion of tree-like model that works for arbitrary modal similarity types. (Hint: in case of $R_\triangle s_0 s_1 \dots s_n$, think of s_0 as being the parent node and of $s_1, \dots, s_n$ as the children.)
(b) Generalize Proposition 2.15 to arbitrary modal similarity types.

2.2 Bisimulations

What do the invariance results of the previous section have in common? They all deal with special sorts of *relations* between two models, namely relations with the following properties: related states carry identical atomic information, and whenever it is possible to make a transition in one model, it is possible to make a matching transition in the other. For example, with generated submodels the inter-model relation is identity, and every transition in one model is matched by an identical transition in the other. With bounded morphisms, the inter-model relation is a function, and the notion of matching involves both the homomorphic link from source to target, and the back condition which reflects target structure in the source.

This observation leads us to the central concept of the chapter: *bisimulations*. Quite simply, a bisimulation is a relation between two models in which related states have identical atomic information and matching transition possibilities. The interesting part of the definition is the way it makes the notion of 'matching transition possibilities' precise.

Definition 2.16 (Bisimulations — the Basic Case) We first give the definition for the basic modal language. Let $\mathfrak{M} = (W, R, V)$ and $\mathfrak{M}' = (W', R', V')$ be two models.

A non-empty binary relation $Z \subseteq W \times W'$ is called a *bisimulation between* $\mathfrak{M}$ and $\mathfrak{M}'$ (notation: $Z : \mathfrak{M} \leftrightarroweq \mathfrak{M}'$) if the following conditions are satisfied.

(i) If wZw' then w and w' satisfy the same proposition letters.
(ii) If wZw' and Rwv, then there exists v' (in $\mathfrak{M}'$) such that vZv' and $R'w'v'$ (the *forth condition*).

(iii) The converse of (ii): if wZw' and $R'w'v'$, then there exists v (in $\mathfrak{M}$) such that vZv' and Rwv (the *back condition*).

When Z is a bisimulation linking two states w in $\mathfrak{M}$ and w' in $\mathfrak{M}'$ we say that w and w' are *bisimilar*, and we write $Z : \mathfrak{M}, w \leftrightarroweq \mathfrak{M}', w'$. If there is a bisimulation Z such that $Z : \mathfrak{M}, w \leftrightarroweq \mathfrak{M}', w'$, we sometimes write $\mathfrak{M}, w \leftrightarroweq \mathfrak{M}', w'$; likewise, if there is some bisimulation between $\mathfrak{M}$ and $\mathfrak{M}'$, we write $\mathfrak{M} \leftrightarroweq \mathfrak{M}'$. ⊣

Think of Definition 2.16 pictorially. Figure 2.3 shows the content of the forth clause. Suppose we know that wZw' and Rwv (the solid arrow in $\mathfrak{M}$ and the Z-link at the bottom of the diagram display this information). Then the forth condition says that it is always possible to find a v' that 'completes the square' (this is shown by the dashed arrow in $\mathfrak{M}'$ and the dotted Z-link at the top of the diagram). Note the symmetry between the back and forth clauses: to visualize the back clause, simply reflect the picture through its vertical axis.

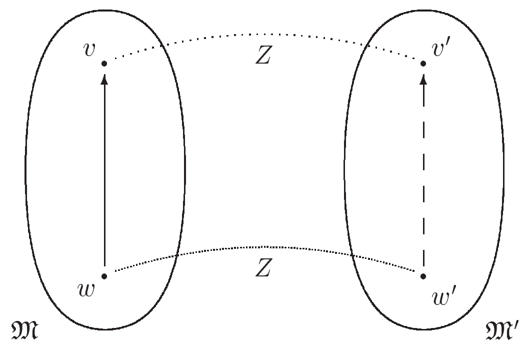


Fig. 2.3. The forth condition.

In effect, bisimulations are a relational generalization of bounded morphisms: we drop the directionality from source to target (and with it the homomorphic condition) and replace it with a back and forth system of matching moves between models.

Example 2.17 The models $\mathfrak{M}$ and $\mathfrak{M}'$ shown in Figure 2.4 are bisimilar. To see this, define the following relation Z between their states: $Z = \{(1, a), (2, b), (2, c), (3, d), (4, e), (5, e)\}$. Condition (i) of Definition 2.16 is obviously satisfied: Z-related states make the same propositional letters true. Moreover, the back-and-forth conditions are satisfied too: any move in $\mathfrak{M}$ can be matched by a similar move in $\mathfrak{M}'$, and conversely, as the reader should check.

This example also shows that bisimulation is a genuine generalization of the constructions discussed in the previous section. Although $\mathfrak{M}$ and $\mathfrak{M}'$ are bisimilar, neither is a generated submodel nor a bounded morphic image of the other. ⊣

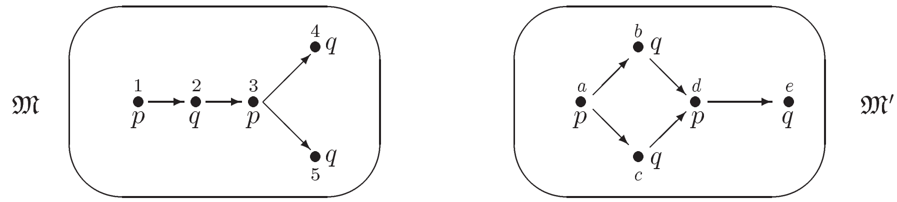


Fig. 2.4. Bisimilar models.

Definition 2.18 (Bisimulations — the General Case) Let τ be a modal similarity type, and let $\mathfrak{M} = (W, R_\triangle, V)_{\triangle\in\tau}$ and $\mathfrak{M}' = (W', R'_\triangle, V')_{\triangle\in\tau}$ be τ-models. A non-empty binary relation $Z \subseteq W \times W'$ is called a *bisimulation* between $\mathfrak{M}$ and $\mathfrak{M}'$ (notation: $Z : \mathfrak{M} \leftrightarroweq \mathfrak{M}'$) if the above condition (i) from Definition 2.16 is satisfied (that is, Z-related states satisfy the same proposition letters) and in addition the following conditions (ii)′ and (iii)′ are satisfied:

(ii)′ If wZw' and $R_\triangle wv_1 \ldots v_n$ then there are $v'_1, \ldots, v'_n$ (in W') such that $R'_\triangle w'v'_1 \ldots v'_n$ and for all i ($1 \leq i \leq n$) $v_iZv'_i$ (the *forth* condition).

(iii)′ The converse of (ii)′: if wZw' and $R_\triangle w'v'_1 \ldots v'_n$ then there are $v_1, \ldots, v_n$ (in W) such that $R_\triangle wv_1 \ldots v_n$ and for all i ($1 \leq i \leq n$) $v_iZv'_i$ (the *back* condition). ⊣

Examples of bisimulations abound — indeed, as we have already mentioned, the constructions of the previous section (disjoint unions, generated submodels, isomorphisms, and bounded morphisms), are all bisimulations:

Proposition 2.19 *Let τ be a modal similarity type, and let $\mathfrak{M}$, $\mathfrak{M}'$ and $\mathfrak{M}_i$ ($i \in I$) be τ-models.*

(i) *If $\mathfrak{M} \cong \mathfrak{M}'$, then $\mathfrak{M} \leftrightarroweq \mathfrak{M}'$.*
(ii) *For every $i \in I$ and every w in $\mathfrak{M}_i$, $\mathfrak{M}_i, w \leftrightarroweq \biguplus_i \mathfrak{M}_i, w$.*
(iii) *If $\mathfrak{M}' \rightarrowtail \mathfrak{M}$, then $\mathfrak{M}', w \leftrightarroweq \mathfrak{M}, w$ for all w in $\mathfrak{M}'$.*
(iv) *If $f : \mathfrak{M} \twoheadrightarrow \mathfrak{M}'$, then $\mathfrak{M}, w \leftrightarroweq \mathfrak{M}', f(w)$ for all w in $\mathfrak{M}$.*

Proof. We only prove the second item, leaving the others as Exercise 2.2.2. Assume we are working in the basic modal language. Define a relation Z between $\mathfrak{M}_i$ and $\biguplus_i \mathfrak{M}_i$ by putting $Z = \{(w, w) \mid w \in \mathfrak{M}_i\}$. Then Z is a bisimulation. To see this, observe that clause (i) of Definition 2.16 is trivially fulfilled, and as to clauses (ii) and (iii), any R-step in $\mathfrak{M}_i$ is reproduced in $\biguplus_i \mathfrak{M}_i$, and by the disjointness condition every R-step in $\biguplus_i \mathfrak{M}_i$ that departs from a point that was originally in $\mathfrak{M}_i$, stems from a corresponding R-step in $\mathfrak{M}_i$. The reader should extend this argument to arbitrary similarity types. ⊣

We will now show that modal satisfiability is invariant under bisimulations (and hence, by Proposition 2.19, provide an alternative proof that modal satisfiability is invariant under disjoint unions, generated submodels, isomorphisms, and bounded morphisms). The key thing to note about the following proof is how straightforward it is — the back and forth clauses in the definition of bisimulation are *precisely* what is needed to push the induction through.

Theorem 2.20 *Let τ be a modal similarity type, and let $\mathfrak{M}$, $\mathfrak{M}'$ be τ-models. Then, for every $w \in W$ and $w' \in W'$, $w \leftrightarroweq w'$ implies that $w \leftrightsquigarrow w'$. In words, modal formulas are invariant under bisimulation.*

Proof. By induction on ϕ. The case where ϕ is a proposition letter follows from clause (i) of Definition 2.16, and the case where ϕ is $\bot$ is immediate. The boolean cases are immediate from the induction hypothesis.

As for formulas of the form $\Diamond\psi$, we have $\mathfrak{M}, w \Vdash \Diamond\psi$ iff there exists a v in $\mathfrak{M}$ such that Rwv and $\mathfrak{M}, v \Vdash \psi$. As $w \leftrightarroweq w'$ we find by clause (ii) of Definition 2.16 that there exists a v' in $\mathfrak{M}'$ such that $R'w'v'$ and $v \leftrightarroweq v'$. By the induction hypothesis, $\mathfrak{M}', v' \Vdash \psi$, hence $\mathfrak{M}', w' \Vdash \Diamond\psi$. For the converse direction use clause (iii) of Definition 2.16.

The argument for the general modal case, with triangles $\triangle$, is an easy extension of that just given, as the reader should check. $\dashv$

This finishes our discussion of the basics of bisimulation — so let's now try and understand the concept more deeply. Some of the remarks that follow are conceptual, and some are technical, but they all point to ideas that crop up throughout the book.

Remark 2.21 (Bisimulation, Locality, and Computation) In the Preface we suggested that the reader think of modal formulas as automata. Evaluating a modal formula amounts to running an automaton: we place it at some state inside a structure and let it search for information. The automaton is only permitted to explore by making transitions to neighboring states; that is, it works locally.

Suppose such an automaton is standing at a state w in a model $\mathfrak{M}$, and we pick it up and place it at a state w' in a different model $\mathfrak{M}'$; would it notice the switch? If w and w' are bisimilar, *no*. Our automaton cares only about the information at the current state and the information accessible by making a transition — it is indifferent to everything else. Thus the definition of bisimulation spells out exactly what we have to do if we want to fool such an automaton as to where it is being evaluated. Viewed this way, it is clear that the concept of bisimulation is a direct reflection of the locality of the modal satisfaction definition.

But there is a deeper link between bisimulation and computation than our informal talk of automaton might suggest. As we discussed in Example 1.3, labelled

Fig. 2.5. Equivalent but not bisimilar.

transition systems (LTSs) are a standard way of thinking about computation: when we traverse an LTS we build a sequence of state transitions — or to put it another way, we compute. When are two LTSs computationally equivalent? More precisely, if we ignore practical issues (such as how long it takes to actually perform a computation) when can two different LTSs be treated as freely exchangeable (‘observationally equivalent’) black boxes? One natural answer is: when they are bisimilar. Bisimulation turns out to be a very natural notion of equivalence for both mathematical and computational investigations. For more on the history of bisimulation and the connection with computer science, see the Notes. $\dashv$

Remark 2.22 (Bisimulation and First-Order Logic) According to Theorem 2.20 modal formulas cannot distinguish between bisimilar states or between bisimilar models, even though these states or models may be quite different. It follows that modal logic is very different from first-order logic, for arbitrary first-order formulas are certainly *not* invariant under bisimulations. For example, the model $\mathfrak{M}'$ of Example 2.17 satisfies the formula

$$\exists y_1 y_2 y_3\,(y_1 \neq y_2 \wedge y_1 \neq y_3 \wedge y_2 \neq y_3 \wedge Rxy_1 \wedge Rxy_2 \wedge Ry_1y_3 \wedge Ry_2y_3),$$

if we assign the state a to the free variable x. This formula says that there is a diamond-shaped configuration of points, which is true of the point a in $\mathfrak{M}'$, but not of the state 1 in $\mathfrak{M}$. But as far as modal logic is concerned, $\mathfrak{M}'$ and $\mathfrak{M}$, being bisimilar, are indistinguishable. In Section 2.4 we will start examining the links between modal logic and first-order logic more systematically. $\dashv$

Now for a fundamental question: is the converse of Theorem 2.20 true? That is, if two models are modally equivalent, must they be bisimilar? The answer is *no*.

Example 2.23 Consider the basic modal language. We may just as well work with an empty set of proposition letters here. Define models $\mathfrak{M}$ and $\mathfrak{N}$ as in Figure 2.5, where arrows denote R-transitions. Each of $\mathfrak{M}$ and $\mathfrak{N}$ has, for each $n > 0$, a finite branch of length n; the difference between the models is that, in addition, $\mathfrak{N}$ has an infinite branch.

One can show that for all modal formulas ϕ, $\mathfrak{M}, w \Vdash \phi$ iff $\mathfrak{N}, w' \Vdash \phi$ (this is easy if one is allowed to use some results that we will prove further on, namely Propositions 2.31 and 2.33, but it is not particularly hard to prove from first principles, and the reader may like to try this). But even though w and w' are modally equivalent, there is no bisimulation linking them. To see this, suppose that there was such a bisimulation Z: we will derive a contradiction from this supposition.

Since w and w' are linked by Z, there has to be a successor of w, say v_0, which is linked to the first point v'_0 on the infinite path from w'. Suppose that n is the length of the (maximal) path leading from w through v_0, and let w, v_0, …, v_{n-1} be the successive points on this path. Using the bisimulation conditions $n - 1$ times, we find points v'_1, …, v'_{n-1} on the infinite path emanating from w', such that $v'_0 R' v'_1 \dots R' v'_{n-1}$ and $v_i Z v'_i$ for each i. Now v'_{n-1} has a successor, but v_{n-1} does not; hence, there is no way that these two points can be bisimilar. ⊣

Nonetheless, it is possible to prove a restricted converse to Theorem 2.20, namely the Hennessy-Milner Theorem. Let τ be a modal similarity type, and $\mathfrak{M}$ a τ-model. $\mathfrak{M}$ is *image-finite* if for each state u in $\mathfrak{M}$ and each relation R in $\mathfrak{M}$, the set $\{ (v_1, \dots, v_n) \mid Ruv_1 \dots v_n \}$ is finite; observe that we are *not* putting any restrictions on the total number of different relations R in the model $\mathfrak{M}$ — just that each of them is image-finite.

Theorem 2.24 (Hennessy-Milner Theorem) *Let τ be a modal similarity type, and let $\mathfrak{M}$ and $\mathfrak{M}'$ be two image-finite τ-models. Then, for every $w \in W$ and $w' \in W'$, $w \leftrightarrows w'$ iff $w \leftrightsquigarrow w'$.*

Proof. Assume that our similarity type τ only contains a single diamond (that is, we will work in the basic modal language). The direction from left to right follows from Theorem 2.20; for the other direction, we will prove that the relation $\leftrightsquigarrow$ of modal equivalence itself satisfies the conditions of Definition 2.16 — that is, we show that the relation of modal equivalence on these models is itself a bisimulation. (This is an important idea; we will return to it in Section 2.5.)

The first condition is immediate. For the second one, assume that $w \leftrightsquigarrow w'$ and Rwv. We will try to arrive at a contradiction by assuming that there is no v' in $\mathfrak{M}'$ with $R'w'v'$ and $v \leftrightsquigarrow v'$. Let $S' = \{u' \mid R'w'u'\}$. Note that S' must be non-empty, for otherwise $\mathfrak{M}', w' \Vdash \Box\bot$, which would contradict $w \leftrightsquigarrow w'$ since $\mathfrak{M}, w \Vdash \Diamond\top$. Furthermore, as $\mathfrak{M}'$ is image-finite, S' must be finite, say $S' = \{w'_1, \dots, w'_n\}$. By assumption, for every $w'_i \in S'$ there exists a formula ψ_i such that $\mathfrak{M}, v \Vdash \psi_i$ but $\mathfrak{M}', w'_i \nVdash \psi_i$. It follows that

$$\mathfrak{M}, w \Vdash \Diamond(\psi_1 \wedge \cdots \wedge \psi_n) \text{ and } \mathfrak{M}', w' \nVdash \Diamond(\psi_1 \wedge \cdots \wedge \psi_n),$$

which contradicts our assumption that $w \leftrightsquigarrow w'$. The third condition of Definition 2.16 may be checked in a similar way. Extending the proof to other similarity types is routine. ⊣

Theorem 2.20 (together with the Hennessy-Milner Theorem) on the one hand, and Example 2.23 on the other, mark important boundaries. Clearly, bisimulations have something important to say about modal expressivity over models, but they don't tell us everything. Two pieces of the jigsaw puzzle are missing. For a start, we are still considering modal languages in isolation: as yet, we have made no attempt to systematically link them to first-order logic. We will remedy this in Section 2.4 and this will eventually lead us to a beautiful result, the Van Benthem Characterization Theorem (Theorem 2.68): modal logic is the bisimulation invariant fragment of first-order logic.

The second missing piece is the notion of an *ultrafilter extension*. We will introduce this concept in Section 2.5, and this will eventually lead us to Theorem 2.62. Informally, this theorem says: modal equivalence implies bisimilarity-somewhere-else. Where is this mysterious 'somewhere else'? In the ultrafilter extension. As we will see, although modally equivalent models need not be bisimilar, they must have bisimilar ultrafilter extensions.

Remark 2.25 (Bisimulations for the Basic Temporal Language, PDL, and Arrow Logic) Although we have already said the most fundamental things that need to be said on this topic (Definition 2.18 and Theorem 2.20 covers these languages), a closer look reveals some interesting results for PDL and arrow logic. But let us first discuss the basic temporal language.

First we issue our (by now customary) warning. When working with the basic temporal language, we usually work with models (W, R, V) and implicitly take R_P to be $R^{\smile}$. Thus we need a notion of bisimulation which takes $R^{\smile}$ into account, and so we define a *temporal* bisimulation between models (W, R, V) and (W', R', V') to be a relation Z between the states of the two models that satisfies the clauses of Definition 2.16, and in addition the following two clauses (iv) and (v) requiring that backward steps in one model should be matched by similar steps in the other model:

(iv) If wZw' and Rvw, then there exists v' (in $\mathfrak{M}'$) such that vZv' and $R'v'w'$.

(v) The converse of (iv): if wZw' and $R'v'w'$, then there exists v (in $\mathfrak{M}$) such that vZv' and Rvw.

If we don't do this, we are in trouble. For example, if $\mathfrak{M}$ is a model whose underlying frame is the integers, and $\mathfrak{M}'$ is the submodel of $\mathfrak{M}$ generated by 0, then these two models are bisimilar in the sense of Definition 2.16, and hence equivalent as far as the basic *modal* language is concerned. But they are not equivalent as far as the basic *temporal* language is concerned: $\mathfrak{M}, 0 \Vdash P\top$, but $\mathfrak{M}, 0 \nVdash P\top$.

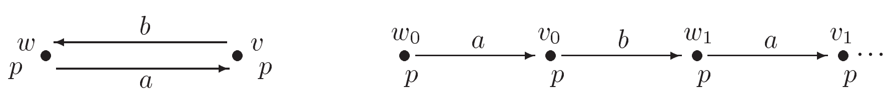
Given our previous discussion, this is unsurprising. What is (pleasantly) surprising is that things do not work this way in PDL. Suppose we are given two regular models. Checking that these models are bisimilar for the language of PDL means checking that bisimilarity holds for all the (infinitely many) relations that exist in regular models (see Definition 1.26). But as it turns out, most of this work is unnecessary. Once we have checked that bisimilarity holds for all the relations which interpret the basic programs, we don't have to check anything else: the relations corresponding to complex programs will *automatically* be bisimilar. In Section 2.7 we will introduce some special terminology to describe this: the operations in regular PDL's modality building repertoire ($\cup$, ;, and $*$) will be called *safe for bisimulation*. Note that taking the converse of a relation is *not* an operation that is safe for bisimulation (in effect, that's what we just noted when discussing the basic temporal language); see Exercise 2.2.6.

What about arrow logic? The required notion of bisimulation is given by Definition 2.18; note that the clause for 1' reads that for bisimilar points a and a' we have Ia iff $I'a$. $\dashv$

Remark 2.26 (The Algebra of Bisimulations) Bisimulations give rise to algebraic structure quite naturally. For instance, if Z_0 is a bisimulation between $\mathfrak{M}_0$ and $\mathfrak{M}_1$, and Z_1 a bisimulation between $\mathfrak{M}_1$ and $\mathfrak{M}_2$, then the composition of Z_0 and Z_1 is a bisimulation linking $\mathfrak{M}_0$ and $\mathfrak{M}_2$. It is also a rather easy observation that the set of bisimulations between two models is closed under taking arbitrary (finite or infinite) unions. This shows that if two points are bisimilar, there is always a *maximal* bisimulation linking them; see Exercise 2.2.8. Further information on closure properties of the set of bisimulations between two models can be found in Section 2.7. $\dashv$

Exercises for Section 2.2

2.2.1 Consider a modal similarity type with two diamonds $\langle a \rangle$ and $\langle b \rangle$, and with $\Phi = \{p\}$. Show that the following two models are bisimilar.



2.2.2 This exercise asks the reader to complete in detail the proof of Proposition 2.19, which links bisimulations and the model constructions discussed in the previous section. You should prove these results for arbitrary similarity types.

(a) Show that if $\mathfrak{M} \cong \mathfrak{M}'$, then $\mathfrak{M} \underline{\leftrightarrow} \mathfrak{M}'$
(b) Show that if $\biguplus_i \mathfrak{M}_i$ is the disjoint union of the models $\mathfrak{M}_i$ ($i \in I$), then, for each i, $\mathfrak{M}_i \underline{\leftrightarrow} \biguplus_i \mathfrak{M}_i$
(c) Show that if $\mathfrak{M}'$ is a generated submodel of $\mathfrak{M}$, then $\mathfrak{M}' \underline{\leftrightarrow} \mathfrak{M}$
(d) Show that if $\mathfrak{M}'$ is a bounded morphic image of $\mathfrak{M}$, then $\mathfrak{M}' \underline{\leftrightarrow} \mathfrak{M}$

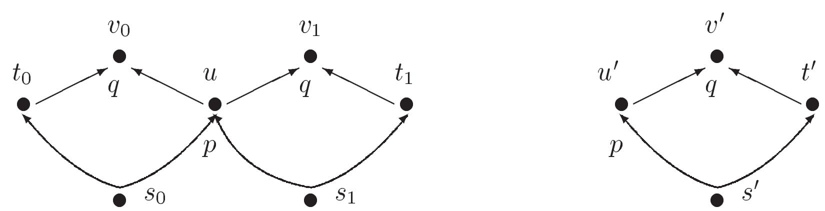
2.2.3 This exercise is about *temporal* bisimulations.

(a) Show *from first principles* that the truth of basic temporal formulas is invariant under temporal bisimulations. (That is, don't appeal to any of the results proved in this section.)
(b) Let $\mathfrak{M}$ and $\mathfrak{M}'$ be *finite* rooted models for basic temporal logic with F and P. Let w and w' be the roots of $\mathfrak{M}$ and $\mathfrak{M}'$, respectively. Prove that if w and w' satisfy the same basic temporal formulas with F and P, then there exists a basic temporal bisimulation that relates w and w'.

2.2.4 Consider the binary until operator U. In a model $\mathfrak{M} = (W, R, V)$ its truth definition reads:

$$\mathfrak{M}, t \Vdash U(\phi, \psi) \quad \text{iff} \quad \text{there is a } v \text{ such that } Rtv \text{ and } v \Vdash \phi\text{, and for all } u \text{ such that } Rtu \text{ and } Ruv\text{: } u \Vdash \psi.$$

Prove that U is not definable in the basic modal language. Hint: think about the following two models, but with arrows added to make sure that the relations are transitive:



2.2.5 Consider the following two models, which we are going to use to interpret the basic *temporal* language: $\mathfrak{M}_0 = (\mathbb{R}, <, V_0)$ and $\mathfrak{M}_1 = (\mathbb{R}, <, V_1)$, where V_0 makes q true at all non-zero integers and V_1 in addition makes q true at all points of the form $1/z$ with z a non-zero integer number.

(a) Prove that there is a temporal bisimulation between $\mathfrak{M}_0$ and $\mathfrak{M}_1$, linking 0 (in the one model) to 0 (in the other model).
(b) Let Π be the *progressive* operator defined by the following truth table:

$$\mathfrak{M}, s \Vdash \Pi\phi \quad \text{iff} \quad \text{there are } t \text{ and } u \text{ such that } t < s < u \text{ and } \mathfrak{M}, x \Vdash \phi \text{ for all } x \text{ between } t \text{ and } u.$$

Prove that this operator is not definable in the basic temporal language.

2.2.6 Suppose we have two bisimilar LTSs. Show that bisimilar states in these LTSs satisfy exactly the same formulas of PDL.

2.2.7 Prove that two square arrow models $\mathfrak{M} = (\mathfrak{S}_U, V)$ and $\mathfrak{M}' = (\mathfrak{S}_{U'}, V')$ are bisimilar if and only if there is a relation Z between *pairs* over U and *pairs* over U' such that

(i) if $(u, v)Z(u', v')$, then $(u, v) \in V(p)$ iff $(u', v') \in V'(p)$,
(ii) if $(u, v)Z(u', v')$, then $u = v$ iff $u' = v'$,
(iii) if $(u, v)Z(u', v')$, then $(v, u)Z(v', u')$,
(iv) if $(u, v)Z(u', v')$, then for any $w \in U$ there exists a $w' \in U'$ such that both $(u, w)Z(u', w')$ and $(w, v)Z(w', v')$,
(v) and vice versa.

Must any two bisimilar square arrow models be isomorphic? (Hint: think of $V(p)$ and $V'(p)$ as the natural ordering relations of the rational and the real numbers, respectively.)

2.2.8 Suppose that $\{Z_i \mid i \in I\}$ is a non-empty collection of bisimulations between $\mathfrak{M}$ and $\mathfrak{M}'$. Prove that the relation $\bigcup_{i \in I} Z_i$ is also a bisimulation between $\mathfrak{M}$ and $\mathfrak{M}'$. Conclude that if $\mathfrak{M}$ and $\mathfrak{M}'$ are bisimilar, then there is a maximal bisimulation between $\mathfrak{M}$ and $\mathfrak{M}'$; that is, a bisimulation Z_m such that for any bisimulation $Z : \mathfrak{M} \leftrightarrows \mathfrak{M}'$ we have $Z \subseteq Z_m$.

2.3 Finite Models

Preservation and invariance results can be viewed either positively or negatively. Viewed negatively, they map the limits of modal expressivity: they tell us, for example, that modal languages are incapable of distinguishing a model from its generated submodels. Viewed positively, they are a toolkit for transforming models into more desirable forms without affecting satisfiability. Proposition 2.15 has already given us a taste of this perspective (we showed that modal languages have the tree model property) and it will play an important role when we discuss completeness in Chapter 4.

The results of this section are similarly double-edged. We are going to investigate modal expressivity over finite models, and the basic result we will prove is that modal languages have the *finite model property*: if a modal formula is satisfiable on an arbitrary model, then it is satisfiable on a finite model.

Definition 2.27 (Finite Model Property) Let τ be a modal similarity type, and let M be a class of τ-models. We say that τ has the *finite model property with respect to* M if the following holds: if ϕ is a formula of similarity type τ, and ϕ is satisfiable in some model in M, then ϕ is satisfiable in a *finite* model in M. $\dashv$

In this section we will mostly be concerned with the special case in which M in Definition 2.27 is the collection of *all* τ-models, so to simplify terminology we will use the term 'finite model property' for this special case. The fact that modal languages have the finite model property (in this sense) can be viewed as a limitative result: modal languages simply lack the expressive strength to force the existence of infinite models. (By way of contrast, it is easy to write down first-order formulas which can only be satisfied on infinite models.) On the other hand, the result is a source of strength: we do not need to bother about (arbitrary) infinite models, for we can always find an equivalent finite one. This opens the door to the decidability results of Chapter 6. (The satisfiability problem for first-order logic, as the reader probably knows, is undecidable over arbitrary models.)

We will discuss two methods for building finite models for satisfiable modal formulas. The first is to (carefully!) select a finite *submodel* of the satisfying model, the second (called the filtration method) is to define a suitable *quotient* structure.

Selecting a finite submodel

The selection method draws together four observations. Here is the first. We know that modal satisfaction is intrinsically *local*: modalities scan the states accessible from the current state. How much of the model can a modal formula see from the current state? That obviously depends on how deeply the modalities it contains are nested.

Definition 2.28 (Degree) We define the *degree* of modal formulas as follows.

$$\begin{aligned}\deg(p) &= 0\\ \deg(\bot) &= 0\\ \deg(\neg\phi) &= \deg(\phi)\\ \deg(\phi\vee\psi) &= \max\{\deg(\phi),\deg(\psi)\}\\ \deg(\triangle(\phi_1,\ldots,\phi_n)) &= 1+\max\{\deg(\phi_1),\ldots,\deg(\phi_n)\}.\end{aligned}$$

In particular, the degree of a basic modal formula $\Diamond\phi$ is $1+\deg(\phi)$. ⊣

Second, we observe the following:

Proposition 2.29 *Let τ be a finite modal similarity type, and assume that our collection of proposition letters is finite as well.*

- (i) *For all n, up to logical equivalence there are only finitely many formulas of degree at most n.*
- (ii) *For all n, and every τ-model $\mathfrak{M}$ and state w of $\mathfrak{M}$, the set of all τ-formulas of degree at most n that are satisfied by w, is equivalent to a single formula.*

Proof. We prove the first item by induction on n. The case $n = 0$ is obvious. As for the case $n+1$, observe that every formula of degree $\leq n+1$ is a boolean combination of proposition letters and formulas of the form $\Diamond\psi$, where $\deg(\psi) \leq n$. By the induction hypothesis there can only be finitely many non-equivalent such formulas ψ. Thus there are only finitely many non-equivalent boolean combinations of proposition letters and formulas $\Diamond\psi$, where ψ has degree at most n. Hence, there are only finitely many non-equivalent formulas of degree at most $n + 1$.

Item (ii) is immediate from item (i). ⊣

Third, we observe that there is a natural way of finitely approximating a bisimulation. These finite approximations will prove crucial in our search for finite models.

Definition 2.30 (n-Bisimulations) Here we define n-bisimulations for modal similarity types containing only diamonds, leaving the definition of the general case as part of Exercise 2.3.2. Let $\mathfrak{M}$ and $\mathfrak{M}'$ be models, and let w and w' be states of $\mathfrak{M}$ and $\mathfrak{M}'$, respectively. We say that w and w' are n-*bisimilar* (notation:

$w \leftrightarroows_n w'$) if there exists a sequence of binary relations $Z_n \subseteq \cdots \subseteq Z_0$ with the following properties (for $i + 1 \leq n$):

(i) wZ_nw'
(ii) If vZ_0v' then v and v' agree on all proposition letters;
(iii) If $vZ_{i+1}v'$ and Rvu, then there exists u' with $R'v'u'$ and uZ_iu';
(iv) If $vZ_{i+1}v'$ and $R'v'u'$, then there exists u with Rvu and uZ_iu'. $\dashv$

The intuition is that if $w \leftrightarroows_n w'$, then w and w' *bisimulate up to depth* n. Clearly, if $w \leftrightarroows w'$, then $w \leftrightarroows_n w'$ for all n — but the converse need not hold; see Exercise 2.3.1.

Fourth, we observe that for languages containing only finitely many proposition letters, there is an *exact* match between modal equivalence and n-bisimilarity for all n. That is, for such languages not only does n-bisimilarity for all n imply modal equivalence, but the converse holds as well.

Proposition 2.31 *Let τ be a finite modal similarity type, Φ a finite set of proposition letters, and let $\mathfrak{M}$ and $\mathfrak{M}'$ be models for this language. Then for every w in $\mathfrak{M}$ and w' in $\mathfrak{M}'$, the following are equivalent.*

(i) $w \leftrightarroows_n w'$
(ii) *w and w' agree on all modal formulas of degree at most n.*

It follows that 'n-bisimilarity for all n' and modal equivalence coincide as relations between states.

Proof. The implication (i) $\Rightarrow$ (ii) may be proved by induction on n. For the converse implication one can use an argument similar to the one used in the proof of Theorem 2.24; we leave the proof as part of Exercise 2.3.2. $\dashv$

It is time to draw these observations together. The following definition and lemma, which are about *rooted* models, give us half of what we need to build finite models.

Definition 2.32 Let τ be a modal similarity type containing only diamonds. Let $\mathfrak{M} = (W, R_1, \ldots, R_n, \ldots)$ be a rooted τ-model with root w. The notion of the *height* of states in $\mathfrak{M}$ is defined by induction. The only element of height 0 is the root of the model; the states of height $n + 1$ are those immediate successors of elements of height n that have not yet been assigned a height smaller than $n + 1$. The *height of a model* $\mathfrak{M}$ is the maximum n such that there is a state of height n in $\mathfrak{M}$, if such a maximum exists; otherwise the height of $\mathfrak{M}$ is infinite.

For a natural number k, the *restriction* of $\mathfrak{M}$ to k (notation: $\mathfrak{M} \upharpoonright k$) is defined as the submodel containing only states whose height is at most k. More precisely, $(\mathfrak{M} \upharpoonright k) = (W_k, R_{1k}, \ldots, R_{nk}, \ldots, V_k)$, where $W_k = \{v \mid \text{height}(v) \leq k\}$, $R_{nk} = R_n \cap (W_k \times W_k)$, and for each p, $V_k(p) = V(p) \cap W_k$. $\dashv$

In words: the restriction of $\mathfrak{M}$ to k contains all states that can be reached from the root in at most k steps along the accessibility relations. Typically, this will not give a *generated* submodel, so why does it interest us? Because, as we can now show, given a formula ϕ of degree k that is satisfiable in some rooted model $\mathfrak{M}$, the restriction of $\mathfrak{M}$ to k contains all the states we need to satisfy ϕ. To put it another way: we are free to simply delete all states that lie beyond the 'k-horizon.'

Lemma 2.33 *Let τ be a modal similarity type that contains only diamonds. Let $\mathfrak{M}$ be a rooted τ-model, and let k be a natural number. Then, for every state w of $(\mathfrak{M} \upharpoonright k)$, we have $(\mathfrak{M} \upharpoonright k), w \leftrightarroweq_l \mathfrak{M}, w$, where $l = k - \mathrm{height}(w)$.*

Proof. Take the identity relation on $(\mathfrak{M} \upharpoonright k)$. We leave the reader to work out the details as Exercise 2.3.3. The following comment may be helpful: in essence this lemma tells us that if we are only interested in the satisfiability of modal formulas of degree at most k, then generating submodels of height k suffices to maintain satisfiability. $\dashv$

Putting together Proposition 2.31 and Lemma 2.33, we conclude that every satisfiable modal formula can be satisfied on a model of finite *height*. This is clearly useful, but we are only halfway to our goal: the resulting model may still be infinite, as it may be infinitely branching. We obtain the finite model we are looking for by a further selection of points; in effect this discards unwanted branches and leads to the desired finite model.

Theorem 2.34 (Finite Model Property — via Selection) *Let τ be a modal similarity type containing only diamonds, and let ϕ be a τ-formula. If ϕ is satisfiable, then it is satisfiable on a finite model.*

Proof. Fix a modal formula ϕ with $\deg(\phi) = k$. We restrict our modal similarity type τ and our collection of proposition letters to the modal operators and proposition letters actually occurring in ϕ. Let $\mathfrak{M}_1, w_1$ be such that $\mathfrak{M}_1, w_1 \Vdash \phi$. By Proposition 2.15, there exists a tree-like model $\mathfrak{M}_2$ with root w_2 such that $\mathfrak{M}_2, w_2 \Vdash \phi$. Let $\mathfrak{M}_3 := (\mathfrak{M}_2 \upharpoonright k)$. By Lemma 2.33 we have $\mathfrak{M}_2, w_2 \leftrightarroweq_k \mathfrak{M}_3, w_2$, and by Proposition 2.31 it follows that $\mathfrak{M}_3, w_2 \Vdash \phi$.

By induction on $n \leq k$ we define finite sets of states $S_0, \ldots, S_k$ and a (final) model $\mathfrak{M}_4$ with domain $S_0 \cup \cdots \cup S_k$; the points in each S_n will have height n.

Define S_0 to be the singleton $\{w_2\}$. Next, assume that $S_0, \ldots, S_n$ have already been defined. Fix an element v of S_n. By Proposition 2.29 there are only finitely many non-equivalent modal formulas whose degree is at most k, say $\psi_1, \ldots, \psi_m$. For each such formula that is of the form $\langle a \rangle \chi$ and holds in $\mathfrak{M}_3$ at v, select a state u from $\mathfrak{M}_3$ such that $R_a vu$ and $\mathfrak{M}_3, u \Vdash \chi$. Add all these us to S_{n+1}, and repeat this selection process for every state in S_n. S_{n+1} is defined as the set of all points that have been selected in this way.

Finally, define $\mathfrak{M}_4$ as follows. Its domain is $S_0 \cup \cdots \cup S_k$; as each S_i is finite, $\mathfrak{M}_4$ is finite. The relations and valuation are obtained by restricting the relations and valuation of $\mathfrak{M}_3$ to the domain of $\mathfrak{M}_4$. By Exercise 2.3.4 we have that $\mathfrak{M}_4, w_2 \leftrightarrows_k \mathfrak{M}_3, w_2$, and hence $\mathfrak{M}_4, w_2 \Vdash \phi$, as required. $\dashv$

How well does the selection method generalize to other modal languages? For certain purposes it is fine. For example, to deal with arbitrary modal similarity types, the notion of a tree-like model needs to be adapted (in fact, we explained how to do this in Exercise 2.1.7), but once this has been done we can prove a general version of Proposition 2.15. Next, the notion of n-bisimilarity needs to be adapted to other similarity types, but that too is straightforward (it is part of Exercise 2.3.2). Finally, the selection process in the proof of Theorem 2.34 needs adaptation, but this is unproblematic. In short, we can show that the finite model property holds for arbitrary similarity types using the selection method.

The method has a drawback: the input model for our construction may satisfy important relational properties (such as being symmetric), but the end result is always a finite tree-like model, and the desired relational properties may be (and often are) lost. So if we want to establish the finite model property with respect to a class of models satisfying additional properties — something that is very important in practice — we may have to do additional work once we have obtained our finite tree-like model. In such cases, the selection method tends to be harder to use than the filtration method (which we discuss next). Nonetheless, the idea of (intelligently!) selecting points to build submodels is important, and (as we will see in Section 6.6 when we discuss NP-completeness) the idea really comes into its own when the model we start with is already finite.

Finite models via filtrations

We now examine the classic modal method for building finite models: filtration. Whereas the selection method builds finite models by *deleting* superfluous material from large, possibly infinite models, the filtration method produces finite models by taking a large, possibly infinite model and *identifying* as many states as possible. We first present the filtration method for the basic modal language.

Definition 2.35 A set of formulas Σ is *closed under subformulas* (or: *subformula closed*) if for all formulas ϕ, ϕ': if $\phi \vee \phi' \in \Sigma$ then so are ϕ and ϕ'; if $\neg\phi \in \Sigma$ then so is ϕ; and if $\triangle(\phi_1, \ldots, \phi_n) \in \Sigma$ then so are $\phi_1, \ldots, \phi_n$. (For the basic modal language, this means that if $\Diamond\phi \in \Sigma$, then so is ϕ.) $\dashv$

Definition 2.36 (Filtrations) We work in the basic modal language. Let $\mathfrak{M} = (W, R, V)$ be a model and Σ a subformula closed set of formulas. Let $\leftrightsquigarrow_\Sigma$ be the

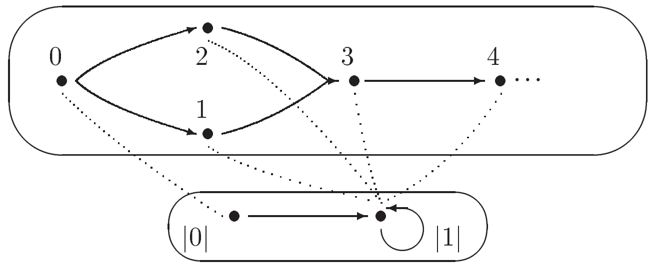


Fig. 2.6. A model and its filtration

relation on the states of $\mathfrak{M}$ defined by:

$$w \leftrightsquigarrow_{\Sigma} v \text{ iff for all } \phi \text{ in } \Sigma\text{: } (\mathfrak{M}, w \Vdash \phi \text{ iff } \mathfrak{M}, v \Vdash \phi).$$

Note that $\leftrightsquigarrow_{\Sigma}$ is an equivalence relation. We denote the equivalence class of a state w of $\mathfrak{M}$ with respect to $\leftrightsquigarrow_{\Sigma}$ by $|w|_{\Sigma}$, or simply by $|w|$ if no confusion will arise. The mapping $w \mapsto |w|$ that sends a state to its equivalence class is called the *natural map*.

Let $W_{\Sigma} = \{|w|_{\Sigma} \mid w \in W\}$. Suppose $\mathfrak{M}^f_{\Sigma}$ is any model (W^f, R^f, V^f) such that:

(i) $W^f = W_{\Sigma}$.
(ii) If Rwv then $R^f|w||v|$.
(iii) If $R^f|w||v|$ then for all $\Diamond\phi \in \Sigma$, if $\mathfrak{M}, v \Vdash \phi$ then $\mathfrak{M}, w \Vdash \Diamond\phi$.
(iv) $V^f(p) = \{|w| \mid \mathfrak{M}, w \Vdash p\}$, for all proposition letters p in Σ.

Then $\mathfrak{M}^f_{\Sigma}$ is called a *filtration of* $\mathfrak{M}$ *through* Σ. $\dashv$

Because of item (ii), the natural map associated with any filtration is guaranteed to be a homomorphism (see Definition 2.7). And at first glance it may seem that it is even guaranteed to be a bounded morphism (see Definition 2.10), for item (iii) seems reminiscent of the back condition. Unfortunately, this is *not* the case, as the following example shows.

Example 2.37 Let $\mathfrak{M}$ be the model $(\mathbb{N}, R, V)$, where $R = \{(0,1), (0,2), (1,3)\} \cup \{(n, n+1) \mid n \geq 2\}$, and V has $V(p) = \mathbb{N} \setminus \{0\}$ and $V(q) = \{2\}$.

Further, assume that $\Sigma = \{\Diamond p, p\}$. Clearly Σ is subformula closed. Then, the model $\mathfrak{N} = (\{|0|, |1|\}, \{(|0|, |1|), (|1|, |1|)\}, V')$, where $V'(p) = \{|1|\}$, is a filtration of $\mathfrak{M}$ through Σ. See Figure 2.6.

Clearly, $\mathfrak{N}$ can *not* be a bounded morphic image of $\mathfrak{M}$: any bounded morphism would have to preserve the formula q, and the natural map does not preserve q, and need not, because q is not an element of our subformula closed set Σ. $\dashv$

But in many other respects filtrations are well-behaved. For a start, the method gives us a bound (albeit an exponential one) on the size of the resulting finite model:

Proposition 2.38 *Let Σ be a finite subformula closed set of basic modal formulas. For any model $\mathfrak{M}$, if $\mathfrak{M}^f$ is a filtration of $\mathfrak{M}$ through a subformula closed set Σ, then $\mathfrak{M}^f$ contains at most $2^{\mathrm{card}(\Sigma)}$ nodes (where $\mathrm{card}(\Sigma)$ denotes the size of Σ).*

Proof. The states of $\mathfrak{M}^f$ are the equivalence classes in W_Σ. Let g be the function with domain W_Σ and range $\mathcal{P}(\Sigma)$ defined by $g(|w|) = \{\phi \in \Sigma \mid \mathfrak{M}, w \Vdash \phi\}$. It follows from the definition of $\leftrightsquigarrow_\Sigma$ that g is well defined and injective. Thus $\mathrm{card}(W_\Sigma) \leq \mathrm{card}(\mathcal{P}(\Sigma)) = 2^{\mathrm{card}(\Sigma)}$. $\dashv$

Moreover — crucially — filtrations preserve satisfaction in the following sense.

Theorem 2.39 (Filtration Theorem) *Consider the basic modal language. Let $\mathfrak{M}^f$ $(= (W_\Sigma, R^f, V^f))$ be a filtration of $\mathfrak{M}$ through a subformula closed set Σ. Then for all formulas $\phi \in \Sigma$, and all nodes w in $\mathfrak{M}$, we have $\mathfrak{M}, w \Vdash \phi$ iff $\mathfrak{M}^f, |w| \Vdash \phi$.*

Proof. By induction on ϕ. The base case is immediate from the definition of V^f. The boolean cases are straightforward; the fact that Σ is closed under subformulas allows us to apply the inductive hypothesis.

So suppose $\Diamond\phi \in \Sigma$ and $\mathfrak{M}, w \Vdash \Diamond\phi$. Then there is a v such that Rwv and $\mathfrak{M}, v \Vdash \phi$. As $\mathfrak{M}^f$ is a filtration, $R^f|w||v|$. As Σ is subformula closed, $\phi \in \Sigma$, thus by the inductive hypothesis $\mathfrak{M}^f, |v| \Vdash \phi$. Hence $\mathfrak{M}^f, |w| \Vdash \Diamond\phi$.

Conversely, suppose $\Diamond\phi \in \Sigma$ and $\mathfrak{M}^f, |w| \Vdash \Diamond\phi$. Thus there is a state $|v|$ in $\mathfrak{M}^f$ such that $R^f|w||v|$ and $\mathfrak{M}^f, |v| \Vdash \phi$. As $\phi \in \Sigma$, by the inductive hypothesis $\mathfrak{M}, v \Vdash \phi$. So the third clause in Definition 2.36 is applicable, and we conclude that $\mathfrak{M}, w \Vdash \Diamond\phi$. $\dashv$

Observe that clauses (ii) and (iii) of Definition 2.36 are designed to make the modal case of the induction step go through in the proof above.

But we still have not done one vital thing: we have not actually shown that filtrations exist! Observe that the clauses (ii) and (iii) in Definition 2.36 only impose conditions on candidate relations R^f — but we have not yet shown that a suitable R^f can always be found. In fact, there are always at least two ways to define binary relations that fulfill the required conditions. Define R^s and R^l as follows:

(i) $R^s|w||v|$ iff $\exists w' \in |w| \exists v' \in |v|\ Rw'v'$.
(ii) $R^l|w||v|$ iff for all formulas $\Diamond\phi$ in Σ: $\mathfrak{M}, v \Vdash \phi$ implies $\mathfrak{M}, w \Vdash \Diamond\phi$.

These relations — which are not necessarily distinct — give rise to the *smallest* and *largest* filtrations respectively.

Lemma 2.40 *Consider the basic modal language. Let $\mathfrak{M}$ be any model, Σ any subformula closed set of formulas, W_Σ the set of equivalence classes induced by $\leftrightsquigarrow_\Sigma$, and V^f the standard valuation on W_Σ. Then both (W_Σ, R^s, V^f) and (W_Σ, R^l, V^f) are filtrations of $\mathfrak{M}$ through Σ. Furthermore, if (W_Σ, R^f, V^f) is any filtration of $\mathfrak{M}$ through Σ then $R^s \subseteq R^f \subseteq R^l$.*

Proof. We show that (W_Σ, R^s, V^f) is a filtration; the rest is left as an exercise. It suffices to show that R^s fulfills clauses (ii) and (iii) of Definition 2.36. But R^s satisfies clause (ii) by definition, so it remains to check clause (iii). Suppose $R^s|w||v|$, and further suppose that $\Diamond\phi \in \Sigma$ and $\mathfrak{M}, v \Vdash \phi$. As $R^s|w||v|$, there exist $w' \in |w|$ and $v' \in |v|$ such that $Rw'v'$. As $\phi \in \Sigma$ and $\mathfrak{M}, v \Vdash \phi$, then because $v \leftrightsquigarrow_\Sigma v'$, we get $\mathfrak{M}, v' \Vdash \phi$. But $Rw'v'$, so $\mathfrak{M}, w' \Vdash \Diamond\phi$. But $\Diamond\phi \in \Sigma$, thus as $w' \leftrightsquigarrow_\Sigma w$ it follows that $\mathfrak{M}, w \Vdash \Diamond\phi$. ⊣

Theorem 2.41 (Finite Model Property — via Filtrations) *Let ϕ be a basic modal formula. If ϕ is satisfiable, then it is satisfiable on a finite model. Indeed, it is satisfiable on a finite model containing at most 2^m nodes, where m is the number of subformulas of ϕ.*

Proof. Assume that ϕ is satisfiable on a model $\mathfrak{M}$; take any filtration of $\mathfrak{M}$ through the set of subformulas of ϕ. That ϕ is satisfied in the filtration is immediate from Theorem 2.39. The bound on the size of the filtration is immediate from Proposition 2.38. ⊣

There are several points worth making about filtrations. The first has to do with the possible loss of properties when moving from a model to one of its filtrations. As we have already discussed, a drawback of the selection method is that it can be hard to preserve such properties. Filtrations are far better in this respect — but they certainly are not perfect. Let us consider the matter more closely.

Suppose (W_Σ, R^f, V^f) is a filtration of (W, R, V). Now, clause (ii) of Definition 2.36 means that the natural map from $\mathfrak{M}$ to $\mathfrak{M}^f$ is a homomorphism with respect to the accessibility relation R. Thus any property of relations which is preserved under such maps will automatically be inherited by any filtration. Obvious examples include reflexivity and right unboundedness $(\forall x \exists y\, Rxy)$.

However, many interesting relational properties are *not* preserved under homomorphisms: transitivity and symmetry are obvious counterexamples. Thus we need to find special filtrations which preserve these properties. Sometimes this is easy; for example, the smallest filtration preserves symmetry. Sometimes we need new ideas to find a good filtration; the classic example involves transitivity. Let's see what this involves.

Lemma 2.42 *Let $\mathfrak{M}$ be a model, Σ a subformula closed set of formulas, and W_Σ*

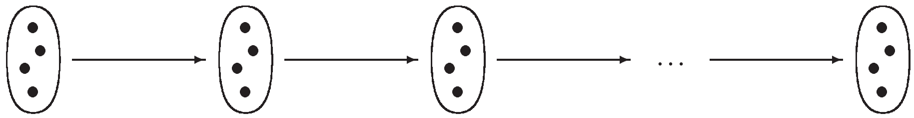


Fig. 2.7. Filtrating a model based on $(\mathbb{Q}, <)$

the set of equivalence classes induced on $\mathfrak{M}$ *by* $\leftrightsquigarrow_\Sigma$. *Let* R^t *be the binary relation on* W_Σ *defined by:*

$$R^t|w||v| \text{ iff for all } \phi, \text{ if } \Diamond\phi \in \Sigma \text{ and } \mathfrak{M}, v \Vdash \phi \vee \Diamond\phi \text{ then } \mathfrak{M}, w \Vdash \Diamond\phi.$$

If R *is transitive then* (W_Σ, R^t, V^f) *is a filtration and* R^t *is transitive.*

Proof. Left as Exercise 2.3.5. ⊣

In short, filtrations are flexible — but it is not a matter of 'plug and play'. Creativity is often required to exploit them.

The second point worth making is that filtrations of an infinite model through a finite set manage to represent an infinite amount of information in a finitary manner. It seems obvious, at least from an intuitive point of view, that this can only be achieved by *identifying* lots of points. As we have seen in Example 2.37, an infinite chain may be collapsed onto a single reflexive point by a filtration. An even more informative example is provided by models based on the rationals. For instance, what happens to the density condition in the filtration? Let $\mathfrak{M} = (\mathbb{Q}, <, V)$; then any (finite) filtration of $\mathfrak{M}$ has the form displayed in Figure 2.7. What is going on here? Instead of viewing models as structures made up of states and relations between them, in the case of filtrations it can be useful to view them as *sets* of states (namely, the sets of identified states) and relations between those sets. The following definition captures this idea.

Definition 2.43 Let (W, R, V) be a transitive frame. A *cluster* on (W, R, V) is a subset C of W that is a maximal equivalence relation under R. That is, the restriction of R to C is an equivalence relation, and this is *not* the case for any other subset D of W such that $C \subset D$.

A cluster is *simple* if it consists of a single reflexive point, and *proper* if it contains more than one point. ⊣

As Figure 2.7 shows, a (finite) filtration of $(\mathbb{Q}, <)$ can be thought of as resulting in a finite linear sequence of clusters, perhaps interspersed with singleton irreflexive points (no two of which can be adjacent). The reader is asked to check this claim in Exercise 2.3.9. Clusters will play an important role in Section 4.5.

To conclude this section we briefly indicate how the filtration method can be extended to other modal languages. Let us first consider modal languages based on arbitrary modal similarity types τ. Fix a τ-model $\mathfrak{M} = (W, R_\triangle, V)_{\triangle \in \tau}$ and a subformula closed set Σ as in Definition 2.36. Suppose $\mathfrak{M}^f_\Sigma = (W_\Sigma, R^f_\triangle, V^f)_{\triangle \in \tau}$ is a τ-model where W_Σ and V^f are as in Definition 2.36, and for $\triangle \in \tau$, $R^f_\triangle$ satisfy

(ii)′ If $R_\triangle w v_1 \dots v_n$ then $R^f |w||v_1| \dots |v_n|$.
(iii)′ If $R^f |w||v_1| \dots |v_n|$, then for all $\phi_1, \dots, \phi_n \in \Sigma$, if $\triangle(\phi_1, \dots, \phi_n) \in \Sigma$ and $\mathfrak{M}, v_1 \Vdash \phi_1, \dots, \mathfrak{M}, v_n \Vdash \phi_n$, then $\mathfrak{M}, w \Vdash \triangle(\phi_1, \dots, \phi_n)$.

Then $\mathfrak{M}^f_\Sigma$ is a *τ-filtration of $\mathfrak{M}$ through Σ.*

With this definition at hand, Proposition 2.38 and Theorem 2.39 can be reformulated and proved for τ-filtrations, and suitable versions of the smallest and largest filtrations can also be defined, resulting in a general modal analog of Theorem 2.41, the Finite Model Property.

What about basic temporal logic, PDL, and arrow logic? It turns out that the filtration method works well for all of these. For basic temporal logic we need to issue the customary warning (we need to be explicit about what the filtration does to $R^{\breve{}}$), but with this observed, matters are straightforward. Exercise 2.3.7 asks the reader to define transitive filtrations for the basic temporal language.

Matters are far more interesting (and difficult) with PDL — but here too, by making use of a clever idea called the Fisher-Ladner closure, it is possible to use a filtration style argument to show that PDL has the finite model property; we will do this in Section 4.8 as part of a completeness proof (Theorem 4.91). Exercise 2.3.10 deals with the finite model property for arrow logic.

Exercises for Section 2.3

2.3.1 Find two models $\mathfrak{M}$ and $\mathfrak{M}'$ and states w and w' in these models such that $w \leftrightarroweq_n w'$ for all n, but it is *not* the case that $w \leftrightarroweq w'$ are bisimilar. (Hint: we drew a picture of such a pair of models in the previous section.)

2.3.2 Generalize the definition of n-bisimulations (Definition 2.30) from diamond-only to arbitrary modal languages. Then prove Proposition 2.31 (that n bisimilarity for all n implies modal equivalence and conversely) for arbitrary modal languages.

2.3.3 Lemma 2.33 tells us that if we are only interested in the satisfiability of modal formulas of degree at most k, we can delete all states that lie beyond the k-horizon without affecting satisfiability. Prove this.

2.3.4 The proof of Theorem 2.34 uses a selection of points argument to establish the finite model property. But no proof details were given for the last (crucial) claim in the proof, namely that $\mathfrak{M}_4, w_2$ is k-bisimilar to $\mathfrak{M}_3, w_2$. Fill in this gap.

2.3.5 First show that not every filtration of a transitive model is transitive. Then prove Lemma 2.42. That is, show that the relation R^t defined there is indeed a filtration, and that any filtration of a transitive model that makes use of R^t is guaranteed to be transitive.

2.3.6 Finish the proof of Lemma 2.40. That is, prove that the filtrations R^s and R^l are indeed the smallest and the largest filtration, respectively. In addition, give an example of a model and a set of formulas for which R^s and R^l coincide.

2.3.7 Show that every transitive model (W, R, V) has a transitive *temporal* filtration. (Take care to specify what the filtration does to $R^{\smile}$.)

2.3.8 Call a frame or model *euclidean* if it satisfies $\forall xyz\,((Rxy \wedge Rxz) \to Ryz)$, and let E be the class of euclidean models. Fix a formula ξ, and let Σ be the smallest subformula closed set of formulas containing ξ that satisfies, for all formulas ψ: if $\Diamond\phi \in \Sigma$, then $\Box\Diamond\psi \in \Sigma$. (Recall that $\Box$ is an abbreviation of $\neg\Diamond\neg$.) Note that in general, Σ will be infinite.

(a) Prove that $\mathsf{E} \Vdash \Diamond\psi \to \Box\Diamond\psi$.
(b) Prove that every euclidean model can be filtrated through Σ to a euclidean model.
(c) Show that every euclidean model satisfies the following modal reduction principles: $\Diamond\Diamond\Diamond \leftrightarrow \Diamond\Diamond$, $\Diamond\Diamond\Box \leftrightarrow \Diamond\Box$, $\Diamond\Box\Diamond \leftrightarrow \Diamond\Diamond$ and $\Diamond\Box\Box \leftrightarrow \Diamond\Box$. That is, prove that the formulas $\Diamond\Diamond\Diamond\phi \leftrightarrow \Diamond\Diamond\phi, \ldots$ are true throughout every euclidean model. Conclude that Σ is finite modulo equivalence on euclidean models.
(d) Prove that the basic modal similarity type has the finite model property with respect to the class of euclidean models. Can you prove this result simply by filtrating through any subformula closed set of formulas containing ξ?

2.3.9 Show that any finite filtration of a model based on the rationals with their usual ordering is a finite linear sequence of clusters, perhaps interspersed with singleton irreflexive points, no two of which can be adjacent.

2.3.10 Consider the similarity type $\tau_{\rightarrow}$ of arrow logic.

(i) Show that $\tau_{\rightarrow}$ has the finite model property with respect to the class of all arrow models.
(ii) Consider the class of arrow models based on arrow frames $\mathfrak{F} = (W, C, R, I)$ such that for all s, t and u in W we have (i) $Cstu$ iff $Csut$ iff $Ctus$ and (ii) $Cstu$ and Iu iff $s = t$. Prove that arrow formulas have the finite model property with respect to this class of arrow models.
(iii) Prove that $\tau_{\rightarrow}$ does not have the finite model property with respect to the class of all square models. (Hint: try to express that the extension of the propositional variable p is a dense, linear ordering.)

2.4 The Standard Translation

In the Preface we warned the reader against viewing modal logic as an isolated formal system (remember Slogan 3?), yet here we are, halfway through Chapter 2, and we still haven't linked modal logic with the wider logical world. We now put this right. We define a link called the *standard translation*. This paves the way for the results on modal expressivity in the sections that follow, for the study of frames in the following chapter, and for the introduction of the guarded fragment in Section 7.4.

We first specify our *correspondence languages* — that is, the languages we will translate modal formulas into.

Definition 2.44 For τ a modal similarity type and Φ a collection of proposition letters, let $\mathcal{L}^1_\tau(\Phi)$ be the first-order language (with equality) which has unary predicates P_0, P_1, P_2, $\ldots$ corresponding to the proposition letters p_0, p_1, p_2, $\ldots$ in Φ, and an $(n+1)$-ary relation symbol $R_\triangle$ for each (n-ary) modal operator $\triangle$ in our similarity type. We write $\alpha(x)$ to denote a first-order formula α with one free variable, x. $\dashv$

We are now ready to define the standard translation.

Definition 2.45 (Standard Translation) Let x be a first-order variable. The *standard translation* ST_x taking modal formulas to first-order formulas in $\mathcal{L}^1_\tau(\Phi)$ is defined as follows:

$$\begin{aligned} ST_x(p) &= Px \\ ST_x(\bot) &= x \neq x \\ ST_x(\neg\phi) &= \neg ST_x(\phi) \\ ST_x(\phi \vee \psi) &= ST_x(\phi) \vee ST_x(\psi) \\ ST_x(\triangle(\phi_1, \ldots, \phi_n)) &= \exists y_1 \ldots \exists y_n \, (R_\triangle x y_1 \ldots y_n \wedge \\ &\qquad ST_{y_1}(\phi_1) \wedge \cdots \wedge ST_{y_n}(\phi_n)), \end{aligned}$$

where $y_1, \ldots, y_n$ are fresh variables (that is, variables that have not been used so far in the translation). When working with the basic modal language, the last clause boils down to:

$$ST_x(\Diamond\phi) = \exists y \, (Rxy \wedge ST_y(\phi)).$$

Note that (to keep notation simple) we prefer to use R rather than $R_\Diamond$, and we will continue to do this. We leave to the reader the task of working out what $ST_x(\nabla(\phi_1, \ldots, \phi_n))$ is, but we will point out that for the basic modal language the required clause is:

$$ST_x(\Box\phi) = \forall y \, (Rxy \to ST_y(\phi)). \quad \dashv$$

Example 2.46 Let's see how this works. Consider the formula $\Diamond(\Box p \to q)$.

$$\begin{aligned} ST_x(\Diamond(\Box p \to q)) &= \exists y_1 \, (Rxy_1 \wedge ST_{y_1}(\Box p \to q)) \\ &= \exists y_1 \, (Rxy_1 \wedge (ST_{y_1}(\Box p) \to ST_{y_1}(q))) \\ &= \exists y_1 \, (Rxy_1 \wedge (\forall y_2 \, (Ry_1y_2 \to ST_{y_2}(p)) \to Qy_1)) \\ &= \exists y_1 \, (Rxy_1 \wedge (\forall y_2 \, (Ry_1y_2 \to Py_2) \to Qy_1)) \end{aligned}$$

Note that (this version of) the standard translation leaves the choice of fresh variables unspecified. For example, $\exists y_{256} \, (Rxy_{256} \wedge (\forall y_{14} \, (Ry_{256}y_{14} \to Py_{14}) \to Qy_{256}))$ is a legitimate translation of $\Diamond(\Box p \to q)$, and indeed there are infinitely many others, all differing only in the bound variables they contain. Later in the section we remove this indeterminacy — elegantly. ⊣

It should be clear that the standard translation makes good sense: it is essentially a first-order reformulation of the modal satisfaction definition. For any modal formula ϕ, $ST_x(\phi)$ will contain exactly one free variable (namely x); the role of this free variable is to mark the current state; this use of a free variable makes it possible for the global notion of first-order satisfaction to mimic the local notion of modal satisfaction. Furthermore, observe that modalities are translated as *bounded quantifiers*, and in particular, quantifiers bounded to act only on related states; this is the obvious way of mimicking the local action of the modalities in first-order logic. Because of its importance it is worth pinning down just why the standard translation works.

Models for modal languages based on a modal similarity type τ and a collection of proposition letters Φ can also be viewed as models for $\mathcal{L}^1_\tau(\Phi)$. For example, if τ contains just a single diamond $\Diamond$, then the corresponding first-order language $\mathcal{L}^1_\tau(\Phi)$ has a binary relation symbol R and a unary predicate symbol corresponding to each proposition letter in Φ — and a first-order model for this language needs to provide an interpretation for these symbols. But a (modal) model $\mathfrak{M} = (W, R, V)$ supplies precisely what is required: the binary relation R can be used to interpret the relation symbol R, and the set $V(p_i)$ can be used to interpret the unary predicate P_i. This should *not* come as a surprise. As we emphasized in Chapter 1 (especially Sections 1.1 and 1.3) there is no mathematical distinction between modal and first-order models — both modal and first-order models are simply relational structures. Thus it makes perfect sense to write things like $\mathfrak{M} \models ST_x(\phi)[w]$, which means that the first-order formula $ST_x(\phi)$ is satisfied (in the usual sense of first-order logic) in the model $\mathfrak{M}$ when w is assigned to the free variable x.

Proposition 2.47 (Local and Global Correspondence on Models) *Fix a modal similarity type τ, and let ϕ be a τ-formula. Then:*

(i) *For all $\mathfrak{M}$ and all states w of $\mathfrak{M}$: $\mathfrak{M}, w \Vdash \phi$ iff $\mathfrak{M} \models ST_x(\phi)[w]$.*
(ii) *For all $\mathfrak{M}$: $\mathfrak{M} \Vdash \phi$ iff $\mathfrak{M} \models \forall x\, ST_x(\phi)$.*

Proof. By induction on ϕ. We leave this to the reader as Exercise 2.4.1. ⊣

Summing up: when interpreted on models, modal formulas are equivalent to first-order formulas in one free variable. Fine — but what does that give us? Lots! Proposition 2.47 is 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.

Example 2.48 First-order logic has the compactness property: if Θ is a set of first-order formulas, and every every finite subset of Θ is satisfiable, then so is Θ

itself. It also has the downward Löwenheim-Skolem property: if a set of first-order formulas has an infinite model, then it has a countably infinite model.

It follows that modal logic must have both these properties (over models) too. Consider compactness. Suppose Σ is a set of modal formulas every finite subset of which is satisfiable — is Σ itself satisfiable? Yes. Consider the set $\{ST_x(\phi) \mid \phi \in \Sigma\}$. As every finite subset of Σ has a model it follows (reading item (i) of Proposition 2.47 left to right) that every finite subset of $\{ST_x(\phi) \mid \phi \in \Sigma\}$ does too, and hence (by first-order compactness) that this whole set is satisfiable in some model, say $\mathfrak{M}$. But then it follows (this time reading item (i) of Proposition 2.47 right to left) that Σ is satisfiable in $\mathfrak{M}$, hence modal satisfiability over models is compact.

And there's interesting traffic from modal logic to first-order logic too. For example, a significant difference between modal and first-order logic is that modal logic is decidable (over arbitrary models) but first-order logic is not. By using our understanding of modal decidability, it is possible to locate novel decidable fragments of first-order logic, a theme we will return to in Section 7.4 when we discuss the guarded fragment. $\dashv$

Just as importantly, the standard translation gives us a new research agenda for investigating modal expressivity: *correspondence theory*. The central aim of this chapter is to explore the expressivity of modal logic over models — but how is expressivity to be measured? Proposition 2.47 suggests an interesting strategy: try to characterize the fragment of first-order logic picked out by the standard translation.

It is obvious on purely syntactic grounds that the standard translation is not surjective (standard translations of modal formulas contain only bounded quantifiers) — but could every first-order formula (in the appropriate correspondence language) be *equivalent* to the translation of a modal formula? No. This is very easy to see: whereas modal formulas are invariant under bisimulations, first-order formulas need not be; thus any first-order formula which is not invariant under bisimulations cannot be equivalent to the translation of a modal formula. We have seen such a formula in Section 2.2, (namely $\exists y_1 y_2 y_3\,(y_1 \neq y_2 \wedge y_1 \neq y_3 \wedge y_2 \neq y_3 \wedge Rxy_1 \wedge Rxy_2 \wedge Ry_1y_3 \wedge Ry_2y_3)$), and it is easy to find simpler examples.

Thus the (first-order formulas equivalent to) standard translations of model formulas are a proper subset of the correspondence language. Which subset? Here's a nice observation. The standard translation can be reformulated so that it maps every modal formula into a very small fragment of $\mathcal{L}^1_\tau(\Phi)$, namely a certain *finite-variable fragment*. Suppose the variables of $\mathcal{L}^1_\tau(\Phi)$ have been ordered in some way. Then the n-variable fragment of $\mathcal{L}^1_\tau(\Phi)$ is the set of $\mathcal{L}^1_\tau(\Phi)$ formulas that contain only the first n variables. As we will now see, by judicious reuse of variables, a modal language with operators of arity at most n can be translated into the $n+1$-variable fragment of $\mathcal{L}^1_\tau(\Phi)$. (Reuse of variables is the name of the game when working with finite variable fragments. For example, we can express the existence of *three* different points in a linear ordering using only *two* variables as follows: $\exists xy\,(x < y \land \exists x\,(y < x))$.)

Proposition 2.49 (i) *Let τ be a modal similarity type that only contains diamonds. Then, every τ-formula ϕ is equivalent to a first-order formula containing at most two variables.*

(ii) *More generally, if τ does not contain modal operators $\triangle$ whose arity exceeds n, all τ-formulas are equivalent to first-order formulas containing at most $n+1$ variables.*

Proof. Assume τ contains only diamonds $\langle a\rangle$, $\langle b\rangle$, $\ldots$; proving the general case is left as Exercise 2.4.2. Fix two distinct individual variables x and y. Define two variants ST_x and ST_y of the standard translation as follows.

$$\begin{aligned}
ST_x(p) &= Px & ST_y(p) &= Py\\
ST_x(\bot) &= x \neq x & ST_y(\bot) &= y \neq y\\
ST_x(\neg\phi) &= \neg ST_x(\phi) & ST_y(\neg\phi) &= \neg ST_y(\phi)\\
ST_x(\phi\lor\psi) &= ST_x(\phi)\lor ST_x(\psi) & ST_y(\phi\lor\psi) &= ST_y(\phi)\lor ST_y(\psi)\\
ST_x(\langle a\rangle\phi) &= \exists y\,(R_a xy \land ST_y(\phi)) & ST_y(\langle a\rangle\phi) &= \exists x\,(R_a yx \land ST_x(\phi)).
\end{aligned}$$

Then, for any τ-formula ϕ, its ST_x-translation contains at most the two variables x and y, and $ST_x(\phi)$ is equivalent to the original standard translation of ϕ. $\dashv$

Example 2.50 Let's see how this modified standard translation works. Consider again the formula $\Diamond(\Box p \to q)$.

$$\begin{aligned}
ST_x(\Diamond(\Box p\to q)) &= \exists y\,(Rxy \land ST_y(\Box p \to q))\\
&= \exists y\,(Rxy \land (\forall x\,(Ryx \to ST_x(p)) \to Qy))\\
&= \exists y\,(Rxy \land (\forall x\,(Ryx \to Px) \to Qy))
\end{aligned}$$

That is, we just keep flipping between the two variables x and y. The result is a translation containing only two variables (instead of the three used in Example 2.46). As a side effect, the indeterminacy associated with the original version of the standard translation has disappeared. $\dashv$

This raises another question: is every first-order formula $\alpha(x)$ in two variables equivalent to the translation of a basic modal formula? Again the answer is *no*. There is even a first-order formula in a single variable x which is not equivalent to any modal formula, namely Rxx. To see this, assume for the sake of a contradiction that ϕ is a modal formula such that $ST_x(\phi)$ is equivalent to Rxx. Let $\mathfrak{M}$ be a singleton reflexive model and let w be the unique state in $\mathfrak{M}$; obviously (irrespective of the valuation) $\mathfrak{M} \models Rxx[w]$. Let $\mathfrak{N}$ be a model based on the strict ordering of the integers; obviously (again, irrespective of the valuation), for every integer v, $\mathfrak{N} \models \neg Rxx[v]$. Let Z be the relation which links every integer with the unique state in $\mathfrak{M}$, and assume that the valuations in $\mathfrak{N}$ and $\mathfrak{M}$ are such that Z is a bisimulation (for example, make all proposition letters true at all points in both models). As $\mathfrak{M} \models Rxx[w]$, it follows by Proposition 2.47 that $\mathfrak{M}, w \Vdash \phi$ (after all, by assumption Rxx is equivalent to $ST_x(\phi)$). But for any integer v, we have that $w \leftrightarroweq v$, hence $\mathfrak{N}, v \Vdash \phi$. Hence (again by Proposition 2.47 and our assumption that $ST_x(\phi)$ is equivalent to Rxx) we have that $\mathfrak{N} \models Rxx[v]$, contradicting the fact that $\mathfrak{N} \models \neg Rxx[v]$.

We will not discuss correspondence theory any further here, but in Section 2.6 we will prove one of its central results, the Van Benthem Characterization Theorem: a first-order formula is equivalent to the translation of a modal formula if and only if it is invariant under bisimulations.

Proposition 2.47 is also going to help us investigate modal expressivity in other ways, notably via the concept of definability.

Definition 2.51 Let τ be a modal similarity type, C a class of τ-models, and Γ a set of formulas over τ. We say that Γ *defines* or *characterizes* a class K of models *within* C if for all models $\mathfrak{M}$ in C we have that $\mathfrak{M}$ is in K iff $\mathfrak{M} \Vdash \Gamma$. If C is the class of all τ-models, we simply say that Γ defines or characterizes K; we omit brackets whenever Γ is a singleton. We will say that a formula ϕ defines a *property* whenever ϕ defines the class of models satisfying that property. $\dashv$

It is immediate from Proposition 2.47 that if a class of models is definable by a set of modal formulas, then it is also definable by a set a first-order formulas — but this is too obvious to be interesting. The important way in which Proposition 2.47 helps, is by making it possible to exploit standard model construction techniques from first-order model theory. For example, in Section 2.6 we will prove Theorem 2.75 which says that a class of (pointed) models is modally definable if and only if it is closed under bisimulations and ultraproducts (an important construction known from first-order model theory; see Appendix A), and its complement is closed under ultrapowers (another standard model theoretic construction). It would be difficult to overemphasize the importance of the standard translation; it is remarkable that such a simple idea can lead to so much.

To conclude this section, let's see how to adapt these ideas to the basic temporal language, PDL, and arrow logic. The case of basic temporal logic is easy: all we have to do is add a clause for translating the backward looking operator P:

$$ST_x(P\phi) = \exists y\,(Ryx \wedge ST_y(\phi)).$$

Note that we are using the more sophisticated approach introduced in the proof of Proposition 2.49: flipping between two translations ST_x and ST_y. (Thus we really need to add a mirror clause which flips the variables back.) So, just like the basic modal language, the basic temporal language can be mapped into a two variable fragment of the correspondence language. Moreover (again, as with the basic modal language) not every first-order formula in two variables is equivalent to (the translation of) a basic temporal formula (see Exercise 2.4.3).

Propositional dynamic logic calls for more drastic changes. Let's first look at the $*$-free fragment — that is, at PDL formulas without occurrences of the Kleene star. In PDL both formulas and modalities are recursively structured, so we're going to need two interacting translation functions: one to handle the formulas, the other to handle the modalities. The only interesting clause in the formula translation is the following:

$$ST_x(\langle\pi\rangle\phi) = \exists y\,(ST_{xy}(\pi) \wedge ST_y(\phi)).$$

That is, instead of returning a fixed relation symbol (say R), the formula translation ST_x calls on ST_{xy} to start recursively decomposing the program π. Why does this part of the translation require two free variables? Because its task is to define a binary relation.

$$\begin{array}{rcl} ST_{xy}(a) & = & R_a xy \text{ (and similarly for other pairs of variables)} \\ ST_{xy}(\pi_1 \cup \pi_2) & = & ST_{xy}(\pi_1) \vee ST_{xy}(\pi_2) \\ ST_{xy}(\pi_1\,;\pi_2) & = & \exists z\,(ST_{xz}(\pi_1) \wedge ST_{zy}(\pi_2)). \end{array}$$

It follows that we can translate the $*$-free fragment of PDL into a *three* variable fragment of the correspondence language. The details are worth checking; see Exercise 2.4.4.

But the really drastic change comes when we consider the full language of PDL (that is, with Kleene star). Recall that a program α^* is interpreted using the reflexive, transitive closure of R_α. But the reflexive, transitive closure of an arbitrary relations is *not* a first-order definable relation (see Exercise 2.4.5). So the standard translation for PDL needs to take us to a richer background logic than first-order logic, one that can express this concept. Which one should we use? There are many options here, but to motivate our actual choice recall the definition of the meaning of a PDL program α^*:

$$R_{\alpha^*} = \bigcup_n (R_\alpha)^n,$$

where R_α^n is defined by

$$R^0xy \text{ iff } x = y \quad \text{and} \quad R^{n+1}xy \text{ iff } \exists z\,(R^n xz \wedge Rzy).$$

Thus, if we were allowed to write infinitely long disjunctions, it would be easy to capture the meaning of an iterated program α^*:

$$(R_\alpha)^*xy \text{ iff } (x = y)\ \vee R_\alpha xy \vee \bigvee_{n\geq 1} \exists z_1 \ldots z_n\,(R_\alpha xz_1 \wedge \cdots \wedge R_\alpha z_n y).$$

In *infinitary logic* we can do this. More precisely, in $\mathcal{L}_{\omega_1\omega}$ we are allowed to form formulas as in first-order logic, and, in addition, to build countably infinite disjunctions and conjunctions. We will take $\mathcal{L}_{\omega_1\omega}$ as the target logic for the standard translation of PDL. We have seen most of the clauses we need: we use the clauses for the $*$-free fragment given above, and in addition the following clause to cater for the Kleene star:

$$ST_{xy}(\alpha^*) = \\ (x = y) \vee ST_{xy}(\alpha) \vee \bigvee_{n\geq 1} \exists z_1 \ldots z_n \, (ST_{xz_1}(\alpha) \wedge \cdots \wedge ST_{z_n y}(\alpha)).$$

This example of PDL makes an important point vividly: we cannot always hope to embed modal logic into first-order logic. Indeed in the following chapter we will see that when it comes to analyzing the expressive power of modal logic at the level of frames, the natural correspondence language (even for the basic modal language) is second-order logic.

There is nothing particularly interesting concerning the standard translation for the arrow language of Example 1.16. However, this changes when we turn to *square* models: in Exercise 2.4.6 the reader is asked to prove that on this class of models, the arrow language corresponds to a first-order language with *binary* predicate symbols, and that, in fact, it is expressively *equivalent* to the three variable fragment of such a language.

Exercises for Section 2.4

2.4.1 Prove Proposition 2.47. That is, check that the standard translation really is correct.

2.4.2 Prove Proposition 2.49 for arbitrary modal languages. That is, show that if τ does not contain modal operators $\triangle$ whose arity exceeds n, all τ-formulas are equivalent to first-order formulas containing at most $n + 1$ variables.

2.4.3 Show that there are first-order formulas $\alpha(x)$ using at most two variables that are not equivalent to the standard translation of a basic temporal formula.

2.4.4 In this exercise you should fill in some of the details for the standard translation for PDL.

(a) Check that the translation for the $*$-free fragment of PDL really does map all such formulas into the three variable fragment of the corresponding first-order language.
(b) Show that in fact, there is a translation into the *two* variable fragment of this corresponding first-order language.

2.4.5 The aim of this exercise is to show that taking the reflexive, transitive closure of a binary relation is not a first-order definable operation.

(a) Show that the class of connected graphs is not first-order definable:

(i) For $l \in \mathbb{N}$, let $\mathfrak{C}_l$ be the graph given by a cycle of length $l + 1$:

$$\mathfrak{C}_l = (\{0, \ldots, l\}, \{(i, i + 1), (i + 1, i) \mid 0 \leq i < l\} \cup \{(0, l), (l, 0)\})$$

Show that for every $k \in \mathbb{N}$ and $l \geq 2^k$ the graph $\mathfrak{C}_l$ satisfies the same first-order sentences of quantifier rank at most k as the disjoint union $\mathfrak{C}_l \uplus \mathfrak{C}_l$.

(ii) Conclude that the class of connected graphs is not first-order definable.

(b) Use item (a) to conclude that the reflexive transitive closure of a relation is not first-order definable.

2.4.6 Consider the class of square models for arrow logic. Observe that a square model $\mathfrak{M} = (\mathfrak{S}_U, V)$ can be seen as a first-order model $\mathfrak{M}^* = (U, V(p))_{p \in \Phi}$ if we let each propositional variable $p \in \Phi$ correspond to a *dyadic* relation symbol P.

(a) Work out this observation in the following sense. Define a suitable translation $(\cdot)^*$ mapping an arrow formula ϕ to a formula $\phi^*(x_0, x_1)$ in this 'dyadic correspondence language'. Prove that this translation has the property that for all arrow formulas ϕ and all square models $\mathfrak{M}$ the following correspondence holds:

$$\mathfrak{M}, (a_0, a_1) \Vdash \phi \text{ iff } \mathfrak{M}^* \models \phi^*(x_0, x_1)[a_0, a_1].$$

(b) Show that this translation can be done within the three variable fragment of first-order logic.

(c) Prove that conversely, every formula $\alpha(x_0, x_1)$ that uses only three variables, in a first-order language with binary predicates only, is equivalent to the translation of an arrow formula on the class of square models.

2.5 Modal Saturation via Ultrafilter Extensions

Bisimulations and the standard translation are two of the tools we need to understand modal expressivity over models. This section introduces the third: *ultrafilter extensions*. To motivate their introduction, we will first discuss *Hennessy-Milner model classes* and *modally saturated models*; both generalize ideas met in our earlier discussion of bisimulations. We will then introduce ultrafilter extensions as a way of building modally saturated models, and this will lead us to an elegant result: modal equivalence implies bisimilarity-somewhere-else.

M-saturation

Theorem 2.20 tells us that bisimilarity implies modal equivalence, but we have already seen that the converse does not hold in general (recall Figure 2.5). The Hennessy-Milner theorem shows that the converse does hold in the special case of image-finite models. Let's try and generalize this theorem.

First, when proving Theorem 2.24, we exploited the fact that, between image-finite models, the relation of modal equivalence *itself* is a bisimulation. Classes of models for which this holds are evidently worth closer study.

Definition 2.52 (Hennessy-Milner Classes) Let τ be a modal similarity type, and K a class of τ-models. K is a *Hennessy-Milner* class, or *has the Hennessy-Milner property*, if for every two models $\mathfrak{M}$ and $\mathfrak{M}'$ in K and any two states w, w' of $\mathfrak{M}$ and $\mathfrak{M}'$, respectively, $w \leftrightsquigarrow w'$ implies $\mathfrak{M}, w \leftrightarroweq \mathfrak{M}, w'$. ⊣

For example, by Theorem 2.24, the class of image-finite models has the Hennessy-Milner property. On the other hand, no class of models containing the two models in Figure 2.5 has the Hennessy-Milner property.

We generalize the notion of image-finiteness; doing so leads us to the concept of *modally-saturated* or (briefly) *m-saturated* models. Suppose we are working in the basic modal language. Let $\mathfrak{M} = (W, R, V)$ be a model, let w be a state in W, and let $\Sigma = \{\phi_0, \phi_1, \ldots\}$ be an infinite set of formulas. Suppose that w has successors $v_0, v_1, v_2, \ldots$ where (respectively) ϕ_0, $\phi_0 \wedge \phi_1$, $\phi_0 \wedge \phi_1 \wedge \phi_2, \ldots$ hold. If there is no successor v of w where *all* formulas from Σ hold *at the same time*, then the model is in some sense incomplete. A model is called m-saturated if incompleteness of this kind does not occur.

To put it another way: suppose that we are looking for a successor of w at which every formula ϕ_i of the infinite set of formulas $\Sigma = \{\phi_0, \phi_1, \ldots\}$ holds. M-saturation is a kind of compactness property, according to which it suffices to find satisfying successors of w for arbitrary finite approximations of Σ.

Definition 2.53 (M-saturation) Let $\mathfrak{M} = (W, R, V)$ be a model of the basic modal similarity type, X a subset of W and Σ a set of modal formulas. Σ is *satisfiable* in the set X if there is a state $x \in X$ such that $\mathfrak{M}, x \models \phi$ for all ϕ in Σ; Σ is *finitely satisfiable* in X if every finite subset of Σ is satisfiable in X.

The model $\mathfrak{M}$ is called *m-saturated* if it satisfies the following condition for every state $w \in W$ and every set Σ of modal formulas.

If Σ is finitely satisfiable in the set of successors of w,
then Σ is satisfiable in the set of successors of w.

The definition of m-saturation for arbitrary modal similarity types runs as follows. Let τ be a modal similarity type, and let $\mathfrak{M}$ be a τ-model. $\mathfrak{M}$ is called *m-saturated* if, for every state w of $\mathfrak{M}$ and every (n-ary) modal operator $\triangle \in \tau$ and sequence $\Sigma_1, \ldots, \Sigma_n$ of sets of modal formulas we have the following.

If for every sequence of finite subsets $\Delta_1 \subseteq \Sigma_1, \ldots, \Delta_n \subseteq \Sigma_n$ there are states $v_1, \ldots, v_n$ such that $Rwv_1 \ldots v_n$ and $v_1 \Vdash \Delta_1, \ldots, v_n \Vdash \Delta_n$,
then there are states $v_1, \ldots, v_n$ in $\mathfrak{M}$ such that $Rwv_1 \ldots v_n$ and $v_1 \Vdash \Sigma_1, \ldots, v_n \Vdash \Sigma_n$. ⊣

Proposition 2.54 *Let τ be a modal similarity type. Then the class of m-saturated τ-models has the Hennessy-Milner property.*

Proof. We only prove the proposition for the basic modal language. Let $\mathfrak{M} = (W, R, V)$ and $\mathfrak{M}' = (W', R', V')$ be two m-saturated models. It suffices to prove that the relation $\leftrightsquigarrow$ of modal equivalence between states in $\mathfrak{M}$ and states in $\mathfrak{M}'$ is a bisimulation. We confine ourselves to a proof of the forth condition of a bisimulation, since the condition concerning the propositional variables is trivially satisfied, and the back condition is completely analogous to the case we prove.

So, assume that $w, v \in W$ and $w' \in W'$ are such that Rwv and $w \leftrightsquigarrow w'$. Let Σ be the set of formulas true at v. It is clear that for every finite subset Δ of Σ we have $\mathfrak{M}, v \Vdash \bigwedge \Delta$, hence $\mathfrak{M}, w \Vdash \Diamond \bigwedge \Delta$. As $w \leftrightsquigarrow w'$, it follows that $\mathfrak{M}', w' \Vdash \Diamond \bigwedge \Delta$, so w' has an R'-successor v_Δ such that $\mathfrak{M}', v_\Delta \Vdash \bigwedge \Delta$. In other words, Σ is finitely satisfiable in the set of successors of w'; but, then, by m-saturation, Σ itself is satisfiable in a successor v' of w'. Thus $v \leftrightsquigarrow v'$. $\dashv$

Ultrafilter extensions

So the class of m-saturated models satisfies the Hennessy-Milner property — but how do we actually *build* m-saturated models? To this end, we will now introduce the last of the 'big four' model constructions: *ultrafilter extensions*. The ultrafilter extension of a structure (model or frame) is a kind of *completion* of the original structure. The construction adds states to a model in order to make it m-saturated. Sometimes the result is a model isomorphic to the original (for example, when the original model is finite) but when working with infinite models, the ultrafilter extension always adds lots of new points. power set algebra of a frame; we have met this operation already in Section 1.4 when we introduced general frames, but we repeat the definition here.

Definition 2.55 Let τ be a modal similarity type, and $\mathfrak{F} = (W, R_\triangle)_{\triangle \in \tau}$ a τ-frame. For each $(n+1)$-ary relation $R_\triangle$, we define the following two operations $m_\triangle$ and $m^\delta_\triangle$ on the power set $\mathcal{P}(W)$ of W.

$$\begin{aligned} m_\triangle(X_1, \ldots, X_n) \quad &:= \quad \{w \in W \mid \text{ there exist } w_1, \ldots, w_n \text{ such that} \\ &\qquad\quad R_\triangle w w_1 \ldots w_n \text{ and } w_i \in X_i \text{ for all } i\} \\ m^\delta_\triangle(X_1, \ldots, X_n) \quad &:= \quad \{w \in W \mid \text{ for all } w_1, \ldots, w_n\text{: if } R_\triangle w w_1 \ldots w_n, \\ &\qquad\quad \text{then there is an } i \text{ with } w_i \in X_i\}. \quad \dashv \end{aligned}$$

In the basic modal language $m_\Diamond(X)$ is the set of points that 'can see' a state in X, and $m^\delta_\Diamond(X)$ is the set of points that 'only see' states in X. It follows that for any model $\mathfrak{M}$

$$V(\Diamond\phi) = m_\Diamond(V(\phi)) \text{ and } V(\Box\phi) = m^\delta_\Diamond(V(\phi)).$$

Similar identities hold for modal operators of higher arity. Furthermore, $m_\triangle$ and $m^\delta_\triangle$ are each other's dual, in the following sense:

Proposition 2.56 *Let τ be a modal similarity type, and $\mathfrak{F} = (W, R_\triangle)_{\triangle\in\tau}$ a τ-frame. For every n-ary modal operator $\triangle$ and for every n-tuple $X_1, \ldots, X_n$ of subsets of W, we have*

$$m^\delta_\triangle(X_1, \ldots, X_n) = W \setminus m_\triangle(W \setminus X_1, \ldots, W \setminus X_n).$$

Proof. Left to the reader. ⊣

We are ready to define ultrafilter extensions. As the name is meant to suggest, the states of the ultrafilter extension of a model $\mathfrak{M}$ are the ultrafilters over the universe of $\mathfrak{M}$. Filters and ultrafilters are discussed in Appendix A. Readers that encounter this notion for the first time, are advised to make the Exercises 2.5.1–2.5.4.

Definition 2.57 (Ultrafilter Extension) Let τ be a modal similarity type, and $\mathfrak{F} = (W,\ R_\triangle)_{\triangle\in\tau}$ a τ-frame. The *ultrafilter extension* $\mathfrak{ue}\,\mathfrak{F}$ of $\mathfrak{F}$ is defined as the frame $(Uf(W), R^{ue}_\triangle)_{\triangle\in\tau}$. Here $Uf(W)$ is the set of ultrafilters over W and $R^{ue}_\triangle u_0 u_1 \ldots u_n$ holds for a tuple $u_0, \ldots, u_n$ of ultrafilters over W if we have that $m_\triangle(X_1, \ldots, X_n) \in u_0$ whenever $X_i \in u_i$ (for all i with $1 \leq i \leq k$).

The *ultrafilter extension* of a τ-model $\mathfrak{M} = (\mathfrak{F}, V)$ is the model $\mathfrak{ue}\,\mathfrak{M} = (\mathfrak{ue}\,\mathfrak{F}, V^{ue})$ where $V^{ue}(p_i)$ is the set of ultrafilters of which $V(p_i)$ is a member. ⊣

What are the intuitions behind this definition? First, note that the main ingredients have a logical interpretation. Any subset of a frame can, in principle, be viewed as (the extension or interpretation of) a *proposition*. A filter over the universe of the frame can thus be seen as a *theory*, in fact as a logically closed theory, since filters are both closed under intersection (conjunction) and upward closed (entailment). Viewed this way, a proper filter is a *consistent* theory, for it does not contain the empty set (falsum). Finally, an ultrafilter is a *complete* theory, or as we will call it, a *state of affairs*: for each proposition (subset of the universe) an ultrafilter decides whether the proposition holds (is a member of the ultrafilter) or not.

How does this relate to ultrafilter extensions? In a given frame $\mathfrak{F}$ not every state of affairs need be 'realized', in the sense that there is a state satisfying all and only the propositions belonging to the state of affairs; only the states of affairs that correspond to the *principal* ultrafilters are realized, namely, as the points of the frame. We build $\mathfrak{ue}\,\mathfrak{F}$ by adding every state of affairs for $\mathfrak{F}$ as a new element of the domain — that is, $\mathfrak{ue}\,\mathfrak{F}$ realizes every proposition in $\mathfrak{F}$.

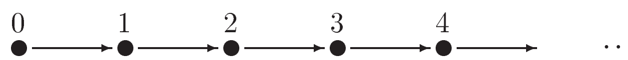
How should we relate these new elements in $\mathfrak{ue}\,\mathfrak{F}$ to each other and to the original elements from $\mathfrak{F}$? The obvious choice is to stipulate that $R^{ue}u_0u_1 \ldots u_n$ if u_0 'sees' the n-tuple $u_1, \ldots, u_n$. That is, whenever $X_1, \ldots, X_n$ are propositions of $u_1, \ldots, u_n$ respectively, then u_0 'sees' this combination: that is, the proposition $m_\triangle(X_1, \ldots, X_n)$ is a member of u_0. The definition of the valuation V^{ue} is self-explanatory.

One final comment: a special role in this section is played by the so-called *principal* ultrafilters over W. Recall that, given an element $w \in W$, the *principal ultrafilter* π_w generated by w is the filter generated by the singleton set $\{w\}$: that is, $\pi_w = \{X \subseteq W \mid w \in X\}$. By identifying a state w of a frame $\mathfrak{F}$ with the principal ultrafilter π_w, it is easily seen that any frame $\mathfrak{F}$ is (isomorphic to) a *submodel* (but in general not a *generated* submodel) of its ultrafilter extension. For we have the following equivalences (here proved for the basic modal similarity type):

$$\begin{aligned} Rwv \quad &\text{iff} \quad w \in m_\diamond(X) \text{ for all } X \subseteq W \text{ such that } v \in X \\ &\text{iff} \quad m_\diamond(X) \in \pi_w \text{ for all } X \subseteq W \text{ such that } X \in \pi_v \\ &\text{iff} \quad R^{ue}\pi_w\pi_v. \end{aligned} \tag{2.1}$$

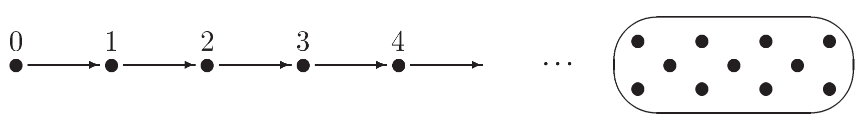
Let's make our discussion more concrete by considering an example.

Example 2.58 Consider the frame $\mathfrak{N} = (\mathbb{N}, <)$ (the natural numbers in their usual ordering):



What is the ultrafilter extension of $\mathfrak{N}$? There are two kinds of ultrafilters over an infinite set: the principal ultrafilters that are in 1–1 correspondence with the points of the set, and the non-principal ones which contain all co-finite sets, and only infinite sets, cf. Exercise 2.5.4. We have just remarked (see (2.1)) that the principal ultrafilters form an isomorphic copy of the frame $\mathfrak{N}$ inside $\mathfrak{ue}\,\mathfrak{N}$. So where are the non-principal ultrafilters situated? The key fact here is that for any pair u, u' of ultrafilters, if u' is non-principal, then $R^{ue}uu'$. To see this, let u' be a non-principal ultrafilter, and let $X \in u'$. As X is infinite, for any $n \in \mathbb{N}$ there is an m such that $n < m$ and $m \in X$. This shows that $m_\diamond(X) = \mathbb{N}$. But $\mathbb{N}$ is an element of every ultrafilter u.

This shows that the ultrafilter extension of $\mathfrak{N}$ looks like a gigantic balloon at the end of an infinite string: it consists of a copy of $\mathfrak{N}$, followed by an large (uncountable) cluster consisting of all the non-principal ultrafilters:



$\dashv$

We will prove two results concerning ultrafilter extensions. The first one, Proposition 2.59, is an invariance result: any state in the original model is modally equivalent to the corresponding principal ultrafilter in the ultrafilter extension. Then, in Proposition 2.61 we show that ultrafilter extensions are m-saturated. Putting these two facts together leads us to the main result of this section: two states are modally equivalent iff their representatives in the ultrafilter extensions are bisimilar.

Proposition 2.59 *Let τ be a modal similarity type, and $\mathfrak{M}$ a τ-model. Then, for any formula ϕ and any ultrafilter u over W, $V(\phi) \in u$ iff $\mathfrak{ue}\,\mathfrak{M}, u \Vdash \phi$. Hence, for every state w of $\mathfrak{M}$ we have $w \leftrightsquigarrow \pi_w$.*

Proof. The second claim of the proposition is immediate from the first one by the observation that $w \Vdash \phi$ iff $w \in V(\phi)$ iff $V(\phi) \in \pi_w$.

The proof of the first claim is by induction on ϕ. The basic case is immediate from the definition of V^{ue}. The proofs of the boolean cases are straightforward consequences of the defining properties of ultrafilters. As an example, we treat negation; suppose that ϕ is of the form $\neg\psi$, then

$$\begin{array}{llll} V(\neg\psi) \in u & \text{iff} & W \setminus V(\psi) \in u & \\ & \text{iff} & V(\psi) \notin u & \\ & \text{iff} & \mathfrak{ue}\,\mathfrak{M}, u \nVdash \psi & \text{(induction hypothesis)} \\ & \text{iff} & \mathfrak{ue}\,\mathfrak{M}, u \Vdash \neg\psi. & \end{array}$$

Next, consider the case where ϕ is of the form $\Diamond\psi$ (we only treat the basic modal similarity type, leaving the general case as an exercise to the reader). Assume first that $\mathfrak{ue}\,\mathfrak{M}, u \Vdash \Diamond\psi$. Then, there is an ultrafilter u' such that $R^{ue}uu'$ and $\mathfrak{ue}\,\mathfrak{M}, u' \Vdash \psi$. The induction hypothesis implies that $V(\psi) \in u'$, so by the definition of R^{ue}, $m_\Diamond(V(\psi)) \in u$. Now the result follows immediately from the observation that $m_\Diamond(V(\psi)) = V(\Diamond\psi)$.

The left-to-right implication requires a bit more work. Assume that $V(\Diamond\psi) \in u$. We have to find an ultrafilter u' such that $V(\psi) \in u'$ and $R^{ue}uu'$. The latter constraint reduces to the condition that $m_\Diamond(X) \in u$ whenever $X \in u'$, or equivalently (see Exercise 2.5.5):

$$u'_0 := \{Y \mid m^\delta_\Diamond(Y) \in u\} \subseteq u'.$$

We will first show that u'_0 is closed under intersection. Let Y, Z be members of u'_0. By definition, $m^\delta_\Diamond(Y)$ and $m^\delta_\Diamond(Z)$ are in u. But then $m^\delta_\Diamond(Y \cap Z) \in u$, as $m^\delta_\Diamond(Y \cap Z) = m^\delta_\Diamond(Y) \cap m^\delta_\Diamond(Z)$, as a straightforward proof shows. This proves that $Y \cap Z \in u'_0$.

Next we make sure that for any $Y \in u'_0$, $Y \cap V(\psi) \neq \varnothing$. Let Y be an arbitrary element of u'_0, then by definition of u'_0, $m^\delta_\Diamond(Y) \in u$. As u is closed under intersection and does not contain the empty set, there must be an element x in $m^\delta_\Diamond(Y) \cap V(\Diamond\psi)$. But then x must have a successor y in $V(\psi)$. Finally, $x \in m^\delta_\Diamond(Y)$ implies $y \in Y$.

¿From the fact that u'_0 is closed under intersection, and the fact that for any $Y \in u'_0$, $Y \cap V(\psi) \neq \varnothing$, it follows that the set $u'_0 \cup \{V(\psi)\}$ has the finite intersection property. So the Ultrafilter Theorem (Fact A.14 in the Appendix) provides us with an ultrafilter u' such that $u'_0 \cup \{V(\psi)\} \subseteq u'$. This ultrafilter u' has the desired properties: it is clearly a successor of u, and the fact that $\mathfrak{ue}\,\mathfrak{M}, u' \Vdash \psi$ follows from $V(\psi) \in u'$ and the induction hypothesis. $\dashv$

Example 2.60 As with the invariance results of Section 2.1 (disjoint unions, generated submodels, and bounded morphisms), our new invariance result can be used to compare the relative expressive power of modal languages. Consider the modal constant $\circlearrowleft$ whose truth definition in a model for the basic modal language is

$$\mathfrak{M}, w \Vdash \circlearrowleft \text{ iff } \mathfrak{M} \models Rxx[v] \text{ for some } v \text{ in } \mathfrak{M}.$$

Can such a modality be defined in the basic modal language? No — a bisimulation based argument given at the end of the previous section already establishes this. Alternatively, we can see this by comparing the pictures of the frames $(\mathbb{N}, <)$ and its ultrafilter extension given in Example 2.58. The former is loop-free (thus in any model over this frame, $\mathfrak{ue}\,\mathfrak{M}, \pi_0 \nVdash \circlearrowleft$), but the later contains uncountably many loops (thus $\mathfrak{ue}\,\mathfrak{M}, \pi_0 \Vdash \circlearrowleft$). So if we want $\circlearrowleft$ we have to add it as a primitive. ⊣

Proposition 2.61 *Let τ be a modal similarity type, and let $\mathfrak{M}$ be a τ-model. Then $\mathfrak{ue}\,\mathfrak{M}$ is m-saturated.*

Proof. We only prove the proposition for the basic modal similarity type. Let $\mathfrak{M} = (W, R, V)$ be a model; we will show that its ultrafilter extension $\mathfrak{ue}\,\mathfrak{M}$ is m-saturated. Consider an ultrafilter u over W, and a set Σ of modal formulas which is finitely satisfiable in the set of successors of u. We have to find an ultrafilter u' such that $R^{ue}uu'$ and $\mathfrak{ue}\,\mathfrak{M}, u' \Vdash \Sigma$. Define

$$\Delta = \{V(\phi) \mid \phi \in \Sigma'\} \cup \{Y \mid m^{\delta}_{\Diamond}(Y) \in u\},$$

where Σ' is the set of (finite) conjunctions of formulas in Σ. We claim that the set Δ has the fip. Since both $\{V(\phi) \mid \phi \in \Sigma'\}$ and $\{Y \mid m^{\delta}_{\Diamond}(Y) \in u\}$ are closed under intersection, it suffices to prove that for an arbitrary $\phi \in \Sigma'$ and an arbitrary set $Y \subseteq W$ for which $m^{\delta}_{\Diamond}(Y) \in u$, we have $V(\phi) \cap Y \neq \varnothing$. But if $\phi \in \Sigma'$, then by assumption, there is a successor u'' of u such that $\mathfrak{ue}\,\mathfrak{M}, u'' \Vdash \phi$, or, in other words, $V(\phi) \in u''$. Then, $m^{\delta}_{\Diamond}(Y) \in u$ implies $Y \in u''$ by Exercise 2.5.5. Hence, $V(\phi) \cap Y$ is an element of the ultrafilter u'' and, therefore, cannot be identical to the empty set.

It follows by the Ultrafilter Theorem that Δ can be extended to an ultrafilter u'. Clearly, u' is the required successor of u in which Σ is satisfied. ⊣

We have finally arrived at the main result of this section: a characterization of modal equivalence as bisimilarity-somewhere-else — namely, between ultrafilter extensions.

Theorem 2.62 *Let τ be a modal similarity type, and let $\mathfrak{M}$ and $\mathfrak{M}'$ be τ-models, and w, w' two states in $\mathfrak{M}$ and $\mathfrak{M}'$, respectively. Then*

$$\mathfrak{M}, w \leftrightsquigarrow \mathfrak{M}', w' \textit{ iff } \mathfrak{ue}\,\mathfrak{M}, \pi_w \leftrightarroweq \mathfrak{ue}\,\mathfrak{M}', \pi_{w'}.$$

Proof. Immediate by Propositions 2.59, 2.61 and 2.54. ⊣

Three remarks. First, it is easy to define ultrafilter extensions and prove an analog of Theorem 2.62 for the basic temporal logic and arrow logic; see Exercises 2.5.8 and 2.5.9. With PDL the situation is a bit more complex; see Exercise 2.5.11. (The problem is that the property of one relation being the reflexive transitive closure of another is not preserved under taking ultrafilter extensions.) Second, we have not seen the last of ultrafilter extensions. Like disjoint unions, generated submodels, and bounded morphisms, ultrafilter extensions are a fundamental modal model construction technique, and we will make use of them when we discuss frames (in Chapter 3) and algebras (in Chapter 5). We will shortly see that ultrafilter extensions tie in neatly with ideas from first-order model theory — and we will use this to prove a second bisimilarity-somewhere-else result, Lemma 2.66. Finally, some readers may still have the feeling that taking the ultrafilter extension of a model is a far less natural construction than the other model operations that we have met. These readers are advised to hold on until (or take a peek ahead towards) Chapter 5, where we will see that ultrafilter extensions are indeed a very natural byproduct of modal logic's duality theory.

Exercises for Section 2.5

2.5.1 Let E be any subset of $\mathcal{P}(W)$, and let F be the filter generated by E.

(a) Prove that indeed, F is a filter over W. (Show that in general, the intersection of a collection of filters is again a filter.)
(b) Show that F is the set of all $X \in \mathcal{P}(W)$ such that either $X = W$ or for some Y_1, $\ldots, Y_n \in E$,

$$Y_1 \cap \cdots \cap Y_n \subseteq X.$$

(c) Prove that F is proper (that is: it does not coincide with $\mathcal{P}(W)$) iff E has the finite intersection property.

2.5.2 Let W be a non-empty set, and let w be an element of W. Show that the principal ultrafilter generated by w, that is, the set $\{X \in \mathcal{P}(W) \mid w \in X\}$, is indeed an ultrafilter over W.

2.5.3 Let F be a filter over W.

(a) Prove that F is an ultrafilter if and only if it is proper and maximal, that is, it has no proper extensions.
(b) Prove that F is an ultrafilter if and only if it is proper and for each pair of subsets X, Y of W we have that $X \cup Y \in F$ iff $X \in F$ or $Y \in F$.

2.5.4 Let W be an infinite set. Recall that $X \subseteq W$ is *co-finite* if $W \setminus X$ is finite.

(a) Prove that the collection of co-finite subsets of W has the finite intersection property.
(b) Show that there are ultrafilters over W that do not contain any finite set.

(c) Prove that an ultrafilter is non-principal if and only if it contains only infinite sets if and only if it contains all co-finite sets.
(d) Prove that any ultrafilter over W has uncountably many elements.

2.5.5 Given a model $\mathfrak{M} = (W, R, V)$ and two ultrafilters u and v over W, show that $R^{ue}uv$ if and only if $\{Y \mid m^{\delta}_{\Diamond}(Y) \in u\} \subseteq v$.

2.5.6 Let $\mathfrak{B} = (B, R)$ be the transitive binary tree; that is, B is the set of finite strings of 0s and 1s, and $R\sigma\tau$ holds if σ is a proper initial segment of τ. The aim of this exercise is to prove that any non-principal ultrafilter over B determines an *infinite* string of 0s and 1s.

To this end, let B^{ω} be the set of finite and infinite strings of 0s and 1s, and R^{ω} the relation on B^{ω} given by $R\sigma\tau$ if σ is an initial segment of τ. Define a function $f : Uf(B) \to B^{\omega}$ such that for all ultrafilters over B we have $uR^{ue}v$ iff $f(u)R^{\omega}f(v)$.

2.5.7 Give an example of a model $\mathfrak{M}$ which is point-generated while its ultrafilter extension is not.

2.5.8 Develop a notion of ultrafilter extension for basic temporal logic, and establish an analog of Theorem 2.62 for basic temporal logic.

2.5.9 Develop a notion of ultrafilter extension for the arrow language introduced in Example 1.14, and establish an analog of Theorem 2.62 for this language.

2.5.10 Show that, in general, first-order formulas are not preserved under ultrafilter extensions. That is, give a model $\mathfrak{M}$, a state w, and a first-order formula $\alpha(x)$ such that $\mathfrak{M} \models \alpha(x)[w]$, but $\mathfrak{ue}\,\mathfrak{M} \not\models \alpha(x)[\pi_w]$, where π_w is the principal ultrafilter generated by w.

2.5.11 Consider a modal similarity type with two diamonds, $\Diamond$ and $\langle * \rangle$, and take any model $\mathfrak{M} = (S, R, R_*, V)$ with

$$\begin{aligned} S &= \mathbb{N} \cup \{\infty\}, \\ R &= \{(n+1, n), (\infty, n) \mid n \in \mathbb{N}\}, \\ R_* &= \{(m, n) \mid m, n \in \mathbb{N}, m \geq n\} \cup (\{\infty\} \times S). \end{aligned}$$

Note that R_* is the reflexive transitive closure of R.

(a) Show that $\mathfrak{M}, \infty \Vdash \Box\langle * \rangle\Box\bot$.
(b) Let u be an arbitrary non-principal ultrafilter over S. Prove that $R^{ue}\pi_{\infty}u$.
(c) Let u be an arbitrary non-principal ultrafilter over S. Prove that u has an R^{ue}-successor in $\mathfrak{ue}\,\mathfrak{M}$, and that each of its R^{ue}-successors is again a non-principal ultrafilter.
(d) Now suppose that we add an new diamond $\langle \star \rangle$ to the language, and that in the model $\mathfrak{ue}\,\mathfrak{M}$ we take $R_{\star}$ to be the reflexive transitive closure of R^{ue}. Show that $\mathfrak{ue}\,\mathfrak{M}, \pi_{\infty} \Vdash \Diamond[\star]\Diamond\top$.
(e) Prove that $R^{ue}_{*} \neq R_{\star}$ (hint: use Proposition 2.59), and conclude that the ultrafilter extension of a regular PDL-model need not be a regular PDL-model.
(f) Prove that every non-principal ultrafilter over S has a *unique* R^{ue}-successor.

2.6 Characterization and Definability

In Section 2.3 we posed two important questions about modal expressivity:

(i) What is the modal fragment of first-order logic? That is, which first-order formulas are equivalent to the standard translation of a modal formula?
(ii) Which properties of models are definable by means of modal formulas?

In this, the first advanced track section of the book, we answer both questions. Our main tool will be a second characterization of modal equivalence as bisimilarity-somewhere-else, the Detour Lemma. Unlike the characterization just proved (Theorem 2.62), the Detour Lemma rests on a number of non-modal concepts and results, all of which are centered on *saturated models* (a standard concept of first-order model theory). We start by introducing saturated models and use them to describe the modal fragment of first-order logic. After that we show how to build saturated models. As corollaries we obtain results on modally definable properties of models. For background information on first-order model theory, see Appendix A.

The Van Benthem Characterization Theorem

To define the notion of saturated models, we need the concept of α-saturation, but before giving a formal definition of the latter, we provide an informal description, which the reader may want to use as a 'working' definition.

Informally, then, the notion of α-saturation can be explained as follows. First of all, let $\Gamma(x)$ be a set of first-order formulas in which a single individual variable x may occur free — such a set of formulas is called a *type*. A first-order model $\mathfrak{M}$ *realizes* $\Gamma(x)$ if there is an element w in $\mathfrak{M}$ such that for all $\gamma \in \Gamma$, $\mathfrak{M} \models \gamma[w]$.

Next, let $\mathfrak{M}$ be a model for a given first-order language $\mathcal{L}^1$ with domain W. For a subset $A \subseteq W$, $\mathcal{L}^1[A]$ is the language obtained by extending $\mathcal{L}^1$ with new constants $\underline{a}$ for all elements $a \in A$. $\mathfrak{M}_A$ is the expansion of $\mathfrak{M}$ to a structure for $\mathcal{L}^1[A]$ in which each $\underline{a}$ is interpreted as a.

Assume that A is of size at most α. For the sake of our informal definition of α-saturation, assume that $\alpha = 3$ and $A = \{a_1, a_2\}$. Let $\Gamma(\underline{a}_1, \underline{a}_2, x)$ be a type of the language $\mathcal{L}^1[A]$; it is not difficult to see that $\Gamma(\underline{a}_1, \underline{a}_2, x)$ is consistent with the first-order theory of $\mathfrak{M}_A$ iff $\Gamma(\underline{a}_1, \underline{a}_2, x)$ is finitely realizable in $\mathfrak{M}_A$, (that is, $\mathfrak{M}_A$ realizes every *finite* subset Δ of $\Gamma(\underline{a}_1, \underline{a}_2, x)$). So, for this particular set $\Gamma(\underline{a}_1, \underline{a}_2, x)$, 3-saturation of $\mathfrak{M}$ means that if $\Gamma(\underline{a}_1, \underline{a}_2, x)$ is finitely realizable in $\mathfrak{M}_A$, then $\Gamma(\underline{a}_1, \underline{a}_2, x)$ is realizable in $\mathfrak{M}_A$.

Yet another way of looking at 3-saturation for this particular set of formulas is the following. Consider a formula $\gamma(\underline{a}_1, \underline{a}_2, x)$, and let $\gamma(x_1, x_2, x)$ be the formula with the fresh variables x_1 and x_2 replacing each occurrence in γ of $\underline{a}_1$ and $\underline{a}_2$, respectively. Then we have the following equivalence:

$\mathfrak{M}_A$ realizes $\{\gamma(\underline{a}_1, \underline{a}_2, x)\}$ iff there is a b such that $\mathfrak{M} \models \gamma(x_1, x_2, x)[a_1, a_2, b]$.

So, a model is α-saturated iff the following holds for every $n < \alpha$, and every set Γ of formulas of the form $\gamma(x_1, \ldots, x_n, x)$.

> *If* $(a_1, \ldots, a_n)$ is an n-tuple such that for every *finite* $\Delta \subseteq \Gamma$ there is a b_Δ such that $\mathfrak{M} \models \gamma(x_1, \ldots, x_n, x)[a_1, \ldots, a_n, b_\Delta]$ for every $\gamma \in \Delta$,
> *then* we have that there is a b such that $\mathfrak{M} \models \gamma(x_1, \ldots, x_n, x)[a_1, \ldots, a_n, b]$ for every $\gamma \in \Gamma$.

This way of looking at α-saturation is useful, for it makes the analogy with m-saturation of the previous section clear. Both m-saturated and countably saturated models are rich in the number of types $\Gamma(x)$ they realize, but the latter are far richer than the former: they realize the maximum number of types.

Now, for the 'official' definition of α-saturation.

Definition 2.63 Let α be a natural number, or ω. A model $\mathfrak{M}$ is α-*saturated* if for every subset $A \subseteq W$ of size less than α, the expansion $\mathfrak{M}_A$ realizes every set $\Gamma(x)$ of $\mathcal{L}^1[A]$-formulas (with only x occurring free) that is consistent with the first-order theory of $\mathfrak{M}_A$. An ω-saturated model is usually called *countably saturated*. $\dashv$

Example 2.64 (i) Every finite model is countably saturated. For, if $\mathfrak{M}$ is finite, and $\Gamma(x)$ is a set of first-order formulas consistent with the first-order theory of $\mathfrak{M}$, there exists a model $\mathfrak{N}$ that is elementarily equivalent to $\mathfrak{M}$ and that realizes $\Gamma(x)$. But, as $\mathfrak{M}$ and $\mathfrak{N}$ are finite, elementary equivalence implies isomorphism, and hence $\Gamma(x)$ is realized in $\mathfrak{M}$.

(ii) The ordering of the rational numbers $(\mathbb{Q}, <)$ is countably saturated as well. The relevant first-order language $\mathcal{L}^1$ has $<$ and $=$. Take a subset A of $\mathbb{Q}$ and let $\Gamma(x)$ be a set of formulas in the resulting expansion $\mathcal{L}^1[A]$ of this first-order language that is consistent with the theory of $(\mathbb{Q}, <, a)_{a \in A}$. Then, there exists a model $\mathfrak{N}$ of the theory of $(\mathbb{Q}, <, a)_{a \in A}$ that realizes $\Gamma(x)$. Now take a countable elementary submodel $\mathfrak{N}'$ of $\mathfrak{N}$ that contains at least one object realizing $\Gamma(x)$. Then $\mathfrak{N}'$ is a countable dense linear ordering without endpoints, and hence the ordering of $\mathfrak{N}'$ is isomorphic to $(\mathbb{Q}, <)$. The interpretations (in $\mathfrak{N}$) of the constants $\underline{a}$ for elements a in A may be copied across to $\mathfrak{N}'$. Hence, as $\mathfrak{N}$ realizes $\Gamma(x)$, so does $\mathfrak{N}'$, and hence, so does $(\mathbb{Q}, <)$, as required.

(iii) The ordering of the natural numbers $(\mathbb{N}, <)$ is not countably saturated. To see this, consider the following set of formulas.

$$\Gamma(x) \quad := \quad \{\exists y_1\, (y_1 < x), \ldots, \exists y_1 \ldots y_n\, (y_1 < \cdots < y_n < x), \ldots\}.$$

$\Gamma(x)$ is clearly consistent with the theory of $(\mathbb{N}, <)$ as each of its finite subsets is realizable in $(\mathbb{N}, <)$. Yet, $\Gamma(x)$ is clearly not realizable in $(\mathbb{N}, <)$. $\dashv$

The following result explains why countably saturated models matter to us.

Theorem 2.65 *Let τ be a modal similarity type. Any countably saturated τ-model is m-saturated. It follows that the class of countably saturated τ-models has the Hennessy-Milner property.*

Proof. We only consider the basic modal language. Assume that $\mathfrak{M} = (W, R, V)$, viewed as a first-order model, is countably saturated. Let a be a state in W, and consider a set Σ of modal formulas which is finitely satisfiable in the successor set of a. Define Σ' to be the set

$$\Sigma' = \{R\underline{a}x\} \cup ST_x(\Sigma),$$

where $ST_x(\Sigma)$ is the set $\{ST_x(\phi) \mid \phi \in \Sigma\}$ of standard translations of formulas in Σ. Clearly, Σ' is consistent with the first-order theory of $\mathfrak{M}_a$: $\mathfrak{M}_a$ realizes every finite subset of Σ', namely in some successor of a. So, by the countable saturation of $\mathfrak{M}$, Σ' itself is realized in some state b. By $\mathfrak{M}_a \models R\underline{a}x[b]$ it follows that b is a successor of a. Then, by Theorem 2.47 and the fact that $\mathfrak{M}_a \models ST_x(\phi)[b]$ for all $\phi \in \Sigma$, it follows that $\mathfrak{M}, b \Vdash \Sigma$. Thus Σ is satisfiable in a successor of a. $\dashv$

In fact, we only need 2-saturation for the proof of Theorem 2.65 to go through. This is because we restricted ourselves to the *basic* modal similarity type. We leave it to the reader to check to which extent the 'amount of saturation' needed to make the proof of Theorem 2.65 go through depends on the rank of the operators of the similarity type.

We have yet to show that countably saturated models actually exist; this issue will be addressed below (see Theorem 2.74). For now, we merely want to record the following important use of saturated models; you may want to recall the definition of an elementary embedding before reading the result (see Appendix A)).

Lemma 2.66 (Detour Lemma) *Let τ be a modal similarity type, and let $\mathfrak{M}$ and $\mathfrak{N}$ be τ-models, and w and v states in $\mathfrak{M}$ and $\mathfrak{N}$, respectively. Then the following are equivalent.*

(i) *For all modal formulas ϕ: $\mathfrak{M}, w \Vdash \phi$ iff $\mathfrak{N}, v \Vdash \phi$.*
(ii) *There exists a bisimulation $Z : \mathfrak{ue}\,\mathfrak{M}, \pi_w \leftrightarrows \mathfrak{ue}\,\mathfrak{N}, \pi_v$.*
(iii) *There exist countably saturated models $\mathfrak{M}^*, w^*$ and $\mathfrak{N}^*, v^*$ and elementary embeddings $f : \mathfrak{M} \preccurlyeq \mathfrak{M}^*$ and $g : \mathfrak{N} \preccurlyeq \mathfrak{N}^*$ such that*
   (a) $f(w) = w^*$ *and* $g(v) = v^*$
   (b) $\mathfrak{M}^*, w^* \leftrightarrows \mathfrak{N}^*, v^*$.

What does the Detour Lemma say in words? Obviously (i) ⇒ (ii) is just our old bisimulation-somewhere-else result (Theorem 2.62). The key new part is the implication (i) ⇒ (iii). This says that if $\mathfrak{M}, w$ and $\mathfrak{N}, v$ are modally equivalent, then both can be extended — more accurately: elementarily extended — to countably saturated models $\mathfrak{M}^*, w^*$ and $\mathfrak{N}^*, v^*$. As $\mathfrak{M}, w$ and $\mathfrak{N}, v$ were modally equivalent, so are $\mathfrak{M}^*, w^*$ and $\mathfrak{N}^*, v^*$; it follows by Theorem 2.65 that the latter two models are bisimilar. In short, this is a second 'bisimilarity somewhere else' result, this time the 'somewhere else' being 'in some suitable ultrapower'. Notice that in order to prove the Detour Lemma all we need to establish is that every model can be elementarily embedded in a countably saturated model — there are standard first-order techniques for doing so, and we will introduce one in the second half of this section.

With the help of the Detour Lemma, we can now precisely characterize the relation between first-order logic, modal logic, and bisimulations. To prove the theorem we need to explicitly define a concept which we have already invoked informally on several occasions.

Definition 2.67 A first-order formula $\alpha(x)$ in $\mathcal{L}^1_\tau$ is *invariant for bisimulations* if for all models $\mathfrak{M}$ and $\mathfrak{N}$, and all states w in $\mathfrak{M}$, v in $\mathfrak{N}$, and all bisimulations Z between $\mathfrak{M}$ and $\mathfrak{N}$ such that wZv, we have $\mathfrak{M} \models \alpha(x)[w]$ iff $\mathfrak{N} \models \alpha(x)[v]$. $\dashv$

Theorem 2.68 (Van Benthem Characterization Theorem) *Let $\alpha(x)$ be a first-order formula in $\mathcal{L}^1_\tau$. Then $\alpha(x)$ is invariant for bisimulations iff it is (equivalent to) the standard translation of a modal τ-formula.*

Proof. The direction from right to left is a consequence of Theorem 2.20. To prove the direction from left to right, assume that $\alpha(x)$ is invariant for bisimulations and consider the set of modal consequences of α:

$$\mathrm{MOC}(\alpha) = \{ST_x(\phi) \mid \phi \text{ is a modal formula, and } \alpha(x) \models ST_x(\phi)\}.$$

Our first claim is that if $\mathrm{MOC}(\alpha) \models \alpha(x)$, then $\alpha(x)$ is equivalent to the translation of a modal formula. To see why this is so, assume that $\mathrm{MOC}(\alpha) \models \alpha(x)$; then, by the Compactness Theorem for first-order logic, for some finite subset $X \subseteq \mathrm{MOC}(\alpha)$ we have $X \models \alpha(x)$. So $\models \bigwedge X \to \alpha(x)$. Trivially $\models \alpha(x) \to \bigwedge X$, thus $\models \alpha(x) \leftrightarrow \bigwedge X$. And as every $\beta \in X$ is the translation of a modal formula, so is $\bigwedge X$. This proves our claim.

So it suffices to show that $\mathrm{MOC}(\alpha) \models \alpha(x)$. Assume $\mathfrak{M} \models \mathrm{MOC}(\alpha)[w]$; we need to show that $\mathfrak{M} \models \alpha(x)[w]$. Let

$$T(x) = \{ST_x(\phi) \mid \mathfrak{M} \models ST_x(\phi)[w]\}.$$

We claim that $T(x) \cup \{\alpha(x)\}$ is consistent. Why? Assume, for the sake of contradiction, that $T(x) \cup \{\alpha(x)\}$ is inconsistent. Then, by compactness, for some finite subset $T_0(x) \subseteq T(x)$ we have $\models \alpha(x) \to \neg \bigwedge T_0(x)$. Hence $\neg \bigwedge T_0(x) \in \mathrm{MOC}(\alpha)$. But this implies $\mathfrak{M} \models \neg \bigwedge T_0(x)[w]$, which contradicts $T_0(x) \subseteq T(x)$ and $\mathfrak{M} \models T(x)[w]$.

So, let $\mathfrak{N}, v$ be such that $\mathfrak{N} \models T(x) \cup \{\alpha(x)\}[v]$. Observe that w and v are modally equivalent: $\mathfrak{M}, w \Vdash \phi$ implies $ST_x(\phi) \in T(x)$, which implies $\mathfrak{N}, v \Vdash \phi$; and likewise, if $\mathfrak{M}, w \nVdash \phi$ then $\mathfrak{M}, w \Vdash \neg\phi$, and $\mathfrak{N}, v \Vdash \neg\phi$. If modal equivalence implied bisimilarity we would be done, because then $\mathfrak{M}, w$ and $\mathfrak{N}, v$ would be bisimilar, and from this we would be able to deduce the desired conclusion $\mathfrak{M}, w \models \alpha(x)[w]$ by invariance under bisimulation. But, in general, modal equivalence does not imply bisimilarity, so this is not a sound argument.

However, we can use the Detour Lemma and make a detour through a Hennessy-Milner class where modal equivalence and bisimilarity do coincide! More precisely, the Detour Lemma yields two countably saturated models $\mathfrak{M}^*, w^* \succcurlyeq \mathfrak{M}, w$ and $\mathfrak{N}^*, v^* \succcurlyeq \mathfrak{N}, v$ such that $\mathfrak{M}^*, w^* \leftrightarroweq \mathfrak{N}^*, v^*$:

$$\begin{array}{ccc} \mathfrak{M}, w & & \mathfrak{N}, v \\ \preccurlyeq \big| & & \big| \preccurlyeq \\ \mathfrak{M}^*, w^* & \leftrightarroweq & \mathfrak{N}^*, v^*. \end{array}$$

This is where we really need the new characterization of modal equivalence in terms of bisimulation-somewhere-else that Theorem 2.74 gives us. We need to 'lift' the first-order formula $\alpha(x)$ from the model $\mathfrak{N}, v$ to the model $\mathfrak{N}^*, v^*$. By definition, the truth of first-order formulas is preserved under elementary embeddings, so that this can indeed be done. However, first-order formulas need not be preserved under ultrafilter extensions (see Exercise 2.5.10), and for that reason we cannot use the ultrafilter extension $\mathfrak{ue}\,\mathfrak{N}, \pi_v$ instead of $\mathfrak{N}^*, v^*$.

Returning to the main argument, $\mathfrak{N} \models \alpha(x)[v]$ implies $\mathfrak{N}^* \models \alpha(x)[v^*]$. As $\alpha(x)$ is invariant for bisimulations, we get $\mathfrak{M}^* \models \alpha(x)[w^*]$. By invariance under elementary embeddings, we have $\mathfrak{M} \models \alpha(x)[w]$. This proves the theorem. $\dashv$

Ultraproducts

The preceding discussion left us with an important technical question: how do we get countably saturated models? Our next aim is to answer this question and thereby prove the Detour Lemma.

The fundamental construction underlying our proof is that of an ultraproduct. Here we briefly recall the basic ideas; further details may be found in Appendix A.

We first apply the construction to sets, and then to models. Suppose $I \neq \varnothing$, U is an ultrafilter over I, and for each $i \in I$, W_i is a non-empty set. Let $C = \prod_{i \in I} W_i$ be the Cartesian product of those sets. That is: C is the set of all functions f with domain I such that for each $i \in I$, $f(i) \in W_i$. For two functions $f, g \in C$ we say that f and g are U-*equivalent* (notation $f \sim_U g$) if $\{i \in I \mid f(i) = g(i)\} \in U$. The result is that $\sim_U$ is an equivalence relation on the set C.

Definition 2.69 (Ultraproduct of Sets) Let f_U be the equivalence class of f modulo $\sim_U$, that is: $f_U = \{g \in C \mid g \sim_U f\}$. The *ultraproduct of* W_i *modulo* U is the set of all equivalence classes of $\sim_U$. it is denoted by $\prod_U W_i$. So

$$\prod_U W_i = \{f_U \mid f \in \prod_{i \in I} W_i\}.$$

In the case where all the sets are the same, say $W_i = W$ for all i, the ultraproduct is called the *ultrapower of* W *modulo* U, and written $\prod_U W$. ⊣

Following the general definition of the ultraproduct of first-order models (Definition A.17), we now define the ultraproduct of modal models.

Definition 2.70 (Ultraproduct of Models) Fix a modal similarity type τ, and let $\mathfrak{M}_i$ ($i \in I$) be τ-models. The *ultraproduct* $\prod_U \mathfrak{M}_i$ *of* $\mathfrak{M}_i$ *modulo* U is the model described as follows.

(i) The universe W_U of $\prod_U \mathfrak{M}_i$ is the set $\prod_U W_i$, where W_i is the universe of $\mathfrak{M}_i$.

(ii) Let V_i be the valuation of $\mathfrak{M}_i$. Then the valuation V_U of $\prod_U \mathfrak{M}_i$ is defined by

$$f_U \in V_U(p) \text{ iff } \{i \in I \mid f(i) \in V_i(p)\} \in U.$$

(iii) Let $\triangle$ be a modal operator in τ, and $R_{\triangle i}$ its associated relation in the model $\mathfrak{M}_i$. The relation $R_{\triangle U}$ in $\prod_U \mathfrak{M}_i$ is given by

$$R_{\triangle U} f_U^1 \dots f_U^{n+1} \text{ iff } \{i \in I \mid R_{\triangle i} f^1(i) \dots f^{n+1}(i)\} \in U.$$

In particular, for a diamond item (iii) boils down to

$$R_{\diamond U} f_U g_U \text{ iff } \{i \in I \mid R_{\diamond i} f(i) g(i)\} \in U. \quad \dashv$$

To show that the above definition is consistent, we should check that V_U and R_U depend only on the equivalence classes $f_U^1, \dots, f_U^{n+1}$.

Proposition 2.71 *Let* $\prod_U \mathfrak{M}$ *be an ultrapower of* $\mathfrak{M}$. *Then, for all modal formulas* ϕ *we have* $\mathfrak{M}, w \Vdash \phi$ *iff* $\prod_U \mathfrak{M}, (f_w)_U \Vdash \phi$, *where* f_w *is the constant function such that* $f_w(i) = w$, *for all* $i \in I$.

Proof. This is left as Exercise 2.6.1. ⊣

To build countably saturated models, we use ultraproducts based on a special kind of ultrafilters. An ultrafilter is *countably incomplete* if it is not closed under countable intersections (of course, it will be closed under finite intersections).

Example 2.72 Consider the set of natural numbers $\mathbb{N}$. Let U be an ultrafilter over $\mathbb{N}$ that does not contain any singletons $\{n\}$. (The reader is asked to prove that such ultrafilters exist in Exercise 2.5.4.) Then, for all n, $(\mathbb{N} \setminus \{n\}) \in U$. But

$$\varnothing = \bigcap_{n \in \mathbb{N}} (\mathbb{N} \setminus \{n\}) \notin U.$$

So U is countably incomplete. $\dashv$

Lemma 2.73 *Let $\mathcal{L}$ be a countable first-order language, U a countably incomplete ultrafilter over a non-empty set I, and $\mathfrak{M}$ an $\mathcal{L}$-model. The ultrapower $\prod_U \mathfrak{M}$ is countably saturated.*

Proof. See Appendix A. $\dashv$

We are now ready to prove the Detour Lemma. In Theorem 2.62 we showed that 'bisimulation somewhere else' can mean 'in the ultrafilter extension'. Now we will show that it can also mean: 'in a suitable ultrapower of the original models.'

Theorem 2.74 *Let τ be a modal similarity type, and let $\mathfrak{M}$ and $\mathfrak{N}$ be τ-models, and w and v states in $\mathfrak{M}$ and $\mathfrak{N}$, respectively. Then the following are equivalent.*

(i) *For all modal formulas ϕ: $\mathfrak{M}, w \Vdash \phi$ iff $\mathfrak{N}, v \Vdash \phi$.*
(ii) *There exist ultrapowers $\prod_U \mathfrak{M}$ and $\prod_U \mathfrak{N}$ and as well as a bisimulation $Z : \prod_U \mathfrak{M}, (f_w)_U \leftrightarroweq \prod_U \mathfrak{N}, (f_v)_U$ linking $(f_w)_U$ and $(f_v)_U$, where f_w (f_v) is the constant function mapping every index to w (v).*

Proof. It is easy to see that (ii) implies (i). By Proposition 2.71 $\mathfrak{M}, w \Vdash \phi$ iff $\prod_U \mathfrak{M}, (f_w)_U \Vdash \phi$. By assumption this is equivalent to $\prod_U \mathfrak{N}, (f_v)_U \Vdash \phi$, and the latter is equivalent to $\mathfrak{N}, v \Vdash \phi$.

To prove the implication from (i) to (ii) we have to do some more work. Assume that for all modal formulas ϕ we have $\mathfrak{M}, w \Vdash \phi$ iff $\mathfrak{N}, v \Vdash \phi$. We need to create bisimilar ultrapowers of $\mathfrak{M}$ and $\mathfrak{N}$.

Take the set of natural numbers $\mathbb{N}$ as our index set, and let U be a countably incomplete ultrafilter over $\mathbb{N}$ (cf. Example 2.72). By Lemma 2.73 the ultrapowers $\prod_U \mathfrak{M}$ and $\prod_U \mathfrak{N}$ are countably saturated. Now $(f_w)_U$ and $(f_v)_U$ are modally equivalent: for all modal formulas ϕ, $\prod_U \mathfrak{M}, (f_w)_U \Vdash \phi$ iff $\prod_U \mathfrak{N}, (f_v)_U \Vdash \phi$. This claim follows from the assumption that w and v are modally equivalent together with Proposition 2.71. Next, apply Theorem 2.65: as $(f_w)_U$ and $(f_v)_U$ are modally equivalent and $\prod_U \mathfrak{M}$ and $\prod_U \mathfrak{N}$ are countably saturated, there exists a bisimulation $Z : \prod_U \mathfrak{M}, (f_w)_U \leftrightarroweq \prod_U \mathfrak{N}, (f_v)_U$. This proves the theorem. $\dashv$

We obtain the Detour Lemma as an immediate corollary of Theorem 2.74 and Theorem 2.62.

Definability

Our next aim is to answer the second of the two questions posed at the start of this section: which properties of models are definable by means of modal formulas? Like the Detour Lemma, the answer is a corollary of Theorem 2.74. We formulate the result in terms of *pointed models*. Given a modal similarity type τ, a pointed model is a pair $(\mathfrak{M}, w)$ where $\mathfrak{M}$ is a τ-model and w is a state of $\mathfrak{M}$. Although the results below can also be given for models, the use of pointed models allows for a smoother formulation, mainly because pointed models reflect the local way in which modal formulas are evaluated.

We need some further definitions. A class of pointed models K is said to be *closed under bisimulations* if $(\mathfrak{M}, w)$ in K and $\mathfrak{M}, w \leftrightarroweq \mathfrak{N}, v$ implies $(\mathfrak{N}, v)$ in K. K is *closed under ultraproducts* if any ultraproduct $\prod_U(\mathfrak{M}_i, w_i)$ of a family of pointed models $(\mathfrak{M}_i, w_i)$ in K belongs to K. If K is a class of pointed τ-models, $\overline{\mathsf{K}}$ denotes the complement of K within the class of all pointed τ-models. Finally, K is *definable by a set of modal formulas* if there is a set of modal formulas Γ such that for any pointed model $(\mathfrak{M}, w)$ we have $(\mathfrak{M}, w)$ in K iff for all $\gamma \in \Gamma$, $\mathfrak{M}, w \Vdash \gamma$; K is definable by a single modal formula iff it is definable by a singleton set.

By Proposition 2.47 definable classes of pointed models must be closed under bisimulations, and by Corollary A.20 they must be closed under ultraproducts as well. Theorems 2.75 and 2.76 below show that these two closure conditions suffice to completely describe the classes of pointed models that are definable by means of modal formulas.

Theorem 2.75 *Let τ be a modal similarity type, and K a class of pointed τ-models. Then the following are equivalent.*

(i) K *is definable by a set of modal formulas.*
(ii) K *is closed under bisimulations and ultraproducts, and* $\overline{\mathsf{K}}$ *is closed under ultrapowers.*

Proof. The implication from (i) to (ii) is easy. For the converse, assume K and $\overline{\mathsf{K}}$ satisfy the stated closure conditions. Observe that $\overline{\mathsf{K}}$ is closed under bisimulations, as K is. Define T as the set of modal formulas holding in K:

$$T = \{\phi \mid \text{for all } (\mathfrak{M}, w) \text{ in } \mathsf{K}\text{: } \mathfrak{M}, w \Vdash \phi\}.$$

We will show that T defines the class K. First of all, by definition every pointed model $(\mathfrak{M}, w)$ in K is a model satisfying T in the sense that $\mathfrak{M}, w \Vdash T$. Second, assume that $\mathfrak{M}, w \Vdash T$; to complete the proof of the theorem we show that $(\mathfrak{M}, w)$ must be in K.

Define Σ to be the modal theory of w; that is, $\Sigma = \{\phi \mid \mathfrak{M}, w \Vdash \phi\}$. It is obvious that Σ is finitely satisfiable in K; for suppose that the set $\{\sigma_1, \ldots, \sigma_n\} \subseteq \Sigma$ is not satisfiable in K. Then the formula $\neg(\sigma_1 \wedge \cdots \wedge \sigma_n)$ would be true on all pointed models in K, so it would belong to T, yet be false in $\mathfrak{M}, w$. But then the following claim shows that Σ is satisfiable in the ultraproduct of pointed models in K.

Claim 1 *Let Σ be a set of modal formulas, and K a class of pointed models in which Σ is finitely satisfiable. Then Σ is satisfiable in some ultraproduct of models in K.*

Proof of Claim. Define an index set I as the collection of all finite subsets of Σ:

$$I = \{\Sigma_0 \subseteq \Sigma \mid \Sigma_0 \text{ is finite}\}.$$

By assumption, for each $i \in I$ there is a pointed model $(\mathfrak{N}_i, v_i)$ in K such that $\mathfrak{N}_i, v_i \Vdash i$. We now construct an ultrafilter U over I such that the ultraproduct $\prod_U \mathfrak{N}_i$ has a state f_U with $\prod_U \mathfrak{N}_i, f_U \Vdash \Sigma$.

For each $\sigma \in \Sigma$, let $\widehat{\sigma}$ be the set of all $i \in I$ such that $\sigma \in i$. Then the set $E = \{\widehat{\sigma} \mid \sigma \in \Sigma\}$ has the finite intersection property because

$$\{\sigma_1, \ldots, \sigma_n\} \in \widehat{\sigma}_1 \cap \cdots \cap \widehat{\sigma}_n.$$

So, by Fact A.14, E can be extended to an ultrafilter U over I. This defines $\prod_U \mathfrak{N}_i$; for the definition of f_U, let W_i denote the universe of the model $\mathfrak{N}_i$ and consider the function $f \in \prod_{i \in I} W_i$ such that $f(i) = v_i$.

It is left to prove that

$$\prod_U \mathfrak{N}_i, f_U \Vdash \Sigma. \tag{2.2}$$

To prove (2.2), observe that for $i \in \widehat{\sigma}$ we have $\sigma \in i$, and so $\mathfrak{N}_i, v_i \Vdash \sigma$. Therefore, for each $\sigma \in \Sigma$

$$\{i \in I \mid \mathfrak{N}_i, v_i \Vdash \sigma\} \supseteq \widehat{\sigma} \text{ and } \widehat{\sigma} \in U.$$

It follows that $\{i \in I \mid \mathfrak{N}_i, v_i \Vdash \sigma\} \in U$, so by Theorem A.19, $\prod_U \mathfrak{N}_i, f_U \Vdash \sigma$. This proves (2.2). $\dashv$

It follows from Claim 1 and the closure of K under taking ultraproducts that Σ is satisfiable in some pointed model $(\mathfrak{N}, v)$ in K. But $\mathfrak{N}, v \Vdash \Sigma$ implies that v and the state w from our original pointed model $(\mathfrak{M}, w)$ are modally equivalent. So by Theorem 2.74 there exists an ultrafilter U' such that

$$\prod_{U'}(\mathfrak{N}, v), (f_v)_U \leftrightarroows \prod_{U'}(\mathfrak{M}, w), (f_w)_U.$$

By closure under ultraproducts, the pointed model $(\prod_{U'}(\mathfrak{N}, v), (f_v)_U)$ belongs to K. Hence by closure under bisimulations, $(\prod_{U'}(\mathfrak{M}, w), (f_w)_U)$ is in K as well. By closure of $\overline{\mathsf{K}}$ under ultrapowers it follows that $(\mathfrak{M}, w)$ is in K. This completes the proof. $\dashv$

Theorem 2.76 *Let τ be a modal similarity type, and K a class of pointed τ-models. Then the following are equivalent.*

(i) K *is definable by means of a single modal formula.*
(ii) *Both* K *and* $\overline{\mathsf{K}}$ *are closed under bisimulations and ultraproducts.*

Proof. The direction from (i) to (ii) is easy. For the converse we assume that K, $\overline{\mathsf{K}}$ satisfy the stated closure conditions. Then both are closed under ultraproducts, hence by Theorem 2.75 there are sets of modal formulas T_1, T_2 defining K and $\overline{\mathsf{K}}$, respectively. Obviously their union is inconsistent in the sense that there is no pointed model $(\mathfrak{M}, w)$ such that $(\mathfrak{M}, w) \Vdash T_1 \cup T_2$. So then, by compactness, there exist $\phi_1, \ldots, \phi_n \in T_1$ and $\psi_1, \ldots, \psi_m \in T_2$ such that for all pointed models $(\mathfrak{M}, w)$

$$\mathfrak{M}, w \Vdash \phi_1 \wedge \cdots \wedge \phi_n \rightarrow \neg\psi_1 \vee \cdots \vee \neg\psi_m. \tag{2.3}$$

To complete the proof we show that K is in fact defined by the conjunction $\phi_1 \wedge \cdots \wedge \phi_n$. By definition, for any $(\mathfrak{M}, w)$ in K we have $\mathfrak{M}, w \Vdash \phi_1 \wedge \cdots \wedge \phi_n$. Conversely, if $\mathfrak{M}, w \Vdash \phi_1 \wedge \cdots \wedge \phi_n$, then, by (2.3), $\mathfrak{M}, w \Vdash \neg\psi_1 \vee \cdots \vee \neg\psi_m$. Hence, $\mathfrak{M}, w \nVdash T_2$. Therefore, $(\mathfrak{M}, w)$ does not belong to $\overline{\mathsf{K}}$, whence $(\mathfrak{M}, w)$ belongs to K. ⊣

Theorems 2.75 and 2.76 correspond to analogous definability results in first-order logic: to get the analogous first-order results, simply replace closure under bisimulations in 2.75 and 2.76 by closure under isomorphisms; see the Notes at the end of the chapter for further details. This close connection to first-order logic may explain why the results of this section seem to generalize to any modal logic that has a standard translation into first-order logic. For example, all of the results of this section can also be obtained for basic temporal logic.

Exercises for Section 2.6

2.6.1 Prove Proposition 2.71: Let $\prod_U \mathfrak{M}$ be an ultrapower of $\mathfrak{M}$. Then, for all modal formulas ϕ we have $\mathfrak{M}, w \Vdash \phi$ iff $\prod_U \mathfrak{M}, (f_w)_U \Vdash \phi$, where f_w is the constant function such that $f_w(i) = w$, for all $i \in I$.

2.6.2 Give simple proofs of Theorem 2.75 and Theorem 2.76 using the analogous proof for first-order logic (see Theorem A.23).

2.6.3 Let I be an index set, and let $\{\mathfrak{M}_i\}_{i \in I}$ and $\{\mathfrak{N}_i\}_{i \in I}$ be two collections of models such that for each $i \in I$, $\mathfrak{M}_i \leftrightarrows \mathfrak{N}_i$. Show that for any ultrafilter U over I, the ultraproducts of the two collections are bisimilar: $\prod_U \mathfrak{M}_i \leftrightarrows \prod_U \mathfrak{N}_i$.

2.6.4 (a) Show that the ultraproduct of point-generated models need not be point-generated.
(b) How is this for transitive models?

2.7 Simulation and Safety

Theorem 2.68 provided a result characterizing the modal fragment of first-order logic as the class of formulas invariant for bisimulations. In this section we present two further results in the same spirit; we focus on these results not just because they are interesting and typical of current work in modal model theory, but also because they provide instructive examples of how to apply the tools and proof strategies we have discussed. We first look at a notion of simulation that has been introduced in various settings, and characterize the modal formulas preserved by simulations. We then examine a question that arises in the setting of dynamic logic and process algebra: which operations on models preserve bisimulation? That is, if we have the back-and-forth clauses holding for R, and we apply an operation O to R which returns a new relation $O(R)$, then when do we also have the back-and-forth-clauses for $O(R)$?

Simulations

A simulation is simply a bisimulation from which half of the atomic clause and the back clause have been omitted.

Definition 2.77 (Simulations) Let τ be a modal similarity type. Let $\mathfrak{M} = (W, R_\triangle, V)_{\triangle\in\tau}$ and $\mathfrak{M}' = (W', R'_\triangle, V')_{\triangle\in\tau}$ be τ-models. A non-empty binary relation $Z \subseteq W \times W'$ is called a τ-*simulation* from $\mathfrak{M}$ to $\mathfrak{M}'$ if the following conditions are satisfied.

(i) If wZw' and $w \in V(p)$, then $w' \in V'(p)$.
(ii) If wZw' and $R_\triangle wv_1 \dots v_n$ then there are $v'_1, \dots, v'_n$ (in W') such that $R'_\triangle w'v'_1 \dots v'_n$ and for all i ($1 \leq i \leq n$) $v_iZv'_i$.

Thus, simulations only require that atomic information is preserved and that the forth condition holds.

If Z is a simulation from w in $\mathfrak{M}$ to w' in $\mathfrak{M}'$, we write $Z : \mathfrak{M}, w \rightrightarrows \mathfrak{M}', w'$; if there is a simulation Z such that $Z : \mathfrak{M}, w \rightrightarrows \mathfrak{M}', w'$, we sometimes write $\mathfrak{M}, w \rightrightarrows \mathfrak{M}', w'$.

A modal formula ϕ is *preserved under* simulations if for all models $\mathfrak{M}$ and $\mathfrak{M}'$, and all states w and w' in $\mathfrak{M}$ and $\mathfrak{M}'$, respectively, $\mathfrak{M}, w \Vdash \phi$ implies $\mathfrak{M}', w' \Vdash \phi$, whenever it is the case that $\mathfrak{M}, w \rightrightarrows \mathfrak{M}', w'$. $\dashv$

In various forms and under various names simulations have been considered in theoretical computer science. In the study of refinement, $\rightrightarrows$ is interpreted as follows: if $\mathfrak{M}, w \rightrightarrows \mathfrak{M}', w'$ then (the system modeled by) $\mathfrak{M}', w'$ refines or implements (the system modeled by) $\mathfrak{M}, w$. And in the database world one looks at simulations the other way around: if $\mathfrak{M}, w \rightrightarrows \mathfrak{M}', w'$, then $\mathfrak{M}', w'$ *constrains* the structure of $\mathfrak{M}, w$

by only allowing those relational patterns that are present in $\mathfrak{M}', w'$ itself. Note that if $\mathfrak{M}, w \rightrightarrows \mathfrak{M}', w'$ then $\mathfrak{M}', w'$ cannot enforce the presence of patterns. (See the Notes for references.) The following question naturally arises: which formulas are preserved when passing from $\mathfrak{M}, w$ to $\mathfrak{M}', w'$ along a simulation? Or, dually, which constraints on $\mathfrak{M}, w$ can be expressed by requiring that $\mathfrak{M}, w \rightrightarrows \mathfrak{M}', w'$?

Clearly simulations do not preserve the truth of all modal formulas. In particular, let $\mathfrak{M}$ be a one-point model with domain $\{w\}$ and empty relation; then, there is a simulation from $\mathfrak{M}, w$ to any state with the same valuation, no matter which model it lives in. Using this observation it is easy to show that universal modal formulas of the form $\Box(\cdots)$ or $\nabla(\cdots)$ are not preserved under simulations. On the other hand, by clause (ii) of Definition 2.77 existential modal formulas of the form $\Diamond(\cdots)$ or $\triangle(\cdots)$ are preserved under simulations. This leads to the conjecture that a modal formula is preserved under simulations if, and only if, it is equivalent to a formula that has been built from proposition letters, using only $\wedge$, $\vee$ and existential modal operators, that is, diamonds or triangles. Below we will prove this conjecture; our proof follows the proof of Theorem 2.68 to a large extent but there is an important difference. Since we are working *within* a modal language, and not in first-order logic, we can make do with a detour via (m-saturated) ultrafilter extensions rather than the (countably saturated) ultrapowers needed in the proof of Theorem 2.68.

Call a modal formula *positive existential* if it has been built up from proposition letters, using only $\wedge$, $\vee$ and existential modal operators $\Diamond$ and $\triangle$.

Theorem 2.78 *Let τ be a modal similarity type, and let ϕ be a τ-formula. Then ϕ is preserved under simulations iff it is equivalent to a positive existential formula.*

Proof. The easy inductive proof that positive existential formulas are preserved under simulations is left to the reader. For the converse, assume that ϕ is preserved under simulations, and consider the set of positive existential consequences of ϕ:

$$\mathrm{PEC}(\phi) = \{\psi \mid \ \psi \text{ is positive existential and } \phi \models \psi\}.$$

We will show that $\mathrm{PEC}(\phi) \models \phi$; then, by compactness, ϕ is equivalent to a positive existential modal formula. Assume that $\mathfrak{M}, w \Vdash \mathrm{PEC}(\phi)$; we need to show that $\mathfrak{M}, w \Vdash \phi$. Let $\Gamma = \{\neg\psi \mid \psi \text{ is positive existential and } \mathfrak{M}, w \nVdash \psi\}$.

Our first claim is that the set $\{\phi\} \cup \Gamma$ is consistent. For, suppose otherwise. Then there are formulas $\neg\psi_1, \ldots, \neg\psi_n \in \Gamma$ such that $\phi \models \psi_1 \vee \cdots \vee \psi_n$. By definition each formula ψ_i is a positive existential formula, hence, so is $\psi_1 \vee \cdots \vee \psi_n$. But then $\mathfrak{M}, w \Vdash \psi_1 \vee \cdots \vee \psi_n$, by assumption; from this it follows that $\mathfrak{M}, w \Vdash \psi_i$ for some i ($1 \leq i \leq n$). This contradicts $\neg\psi_i \in \Gamma$.

As a corollary we find a model $\mathfrak{N}$ and a state v of $\mathfrak{N}$ such that $\mathfrak{N}, v \Vdash \phi \wedge \bigwedge \Gamma$. Clearly, for every positive existential formula ψ, if $\mathfrak{N}, v \Vdash \psi$, then $\mathfrak{M}, w \Vdash \psi$. It follows from Proposition 2.59 that for the ultrafilter extensions $\mathfrak{ue}\,\mathfrak{M}$ and $\mathfrak{ue}\,\mathfrak{N}$

we have the same relation: for every positive existential formula ψ, if $\mathfrak{ue}\,\mathfrak{N}, \pi_v \Vdash \psi$, then $\mathfrak{ue}\,\mathfrak{M}, \pi_w \Vdash \psi$. By exploiting the fact that ultrafilter extensions are m-saturated (Proposition 2.61), it can be shown that this relation is in fact a simulation from $\mathfrak{ue}\,\mathfrak{N}, \pi_v$ to $\mathfrak{ue}\,\mathfrak{M}, \pi_w$; see Exercise 2.7.1.

In a diagram we have now the following situation.

$$\begin{array}{ccc}
\mathfrak{N}, v & & \mathfrak{M}, w \\
\leftrightsquigarrow \Big| & & \Big| \leftrightsquigarrow \\
\mathfrak{ue}\,\mathfrak{N}, \pi_v & \rightrightarrows & \mathfrak{ue}\,\mathfrak{M}, \pi_w .
\end{array}$$

We can carry ϕ around the diagram from $\mathfrak{N}, v$ to $\mathfrak{M}, w$ as follows. $\mathfrak{N}, v \Vdash \phi$ implies $\mathfrak{ue}\,\mathfrak{N}, \pi_v \Vdash \phi$ by Proposition 2.59. Since ϕ is preserved under simulations, we get $\mathfrak{ue}\,\mathfrak{M}, \pi_w \Vdash \phi$. By Proposition 2.59 again we conclude $\mathfrak{M}, w \Vdash \phi$. ⊣

Using Theorem 2.78 we can also answer the second of the two questions raised above. Call a constraint ϕ *expressible* if whenever $\mathfrak{M}, w$ satisfies ϕ and $\mathfrak{N}, v \rightrightarrows \mathfrak{M}, w$, then $\mathfrak{N}, v$ also satisfies ϕ. By Theorem 2.78 the expressible constraints (in first-order logic) are precisely the ones that are (equivalent to) the standard translations of negative universal modal formulas, that is, translations of modal formulas built up from negated proposition letters using only $\vee$, $\wedge$ and universal modal operators $\Box$ and $\triangledown$.

Safety

Recall from Exercise 2.2.6 that bisimulations preserve the truth of formulas from propositional dynamic logic. This result hinges on the fact that bisimulations not only preserve the relations R_a corresponding to atomic programs, but also relations that are definable from these using PDL's relational repertoire $\cup$, ; and *. Put differently, if the back-and-forth conditions in the definition of a bisimulation hold for the relations $R_{a_1}, \ldots, R_{a_n}, \ldots$, then they also hold for any relation that is definable from these using $\cup$, ; and *; these operations are 'safe' for bisimulation.

In this part of the section we work with modal similarity types having diamonds only.

Definition 2.79 Let τ be a modal similarity type, and let $\alpha(x, y)$ denote an $\mathcal{L}^1_\tau(\Phi)$-formula with at most two free variables. Then $\alpha(x, y)$ is called *safe for bisimulations* if the following holds.

> *If* $Z : \mathfrak{M} \leftrightarroweq \mathfrak{M}'$ is a bisimulation with wZw' and for some state v of $\mathfrak{M}$ we have $\mathfrak{M} \models \alpha(x, y)[wv]$,
> *then* there is a state v' of $\mathfrak{M}'$ such that $\mathfrak{M}' \models \alpha(x, y)[w'v']$ and vZv'.

In words, $\alpha(x, y)$ is safe if the back-and-forth clauses hold for $\alpha(x, y)$ whenever they hold for the atomic relations. $\dashv$

Example 2.80 (i) All PDL program constructors (;, $\cup$, and *) are safe for bisimulations. For instance, assume that wZw', where Z is a bisimulation, and $(w, v) \in (R\,;S)$ in $\mathfrak{M}$. Then, there exists u with Rwu and Suv in $\mathfrak{M}$; hence by the back-and-forth conditions for R and S, we find u' with uZu' and $R'w'u'$ in $\mathfrak{M}'$, and a state v' with vZv' and $S'u'v'$ in $\mathfrak{M}'$. Then v' is the required $(R\,;S)$-successor of w' in $\mathfrak{M}'$.

(ii) Atomic tests $(P)?$, defined by $(P)? := \{(x, y) \mid x = y \wedge Py\}$, are safe. For, assume that wZw', where Z is a bisimulation, and $(w, v) \in (P)?$. Then $w = v$ and $\mathfrak{M} \models Px[w]$. By the atomic clause in the definition of bisimulation, this implies $\mathfrak{M}' \models Px[w']$. Hence, $(w', w') \in (P)?$, as required.

(iii) Dynamic negation $\sim(R)$, defined by $\sim(R) = \{(x, y) \mid x = y \wedge \neg\exists z\, Rxz\}$, is safe. For, assume that wZw', where Z is a bisimulation, and $(w, v) \in \sim(R)$ in $\mathfrak{M}$. Then, $w = v$ and w has no R-successors in $\mathfrak{M}$. Now, suppose that w' did have an R'-successor in $\mathfrak{M}'$; then, by the back-and-forth conditions, w would have to have an R-successor in $\mathfrak{M}$ — a contradiction.

(iv) Intersection of relations is not safe; see Exercise 2.7.2. $\dashv$

Which operations are safe for bisimulations? Below, we give a complete answer for the restricted case where we consider first-order definable operations and languages with diamonds only. We need some preparations before we can prove this result.

First, we define a modal formula ϕ to be *completely additive in a proposition letter* p if it satisfies the following.

> For every family of non-empty sets $\{X_i\}_{i \in I}$ such that $V(p) = \bigcup_i X_i$ we have $(W, R_1, \ldots, V), w \Vdash \phi$ iff, for some i, $(W, R_1, \ldots, V_i), w \Vdash p$, where $V_i(p) = X_i$ and $V_i(q) = V(q)$ for $q \neq p$.

Completely additive formulas can be characterized syntactically. To this end, we need the following technical lemma. Let p be a fixed proposition letter. We write $\underline{\leftrightarrow}^-$ to denote the existence of a bisimulation for the modal language without the proposition letter p (exactly which proposition letter is meant will be clear in the applications of the lemma).

Lemma 2.81 *Assume that* $Z : \mathfrak{M}, w_0 \underline{\leftrightarrow}^- \mathfrak{N}, v_0$*, where* $\mathfrak{M}$ *and* $\mathfrak{N}$ *are intransitive tree-like transition systems with* $w_0 R \cdots R w_n$ *(in* $\mathfrak{M}$*),* $v_0 R \cdots R v_n$ *(in* $\mathfrak{N}$*) and* $w_i Z v_i$ *(*$1 \leq i \leq n$*). Then there are extensions* $(\mathfrak{M}^*, w_0)$ *of* $(\mathfrak{M}, w_0)$ *and* $(\mathfrak{N}^*, v_0)$ *of* $(\mathfrak{N}, v_0)$ *(i.e., the universe of* $\mathfrak{M}$ *is a subset of the universe of* $\mathfrak{M}^*$*, and likewise* *for $\mathfrak{N}$ and $\mathfrak{N}^*$) such that*

$$\begin{array}{ccc} (\mathfrak{M}, w_0) & Z : \underline{\leftrightarrow}^- & (\mathfrak{N}, v_0) \\ \underline{\leftrightarrow}\Big| & & \Big|\underline{\leftrightarrow} \\ (\mathfrak{M}^*, w_0) & Z' : \underline{\leftrightarrow}^- & (\mathfrak{N}^*, v_0), \end{array}$$

where Z' is such that for any i ($1 \leq i \leq n$) we have that w_i and v_i are only related to each other.

Proof. See Exercise 2.7.3. ⊣

Lemma 2.82 *A modal formula is completely additive in p iff it is equivalent to a disjunction of path formulas, that is, formulas of the form*

$$\psi_0 \wedge \langle a_1 \rangle (\psi_1 \wedge \cdots \wedge \langle a_n \rangle (\psi_n \wedge p) \cdots), \tag{2.4}$$

where p occurs in none of the formulas ψ_i.

Proof. We only prove the hard direction. Assume that ϕ is completely additive in p. Define

$$\mathrm{COC}(\phi) := \bigvee \{\psi \mid \psi \text{ is of the form (2.4) and } \psi \models \phi\},$$

that is, $\mathrm{COC}(\phi)$ is an infinite disjunction of modal formulas. We will show that $\phi \models \mathrm{COC}(\phi)$; then, by compactness, ϕ is equivalent to a finite disjunction of formulas of the form specified in (2.4), and this proves the lemma.

So, assume that $\mathfrak{M}, w_0 \Vdash \phi$; we need to show $\mathfrak{M}, w_0 \Vdash \mathrm{COC}(\phi)$. It suffices to find a formula ψ of the form specified in (2.4) such that $\mathfrak{M}, w_0 \Vdash \psi$ and $\psi \models \phi$. By Lemma 2.15 we may assume that $\mathfrak{M}$ is an intransitive, tree-like model with root w_0. As ϕ is completely additive in p, we may also assume that $V(p)$ is just a singleton w_n; see Figure 2.8. Consider the following description of the above path leading up to w_n:

$$\begin{aligned} \Psi(x_0, \ldots, x_n) \quad = \quad & \{ST_{x_i}(\psi) \mid \psi \in tp^-(w_i) \text{ and } 0 \leq i \leq n\} \\ & \cup \{R_i x_i x_{i+1} \mid 0 \leq i \leq n-1\} \cup \{P x_n\}, \end{aligned}$$

where we use $tp^-(w_i)$ to denote the set of p free modal formulas satisfied by w_i. The remainder of the proof is devoted to showing that $\Psi(x_0, \ldots, x_n) \models ST_{x_0}(\phi)$, and this will do to prove the lemma. For if $\Psi(x_0, \ldots, x_n) \models ST_{x_0}(\phi)$, then, for some finite subset $\Psi_0(x_0, \ldots, x_n) \subseteq \Psi(x_0, \ldots, x_n)$ we have $\Psi_0(x_0, \ldots, x_n) \models ST_{x_0}(\phi)$, by compactness. Since x_0 is the *only* free variable in $ST_{x_0}(\phi)$, this gives $\exists x_1 \ldots x_n \, \Psi_0(x_0, \ldots, x_n) \models ST_{x_0}(\phi)$. It is easy to see that the latter formula is (the standard translation of) a path formula ψ. Hence, we have found our formula satisfying $\mathfrak{M}, w_0 \Vdash \psi$ and $\psi \models \phi$.

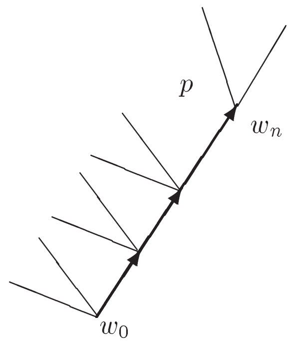


Fig. 2.8. True at only one state.

To show that $\Psi(x_0,\dots,x_n) \models ST_{x_0}(\phi)$ we proceed as follows. Take a model $\mathfrak{N}$ with $\mathfrak{N} \models \Psi(x_0,\dots,x_n)[v_0 v_1 \dots v_n]$; we need to show that $\mathfrak{N} \models ST_{x_0}(\phi)[v_0]$. It follows from the definition of Ψ that each w_i and v_i agree on all p free modal formulas.

We may assume that $\mathfrak{N}$ is an intransitive tree with root v. Take countably saturated elementary extensions $\mathfrak{M}^\dagger, w_0$ and $\mathfrak{N}^\dagger, v_0$ of $\mathfrak{M}, w_0$ and $\mathfrak{N}, v_0$, respectively. Since $\mathfrak{M}^\dagger$ and $\mathfrak{N}^\dagger$ are elementary extensions of $\mathfrak{M}$ and $\mathfrak{N}$, respectively, we may assume a number of things about $(\mathfrak{M}^\dagger, w_0)$ and $(\mathfrak{N}^\dagger, v_0)$ — things that can be expressed by first-order means, and hence are preserved under passing from a model to any of its elementary extensions. First, we may assume that w_0 and v_0 have no incoming R-transitions, for any R, since this can be expressed by means of the collection of all formulas of the form $\forall y\, \neg Ryx$, where R is a binary relation symbol in our language. Second, we may assume that states different from w_0 and v_0 have at most one incoming R-transition, for any R, since this can be expressed by the set of formulas of the form $\forall xyz\, (Ryx \wedge Rzx \to y = z)$. Summarizing, then, $\mathfrak{M}^\dagger, w_0$ and $\mathfrak{N}^\dagger, v_0$ are very much like intransitive trees with roots w_0 and v_0 — but possibly not quite: we have no guarantee that all nodes in $\mathfrak{M}^\dagger$ and $\mathfrak{N}^\dagger$ are actually accessible from w_0 and v_0, respectively, in finitely many steps.

Now, from the fact that w_i and v_i agree on all modal formulas and Theorem 2.65, we obtain a bisimulation $Z^\dagger$ such that $Z^\dagger : \mathfrak{M}^\dagger, w_i \leftrightarroweq^- \mathfrak{N}^\dagger, v_i$. Next, we want to apply Lemma 2.81, but to be able to do so, our models need to be rooted, intransitive trees. We can guarantee this by taking submodels $\mathfrak{M}^{\dagger\circ}$ and $\mathfrak{N}^{\dagger\circ}$ of $\mathfrak{M}^\dagger$ and $\mathfrak{N}^\dagger$ that are generated by w_0 and v_0, respectively. Clearly, for some Z, we have $Z : \mathfrak{M}^{\dagger\circ} \leftrightarroweq^- \mathfrak{N}^{\dagger\circ}$.

By Lemma 2.81 we can move to bisimilar extensions $\mathfrak{M}^{\dagger *}$ and $\mathfrak{N}^{\dagger *}$ of $\mathfrak{M}^{\dagger\circ}$ and $\mathfrak{N}^{\dagger\circ}$, respectively, and find a special bisimulation Z' linking w_i and v_i only to each other (for $1 \le i \le n$), as indicated in Figure 2.9.

We will amend the models $\mathfrak{M}^{\dagger *}$ and $\mathfrak{N}^{\dagger *}$ as follows. We shrink the interpretation

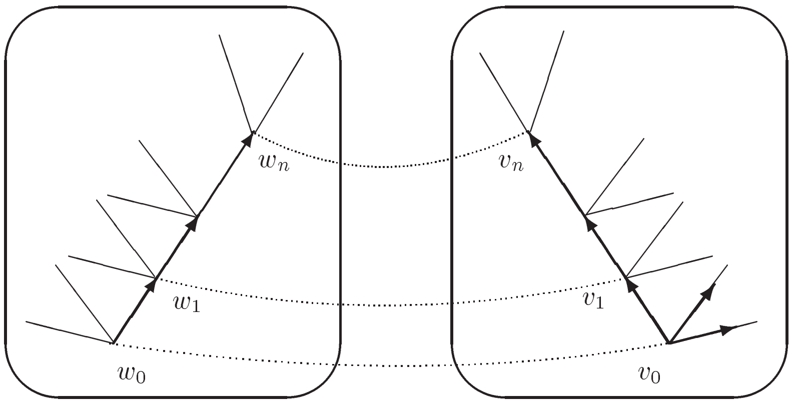


Fig. 2.9. Linking w_i only to v_i ($1 \leq i \leq n$).

of the proposition letter p so that it only holds at w_n and v_n. This allows us to extend Z' to a full directed simulation Z'' for the whole language:

$$
\begin{array}{ccccccc}
(\mathfrak{M}, w_0) & \preccurlyeq & (\mathfrak{M}^{\dagger}, w_0) & Z^{\dagger} : \underline{\leftrightarrow}^{-} & (\mathfrak{N}^{\dagger}, v_0) & \succcurlyeq & (\mathfrak{N}, v_0) \\
& & \underline{\leftrightarrow}\Big| & & \Big|\underline{\leftrightarrow} & & \\
& & (\mathfrak{M}^{\dagger\circ}, w_0) & Z : \underline{\leftrightarrow}^{-} & (\mathfrak{N}^{\dagger\circ}, v_0) & & \\
& & \underline{\leftrightarrow}\Big| & & \Big|\underline{\leftrightarrow} & & \\
& & (\mathfrak{M}^{\dagger *}, w_0) & Z' : \underline{\leftrightarrow}^{-} & (\mathfrak{N}^{\dagger *}, v_0) & & \\
& & \text{Shrink } V(p)\Big| & & \Big|\text{Expand } V(p) & & \\
& & (\mathfrak{M}^{\dagger **}, w_0) & Z'' : \underline{\leftrightarrow} & (\mathfrak{N}^{\dagger **}, v_0). & &
\end{array}
\tag{2.5}
$$

We can chase ϕ around the diagram displayed in (2.5), from $\mathfrak{M}, w_0$ to $\mathfrak{N}, v_0$; see Exercise 2.7.4. This proves the lemma. $\dashv$

Lemma 2.83 *For any program a and any formulas ϕ and ψ, the following identities hold in any model:*

(i) $(\neg\phi)? = {\sim}(\phi)?$
(ii) $(\phi \wedge \psi)? = (\phi)? \,;(\psi)?$
(iii) $(\langle a \rangle \phi)? = {\sim}{\sim}(a\,;(\phi)?)$.

The proof of this lemma is left as Exercise 2.7.5.

Theorem 2.84 *Let τ be a modal similarity type containing only diamonds, and let $\alpha(x, y)$ be a first-order formula in $\mathcal{L}^1_{\tau}(\Phi)$. Then $\alpha(x, y)$ is safe for bisimulations* *iff it can be defined from atomic formulas* R_axy *and atomic tests* $(P)?$ *using only* ;, $\cup$ *and* $\sim$.

Proof. To see that the constructions mentioned are indeed safe, consult Example 2.80. Now, to prove the converse, let $\alpha(x,y)$ be a safe first-order operation, and choose a *new* proposition letter p. Our first observation is that $\exists y\,(\alpha(x,y) \wedge Py)$ is preserved under bisimulations. So by Theorem 2.68, the formula $\exists y\,(\alpha(x,y) \wedge Py)$ is equivalent to a modal formula ϕ.

Next we exploit special properties of ϕ to arrive at our conclusion. First, because of its special form, $\exists y\,(\alpha(x,y) \wedge Py)$ is completely additive in P, and hence, ϕ is completely additive in p. Therefore, by Lemma 2.82 it is (equivalent to) a disjunction of the form specified in (2.4). Then, $\alpha(x,y)$ must be definable using the corresponding union of relations $(\psi_0)?\,;a_1\,;(\psi_1)?\,;\cdots;a_n\,;(\psi_n)?$. Finally, by using Lemma 2.83 all complex tests can be pushed inside until we get a formula of the required form, involving only ;, $\cup$, $\sim$ and ?. $\dashv$

Exercises for Section 2.7

2.7.1 Assume that $\mathfrak{M}$ and $\mathfrak{M}'$ are m-saturated models and suppose that for every positive existential formula ϕ it holds that $\mathfrak{M}, w \Vdash \phi$ only if $\mathfrak{M}', w' \Vdash \phi$ for some w and w'. Prove that $\mathfrak{M}, w \rightrightarrows \mathfrak{M}', w'$.

2.7.2 Prove that intersection of relations is not an operation that is safe for bisimulations (see Example 2.80).

2.7.3 The aim of this exercise is to prove Lemma 2.81: assume that $Z : \mathfrak{M}, w_0 \leftrightarroweq^- \mathfrak{N}, v_0$, where $\mathfrak{M}$ and $\mathfrak{N}$ are intransitive tree-like transition systems with $w_0R_j \cdots R_kw_n$ (in $\mathfrak{M}$), $v_0R_j \cdots R_kv_n$ (in $\mathfrak{N}$) and w_iZv_i ($1 \leq i \leq n$).

- (a) Explain why we may assume that all bisimulation links (between $\mathfrak{M}$ and $\mathfrak{N}$) occur between states at the same height in the tree.
- (b) Next, work your way up along the branch $w_0R_j \cdots R_kw_n$ and remove any double bisimulation links involving the w_i. from the w_i. More precisely, and starting at height 1, assume that w_1Zv_1 and w_1Zv. Add a copy of the submodel generated by w_1 to $\mathfrak{M}$, connect w_0 to the copy w_1' of w_1 by R_j, and 'divert' the bisimulation link w_1Zv to $w_1'Zv$. Show that the resulting model $\mathfrak{M}'$ is bisimilar (in the sense of $\leftrightarroweq$) to $\mathfrak{M}$ and that $\mathfrak{M}'$ is bisimilar to $\mathfrak{N}$ (in the sense of $\leftrightarroweq^-$).
- (c) Similar to the previous item, but now working up the branch $v_0R_j \cdots R_kv_n$ in $\mathfrak{N}$ to eliminate any double bisimulation links ending in one of the v_is ($1 \leq i \leq n$).
- (d) By putting together the previous items conclude that there are extensions $(\mathfrak{M}^*, w_0)$ of $(\mathfrak{M}, w_0)$ and $(\mathfrak{N}^*, v_0)$ of $(\mathfrak{N}, v_0)$ (i.e., the universe of $\mathfrak{M}$ is a subset of the universe of $\mathfrak{M}^*$, and likewise for $\mathfrak{N}$ and $\mathfrak{N}^*$) such that

$$\begin{array}{ccc} (\mathfrak{M}, w_0) & Z : \leftrightarroweq^- & (\mathfrak{N}, v_0) \\ \leftrightarroweq \Big| & & \Big| \leftrightarroweq \\ (\mathfrak{M}^*, w_0) & Z' : \leftrightarroweq^- & (\mathfrak{N}^*, v_0), \end{array}$$

where Z' is such that for any i ($1 \leq i \leq n$) we have that w_i and v_i are only related to each other.

2.7.4 Explain why we can chase ϕ around the diagram displayed in (2.5) to infer $\mathfrak{N}, v_0 \Vdash \phi$ from $\mathfrak{M}, w_0 \Vdash \phi$.

2.7.5 Prove Lemma 2.83.

2.8 Summary of Chapter 2

- *New Models from Old Ones*: Taking disjoint unions, generated submodels, and bounded morphic images are three important ways of building new models from old that leave the truth values of modal formulas invariant.
- *Bisimulations*: Bisimulations offer a unifying perspective on model invariance, and each of the constructions just mentioned is a kind of bisimulation. Bisimilarity implies modal equivalence, but the converse does not hold in general. On image-finite models, however, bisimilarity and modal equivalence coincide.
- *Using Bisimulations*: Bisimulations can be used to establish non-definability results (for example, to show that the global modality is not definable in the basic modal language), or to create models satisfying special relational properties (for example, to show that every satisfiable formula is satisfiable in a tree-like model).
- *Finite Model Property*: Modal languages have the finite model property (f.m.p.). One technique for establishing the f.m.p. is by a selection of states argument involving finite approximations to bisimulations. Another, the filtration method, works by collapsing as many states as possible.
- *Standard Translation*: The standard translation maps modal languages into classical languages (such as the language of first-order logic) in a way that reflects the satisfaction definition. Every modal formula is equivalent to a first-order formula in one free variable; if the similarity type is finite, finitely many variables suffice to translate all modal formulas. Propositional dynamic logic has to be mapped into a richer classical logic capable of expressing transitive closure.
- *Ultrafilter Extensions*: Ultrafilter extensions are built by using the ultrafilters over a given model as the states of a new model, and defining an appropriate relation between them. This leads to the first bisimilarity-somewhere-else result: two states in two models are modally equivalent if and only if their (counterparts in) the ultrafilter extensions of the two models are bisimilar.
- *Van Benthem Characterization Theorem*: The Detour Lemma — a bisimilarity-somewhere-else result in terms of ultrapowers — can be used to prove the Van Benthem Characterization Theorem: the modal fragment of first-order logic is the set of formulas in one free variable that are invariant for bisimulations.

- *Definability*: The Detour Lemma also leads to the following result: the modally definable classes of (pointed) models are those that are closed under bisimulations and ultraproducts, while their complements are closed under ultrapowers.
- *Simulation*: The modal formulas preserved under simulations are precisely the positive existential ones.
- *Safety*: An operation on relations is safe for bisimulations if whenever the back-and-forth conditions hold for the base relations, they also hold for the result of applying the operation to the relations. The first-order operations safe for bisimulations are the ones that can be defined from atoms and atomic tests, using only composition, union, and dynamic negation.

Notes

Kanger, Kripke, Hintikka, and others introduced models to modal logic in the late 1950s and early 1960s, and relational semantics (or Kripke semantics as it was usually called) swiftly became the standard way of thinking about modal logic. In spite of this, much of the material discussed in this chapter dates not from the 1960s, or even the 1970s, but from the late 1980s and 1990s. Why? Because relational semantics was not initially regarded as of independent interest, rather it was thought of as a tool that lead to interesting modal completeness theory and decidability results. Only in the early 1970s (with the discovery of the frame incompleteness results) did modal expressivity become an active topic of research — and even then, such investigations were initially confined to expressivity at the level of frames rather than at the level of models. Thus the most fundamental level of modal semantics was actually the last to be explored mathematically.

Generated submodels and bounded morphisms arose as tools for manipulating the canonical models used in modal completeness theory (we discuss canonical models in Chapter 4). *Point*-generated submodels, however, were already mentioned, under the name of connected model structures, in Kripke [291]. Bounded morphisms go back to at least Segerberg [396], where they are called *pseudo epimorphisms*; this soon got shortened down to *p-morphism*, which remains the most widely used terminology. A very similar, earlier, notion is in de Jongh and Troelstra [103]. The name *bounded morphism* stems from Goldblatt [192]. Disjoint unions and ultrafilter extensions seem to have first been isolated when modal logicians started investigating modal expressivity over frames in the 1970s (along with generated submodels and bounded morphisms they are the four constructions needed in the Goldblatt-Thomason theorem, which we discuss in the following chapter). Neither construction is as useful as generated submodels and bounded morphisms when it comes to proving completeness results, which is probably why they weren't noted earlier. However, both arise naturally in the context of modal duality theory, cf. Goldblatt [190, 191]. Ultrafilter extensions independently came about in the model-theoretic analysis of modal logic, see Fine [140]; the name seems to be due to van Benthem. The unraveling construction (that is, unwinding arbitrary models into trees; see Proposition 2.15) is helpful in many situations. Surprisingly, it was first used as early as in 1959, by Dummett and Lemmon [125], but the method only seems to have become widely known because of Sahlqvist's heavy use of it in his classic 1975 paper [388].

Vardi [434] has stressed the importance of the *tree model property* of modal logic: the property that a formula is satisfiable iff it is satisfiable at the root of a tree-like model. The tree model property paves the way for the use of automata-theoretic tools and tableaux-based proof methods. Moreover, it is essential for explaining the so-called robust decidability of modal logic — the phenomenon that the basic modal logic is decidable itself, and of reasonably low complexity, and that these features are preserved when the basic modal logic is extended by a variety of additional constructions, including counting, transitive closure, and least fixed points.

We discussed two ways of building finite models: the selection method and filtration. However, the use of finite *algebras* predates the use of finite models: they were first used in 1941 by McKinsey [328]; Lemmon [302] used and extended this method in 1966. The use of model-theoretic filtration dates back to Lemmon and Scott's long unpublished monograph *Intensional Logic* [303] (which began circulating in the mid 1960s); it was further developed in Segerberg's *An Essay in Classical Modal Logic* [396], which also seems to have given the method its name (see also Segerberg [394]). We introduced the selection method via the notion of finitely approximating a bisimulation, an idea which seems to have first appeared in 1985 in Hennessy and Milner [225].

The standard translation, in various forms, can be found in the work of a number of writers on modal and tense logic in the 1960s — but its importance only became fully apparent when the first frame incompleteness results were proved. Thomason [426], the paper in which frame incompleteness results was first established, uses the standard translation — and shows why the move to frames and validities requires a *second*-order perspective (something we will discuss in the following chapter). Thus the need became clear for a thorough investigation of the relation between modal and classical logic, and correspondence theory was born. But although other authors (notably Sahlqvist [388]) helped pioneer correspondence theory, it was the work of Van Benthem [35] which made clear the importance of systematic use of the standard translation to access results and techniques from classical modal theory. The observation that at most two variables are needed to translate basic modal formulas into first-order logic is due to Gabbay [158]. The earliest systematic study of finite variable fragments seems to be due to Henkin [223] in the setting of algebraic logic, and Immerman and Kozen [246] study the link with complexity and database theory. Consult Otto [355] for more on finite variable logics. Keisler [272] is still a valuable reference for infinitary logic. A variety of other translations from modal to classical logic have been studied, and for a wide variety of purposes. For example, simply standardly translating modal logics into first-order logic and then feeding the result to a theorem prover is not an efficient way of automating modal theorem proving. But the idea of automating modal reasoning via translation is interesting, and a variety of translations more suitable for this purpose have been devised; see Ohlbach *et al.* [351] for a survey.

Under the name of p-relations, bisimulations were introduced by Johan van Benthem in the course of his work on correspondence theory. Key references here are Van Benthem's 1976 PhD thesis [35]; his 1983 book based on the thesis [35]; and [42], his 1984 survey article on correspondence theory. In keeping with the spirit of the times, most of Van Benthem's early work on correspondence theory dealt with frame definability (in fact he devotes only 6 of the 227 pages in his book to expressivity over models). Nonetheless, much of this chapter has its roots in this early work, for in his thesis Van Benthem introduced the concept of a bisimulation (he used the name *p-relation* in [35, 41], and the name *zigzag relation* in [42]) and proved the Characterization Theorem. His original proof differs from the one given in the text: instead of appealing to saturated models, he employs an elementary chains argument. Explicitly isolating the Detour Lemma (which brings out the importance of ultrapowers) opens the way to Theorems 2.75 and 2.76 on definability and makes explicit the interesting analogies with first-order model theory discussed below. On the other hand, the original proof is more concrete. Both are worth knowing. The first published proof using saturated models seems to be due to Rodenburg [382], who used it to characterize the first-order fragment corresponding to intuitionistic logic.

The back-and-forth clauses of a bisimulation can be adapted to analyze the expressivity of a wide range of extended modal logics (such as those studied in Chapter 7), and such analyses are now commonplace. Bisimulation based characterizations have been given for the modal mu-calculus by Janin and Walukiewicz [249], for temporal logics with since and until by Kurtonina and De Rijke [295], for subboolean fragments of knowledge representation languages by Kurtonina and De Rijke [296], and for CTL^* by Moller and Rabinovich [339]. Related model-theoretic characterizations can be found in Immerman and Kozen [246] (for finite variable logics) and Toman and Niwiński [430] (for temporal query languages). Rosen [384] presents a version of the Characterization Theorem that also works for the case of finite models; the proof given in the text breaks down in the finite case as it relies on compactness and saturated models.

But bisimulations did not just arise in modal logic — they were independently invented in computer science as an equivalence relation on process graphs. Park [358] seems to have been the first author to have used bisimulations in this way. The classic paper on the subject is Hennessy and Milner [225], the key reference for the Hennessy-Milner Theorem. The reader should be warned, however, that just as the notion of bisimulation can be adapted to cover many different modal systems, the notion of bisimulation can be adapted to cover many different concepts of process — in fact, a survey of bisimulation in process algebra in the early 1990s lists over 155 variants of the notion [179]! Our definitions do not exclude bisimulations between a model and itself (*auto-bisimulations*); the quotient of a model with respect to its largest auto-bisimulation can be regarded as a minimal representation of this model. The standard method for computing the largest auto-bisimulation is the so-called Paige-Tarjan algorithm; see the contributions to Ponse, de Rijke and Venema [364] for relevant pointers and surveys.

More recently, bisimulations have become fundamental in a third area, non-well founded set theory. In such theories, the axiom of foundation is dropped, and sets are allowed to be members of themselves. Sets are thought of as graphs, and two sets are considered identical if and only if they are bisimilar. The classic source for this approach is Aczel [2], who explicitly draws on ideas from process theory. A recent text on the subject is Barwise and Moss [26], who link their work with the modal tradition. For recent work on modal logic and non-well founded set theory, see Baltag [19].

The name 'm-saturation' stems from Visser [443], but the notion is older: its first occurrence in the literature seems to be in Fine [140] (under the name 'modally saturated$_2$'). The concept of a Hennessy-Milner class is from Goldblatt [185] and Hollenberg [239]. Theorem 2.62, that equivalence of models implies bisimilarity between their ultrafilter extensions, is due to [239]. Chang and Keisler [89, Chapters 4 and 6] is the classic reference for the ultraproduct construction; their Chapters 2 and 5 also contain valuable material on saturated models. Doets and Van Benthem [120] give an intuitive explanation of the ultraproduct construction.

The results proved in this chapter are often analogs of standard results in first-order model theory, with bisimulations replacing partial isomorphisms. The Keisler-Shelah Theorem (see Chang and Keisler [89, Theorem 6.1.15]) states that two models are elementarily equivalent iff they have isomorphic ultrapowers; a weakened form, due to Doets and Van Benthem [120], replaces 'isomorphic' with 'partially isomorphic'. Theorem 2.74, which is due to De Rijke [109], is a modal analog of this weakened characterization theorem. Proposition 2.31 is similar to characterizations of logical equivalence for first-order logic due to Ehrenfeucht [127] and Fraïssé [149]; in fact, bisimulations can be regarded as the modal cousins of the model theoretic Ehrenfeucht-Fraïssé games. We will return to the theme of analogies between first-order and modal model theory in Section 7.6 when we prove a Lindström theorem for modal logic. See De Rijke [109] and Sturm [418] for further work on modal model theory; De Rijke and Sturm [113] provide global counterparts for the local definability results presented in Section 2.6. One can also characterize modal definability of model classes using ‘modal’ structural operations only, i.e., bisimulations, disjoint unions and ultrafilter extensions, cf. Venema [437].

Sources for the use of simulations in refinement are Henzinger et al. [227] and He Jifeng [252], and for their use in a database setting, consult Buneman et al. [74]; see De Rijke [106] for Theorem 2.78. The Safety Theorem 2.84 is due to Van Benthem [47]. The text follows the original proof fairly closely; an alternative proof has been given by Hollenberg [238], who also proves generalizations.

One final remark. Given the importance of *finite* model theory, the reader may be surprised to find so little in this chapter on the topic. But we don’t neglect finite model theory in this book: virtually all the results proved in Chapter 6 revolve around finite models and the way they are structured. That said, the topic of finite modal model theory has received less attention from modal logicians than it deserves. In spite of Rosen’s [384] proof of the Van Benthem characterization theorem for finite models, and in spite of work on modal 0-1 laws (Halpern and Kapron [211], Goranko and Kapron [197], and Grove *et al.* [206, 205]), finite modal model theory is clearly an area where interesting questions abound.