3

Frames

As we saw in Section 1.3, the concept of *validity*, which abstracts away from the effects of particular valuations, allows modal languages to get to grips with frame structure. As we will now see, this makes it possible for modal languages to *define* classes of frames, and most of the chapter is devoted to exploring this idea.

The following picture will emerge. Viewed as tools for defining frames, every modal formula corresponds to a second-order formula. Although this second-order formula sometimes has a first-order equivalent, even quite simple modal formulas can define classes of frames that no first-order formula can. In spite of this, there are extremely simple first-order definable frame classes which no modal formula can define. In short, viewed as frame description languages, modal languages exhibit an unusual blend of first- and second-order expressive powers.

The chapter has three main parts. The first, consisting of the first four sections, introduces frame definability, explains why it is intrinsically second-order, presents the four fundamental frame constructions and states the *Goldblatt-Thomason Theorem*, and discusses finite frames. The second part, consisting of the next three sections, is essentially a detailed exposition of the *Sahlqvist Correspondence Theorem*, which identifies a large class of modal formulas which correspond to first-order formulas. The final part, consisting of the last section, studies further frame constructions and gives a model-theoretic proof of the Goldblatt-Thomason theorem. With the exception of the last two sections, all the material in this chapter lies on the basic track.

Chapter guide

Section 3.1: Frame Definability (Basic track). This section introduces frame definability, and gives several examples of modally definable frame classes.

Section 3.2: Frame Definability and Second-Order Logic (Basic Track). We explain why frame definability is intrinsically second-order, and give examples of frame classes that are modally definable but not first-order definable.

Section 3.3: Definable and Undefinable Properties (Basic track). We first show that validity is preserved under the formation of *disjoint unions*, *generated subframes* and *bounded morphic images*, and anti-preserved under *ultrafilter extensions*. We then use these constructions to give examples of frame classes that are *not* modally definable, and state the Goldblatt-Thomason Theorem.

Section 3.4: Finite Frames (Basic track). Finite frames enjoy a number of pleasant properties. We first prove a simple analog of the Goldblatt-Thomason Theorem for finite transitive frames. We then introduce the *finite frame property*, and show that a normal modal logic has the finite frame property if and only if it has the finite model property.

Section 3.5: Automatic First-Order Correspondence (Basic track). Here we prepare for the proof of the *Sahlqvist Correspondence Theorem* in the following section. We introduce positive and negative formulas, and show that their monotonicity properties can help eliminate second-order quantifiers.

Section 3.6: Sahlqvist Formulas (Basic track). In this section we prove the Sahlqvist Correspondence Theorem. Our approach is incremental. We first explore the key ideas in the setting of two smaller fragments, and then state and prove the main result.

Section 3.7: More About Sahlqvist Formulas (Advanced track). We first discuss the limitations of the Sahlqvist Correspondence Theorem. We then prove Kracht's Theorem, which provides a syntactic description of the first-order formulas that can be obtained as translations of Sahlqvist formulas.

Section 3.8: Advanced Frame Theory (Advanced track). We finish off the chapter with some advanced material on frame constructions, and prove the Goldblatt-Thomason Theorem model-theoretically.

3.1 Frame Definability

This chapter is mostly about using modal formulas to define classes of frames. In this section we introduce the basic ideas (*definability*, and *first- and second-order frame languages*), and give a number of examples of modally definable frames classes. Most of these examples — and indeed, most of the examples given in this chapter — are important in their own right and will be used in later chapters.

Frame definability rests on the notion of a formula being *valid* on a frame, a concept which was discussed in Section 1.3 (see in particular Definition 1.28). We first recall and extend this definition.

Definition 3.1 (Validity) Let τ be a modal similarity type. A formula ϕ (of this similarity type) is *valid at a state w in a frame* $\mathfrak{F}$ (notation: $\mathfrak{F}, w \Vdash \phi$; here, of course, $\mathfrak{F}$ is a frame of type τ) if ϕ is true at w in every model $(\mathfrak{F}, V)$ based on $\mathfrak{F}$; ϕ is *valid on a frame* $\mathfrak{F}$ (notation: $\mathfrak{F} \Vdash \phi$) if it is valid at every state in $\mathfrak{F}$. A formula ϕ is *valid on a class of frames* K (notation: $\mathsf{K} \Vdash \phi$) if it is valid on every frame $\mathfrak{F}$ in K. We denote the class of frames where ϕ is valid by Fr_ϕ.

These concepts can be extended to sets of formulas in the obvious way. In particular, a set Γ of modal formulas (of type τ) is *valid on a frame* $\mathfrak{F}$ (also of type τ) if every formula in Γ is valid on $\mathfrak{F}$; and Γ is *valid on a class* K *of frames* if Γ is valid on every member of K. We denote the class of frames where Γ is valid by Fr_Γ. $\dashv$

Now for the concept underlying most of our work in this chapter:

Definition 3.2 (Definability) Let τ be a modal similarity type, ϕ a modal formula of this type, and K a class of τ-frames. We say that ϕ *defines* (or *characterizes*) K if for all frames $\mathfrak{F}$, $\mathfrak{F}$ is in K if and only if $\mathfrak{F} \Vdash \phi$. Similarly, if Γ is a set of modal formulas of this type, we say that Γ *defines* K if $\mathfrak{F}$ is in K if and only if $\mathfrak{F} \Vdash \Gamma$.

A class of frames is *(modally) definable* if there is some set of modal formulas that defines it. $\dashv$

In short, a modal formula defines a class of frames if the formula pins down precisely the frames that are in that class via the concept of validity. The following generalization of this concept is sometimes useful:

Definition 3.3 (Relative Definability) Let τ be a modal similarity type, ϕ a modal formula of this type, and C a class of τ-frames. We say that ϕ *defines* (or *characterizes*) a class K of frames *within* C (or *relative to* C) if for all frames $\mathfrak{F}$ in C we have that $\mathfrak{F}$ is in K if and only if $\mathfrak{F} \Vdash \phi$.

Similarly, if Γ is a set of modal formulas of this type, we say that Γ *defines* a class K of frames *within* C (or *relative to* C) if for all frames $\mathfrak{F}$ in C we have that $\mathfrak{F}$ is in K if and only if $\mathfrak{F} \Vdash \Gamma$. $\dashv$

Note that when C is the class of *all* τ-frames, definability within C is our original notion of definability. In Section 3.4 we will investigate which frames are definable within the class of finite transitive frames, but for the most part we will work with the 'absolute' notion of definability given in Definition 3.2.

We often say that a formula ϕ (or a set of formulas Γ) defines a *property* (for example, reflexivity) if it defines the class of frames satisfying that property. For example, we will shortly see that $p \to \Diamond p$ defines the class of reflexive frames; in practice, we would often simply say that $p \to \Diamond p$ defines reflexivity.

Up till now our discussion has been purely modal — but of course, as frames are just relational structures, we are free to define frame classes using a wide variety of *non-modal* languages. For example, the class of reflexive frames is simply the class of all frames that make $\forall x\, Rxx$ true. In this chapter, we are interested in comparing modal languages with the following classical languages as tools for defining frame classes:

Definition 3.4 (Frame Languages) For any modal similarity type τ, the *first-order frame language* of τ is the first-order language that has the identity symbol $=$ together with an $n+1$-ary relation symbol $R_{\triangle}$ for each n-ary modal operator $\triangle$ in τ. We denote this language by $\mathcal{L}^1_\tau$. We often call it the *first-order correspondence language* (for τ).

Let Φ be any set of proposition letters. The *monadic second-order frame language* of τ over Φ is the monadic second-order language obtained by augmenting $\mathcal{L}^1_\tau$ with a Φ-indexed collection of monadic predicate variables. (That is, this language has all the resources of $\mathcal{L}^1_\tau$, and in addition is capable of quantifying over subsets of frames.) We denote this language by $\mathcal{L}^2_\tau(\Phi)$, though sometimes we suppress reference to Φ and write $\mathcal{L}^2_\tau$. Moreover, we often simply call it the *second-order frame language* or the *second-order correspondence language* (for τ), taking it for granted that only monadic second-order quantification is permitted. $\dashv$

Note that the second-order frame language is extremely powerful, even for the basic modal similarity type. For example, if R is interpreted as the relation of set membership, second-order ZF set theory can be axiomatized by a single sentence of this language.

Definition 3.5 (Frame Correspondence) If a class of frames (or more informally, a property) can be defined by a modal formula ϕ and by a formula α from one of these frame languages, then we say that ϕ and α are each others (frame) *correspondents*. $\dashv$

For example, the basic modal formula $p \to \Diamond p$ and the first-order sentence $\forall x Rxx$ are correspondents, for we will shortly see that $p \to \Diamond p$ defines reflexivity. Later in this chapter we will show how to systematically find correspondents of modal formulas by adopting a slightly different perspective on the standard translation introduced in Section 2.4.

In Definition 3.5 we did mention the possibility that modal formulas correspond to a *set* of first-order formulas. Why not? The reason is that this situation simply cannot occur, as we ask the reader to show in Exercise 3.8.3.

There are a number of practical reasons for being interested in frame definability. First, some applications of modal logic are essentially *syntactically* driven; their starting point is some collection of modal formulas expressing axioms, laws, or principles which for some reason we find interesting or significant. Frame definability can be an invaluable tool in such work, for by determining which frame classes these formulas define we obtain a mathematical perspective on their content. On the other hand, some applications of modal logic are essentially *semantically* driven; their starting point is some class of frames of interest. But here too definability is a useful concept. For a start, can the modal language distinguish the 'good' frames from the 'bad' ones? And which properties can the modal language express *within* the class of 'good' frames? Finally, many applied modal languages contain several modalities, whose intended meanings are interrelated. Sometimes it is clear that these relationships should validate certain formulas, and we want to extract the frame-theoretic property they correspond to. On the other hand it may be clear what the relevant frame-theoretic property is (for example, in the basic temporal language we want the P and F operators to scan backwards and forward along the *same* relation) and we want to see whether there is a modal formula that defines this property. In short, thinking in terms of frame definability can be useful for a variety of reasons — and as the following examples will make clear, modal languages can define some very interesting frame classes indeed.

Example 3.6 In Example 1.10 in Section 1.2 we mentioned the following reading of the modalities: read $\Diamond\phi$ as 'it is *possibly* the case that ϕ' and $\Box\phi$ as '*necessarily* ϕ'. We also mentioned that a number of interesting looking principles concerning necessity and possibility could be stated in the basic modal language. Here are three important examples, together with their traditional names:

(T) $p \to \Diamond p$
(4) $\Diamond\Diamond p \to \Diamond p$
(5) $\Diamond p \to \Box\Diamond p$

But now the problems start. While the status of T seems secure (if p holds here-and-now, p must be *possible*) but what about 4 and 5? When we have to deal with embedded modalities, our intuitions tend to fade, even for such simple formulas as 4 and 5; it is not easy to say whether they should be accepted, and if we only have our everyday understanding of the words 'necessarily' and 'possibly' to guide us, it is difficult to determine whether these principles are interrelated. What we need is a *mathematical* perspective on their content, and that is what the frame definability offers. So let's see what frame conditions these principles define.

Our first claim is that for any frame $\mathfrak{F} = (W, R)$, the axiom T corresponds to *reflexivity* of the relation R:

$$\mathfrak{F} \Vdash \text{T iff } \mathfrak{F} \models \forall x\, Rxx. \tag{3.1}$$

The proof of the right to left direction of (3.1) is easy: let $\mathfrak{F}$ be a reflexive frame, and take an arbitrary valuation V on $\mathfrak{F}$, and a state w in $\mathfrak{F}$ such that $(\mathfrak{F}, V), w \Vdash p$. We need to show that $\Diamond p$ holds at some state that is accessible from w — but as R is reflexive, w is accessible from itself, and $w \Vdash \Diamond p$.

For the other direction, we use contraposition: suppose that R is *not* reflexive, that is, there exists a state w which is not accessible from itself. To falsify T in $\mathfrak{F}$, it suffices to find a valuation V and a state v such that p holds at v, but $\Diamond p$ does not. It is pretty obvious that we should choose v to be our irreflexive state w. Now the valuation V has to satisfy two conditions: (1) $w \in V(p)$ and (2) $\{x \in W \mid Rwx\} \cap V(p) = \varnothing$. Consider the *minimal* valuation V satisfying condition (1), that is, take

$$V(p) = \{w\}.$$

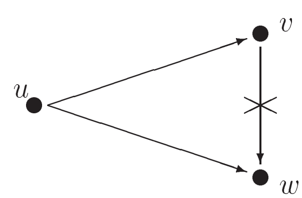
Then it is immediate that $(\mathfrak{F}, V), w \Vdash p$. Now let v be an R-successor of w. As Rww does not hold in $\mathfrak{F}$, v must be distinct from w, so $v \nVdash p$. As v was arbitrary, $w \nVdash \Diamond p$. This proves (3.1).

Likewise, one can prove that for any frame $\mathfrak{F} = (W, R)$

$$\mathfrak{F} \Vdash 4 \text{ iff } R \text{ is transitive, and} \tag{3.2}$$

$$\mathfrak{F} \Vdash 5 \text{ iff } R \text{ is euclidean}, \tag{3.3}$$

where a relation is *euclidean* if it satisfies $\forall xyz\,((Rxy \wedge Rxz) \rightarrow Ryz)$. We leave the proofs of (3.2) and the easy (right to left) direction of (3.3) to the reader. For the left to right direction of (3.3), we again argue by contraposition. Assume that $\mathfrak{F}$ is a non-euclidean frame; then there must be states u, v and w such that Ruv, Ruw, but not Rvw:



We will try to falsify 5 in u; for this purpose we have to find a valuation V such that $(\mathfrak{F}, V), u \Vdash \Diamond p$ and $(\mathfrak{F}, V), u \nVdash \Box\Diamond p$. In other words, we have to make p ***true*** at some R-successor x of u, and *false* at all R-successors of some R-successor y of u. Some reflection shows that appropriate candidates for x and y are w and v, respectively. Note that again the constraints on V are twofold: (1) $w \in V(p)$ and (2) $\{z \mid Rvz\} \cap V(p) = \varnothing$.

Let us take a *maximal* V satisfying condition (2), that is, define

$$V(p) = \{z \in W \mid \text{ it is not the case that } Rvz\}.$$

Now clearly $v \nVdash \Diamond p$, so $u \nVdash \Box\Diamond p$. On the other hand we have $w \Vdash p$, since w is in the set $\{z \in W \mid \text{ it is not the case that } Rvz\}$. So $u \Vdash \Diamond p$. In other words, we have indeed found a valuation V and a state u such that 5 does not hold in u. Therefore, 5 is not valid in $\mathfrak{F}$. This proves (3.3). $\dashv$

Example 3.7 Suppose that we are working with the basic temporal language (see Section 1.3 and in particular Example 1.25) and that we are interested in *dense* bidirectional frames (that is, structures in which between every two points there is a third). This property can be defined using a first-order sentence (namely $\forall xy\,(x < y \to \exists z\,(x < z \wedge z < y))$ but can the basic temporal language define it too?

It can. The following simple formula suffices: $Fp \to FFp$. To see this, let $\mathfrak{T} = (T, <)$ be a frame such that $\mathfrak{T} \Vdash Fp \to FFp$. Suppose that a point $t \in T$ has a $<$-successor t'. To show that t and t' satisfy the density condition, consider the following *minimal* valuation V_m guaranteeing that $(\mathfrak{T}, V_m), t \Vdash Fp$:

$$V_m(p) = \{t'\}.$$

Now, under this valuation $t \Vdash Fp$, and by assumption $\mathfrak{T} \Vdash Fp \to FFp$, hence $t \Vdash FFp$. This means there is a point s such that $t < s$ and $s \Vdash Fp$. But as t' is the *only* state where p holds, this implies that $s < t'$, so s is the intermediate point we were looking for.

Conversely, let $\mathfrak{T} = (T, <)$ be a dense frame, and assume that under some valuation V, Fp holds at some $t \in T$. Then there is a point t' such that $t < t'$ and $t' \Vdash p$. But as $\mathfrak{T}$ is dense, there is a point s such that $t < s < t'$, hence $s \Vdash Fp$ and hence $t \Vdash FFp$.

Note that nothing in the previous argument depended on the fact that we were working with the basic temporal language; the previous argument also shows that density is definable in the basic modal language using the formula $\Diamond p \to \Diamond\Diamond p$. Note that this is the converse of the 4 axiom that defines transitivity. $\dashv$

Example 3.8 Here's a more abstract example. Suppose we are working with a similarity type with three binary operators $\triangle_1$, $\triangle_2$ and $\triangle_3$, and that we are interested in the class of frames in which the three ternary accessibility relations (denoted by R_1, R_2 and R_3, respectively), offer, so to speak, three 'perspectives' on the same relation. To put this precisely, suppose we want the condition

$$R_1 stu \text{ iff } R_2 tus \text{ iff } R_3 ust$$

to hold for all s, t and u in such frame. Can we define this class of frames?

We can. We will show that for all frames $\mathfrak{F} = (W, R_1, R_2, R_3)$ we have

$$\mathfrak{F} \Vdash p \wedge (q \triangle_1 r) \to (q \wedge r \triangle_2 p) \triangle_1 r \text{ iff } \mathfrak{F} \models \forall xyz\,(R_1 xyz \to R_2 yzx). \tag{3.4}$$

(Recall that we use infix notation for dyadic operation symbols.) The easy direction is from right to left. Let $\mathfrak{F}$ be a frame satisfying $\forall xyz\,(R_1 xyz \to R_2 yzx)$. Consider an arbitrary valuation V on $\mathfrak{F}$ and an arbitrary state s such that $(\mathfrak{F}, V), s \Vdash p \wedge (q \triangle_1 r)$. Then, $s \Vdash p$ and there are states t and u with $R_1 stu$, $t \Vdash q$ and $u \Vdash r$. From $R_1 stu$ we derive $R_2 tus$. But then $t \Vdash q \wedge r \triangle_2 p$, so by $R_1 stu$ we have $s \Vdash (q \wedge r \triangle_2 p) \triangle_1 r$.

For the other direction, suppose that the modal formula $p \wedge (q \triangle_1 r) \to (q \wedge r \triangle_2 p) \triangle_1 r$ is valid in $\mathfrak{F}$, and consider states s, t and u in $\mathfrak{F}$ with $R_1 stu$. We will show that $R_2 tus$. Consider a valuation V with $V(p) = \{s\}$, $V(q) = \{t\}$ and $V(r) = \{u\}$. Then $(\mathfrak{F}, V), s \Vdash p \wedge q \triangle_1 r$, so by our assumption, $s \Vdash (q \wedge r \triangle_2 p) \triangle_1 r$. Hence, there must be states t', u' with $R_1 st'u'$, $t' \Vdash q \wedge r \triangle_2 p$ and $u' \Vdash r$. From $t' \Vdash q$ it follows that $t = t'$, so we have $t \Vdash r \triangle_2 p$. Again, using the truth definition we find states s'', u'' with $R_2 tu''s''$, $u'' \Vdash r$ and $s'' \Vdash p$. The latter two facts imply that $u'' = u$ and $s'' = s$. But then we have $R_2 tus$, as required. $\dashv$

From these examples the reader could easily get the impression that modal formulas always correspond to frame properties that are definable in first-order logic. This impression is wrong, and in the next section we will see why.

Exercises for Section 3.1

3.1.1 Consider a language with two diamonds $\langle 1 \rangle$ and $\langle 2 \rangle$. Show that $p \to [2]\langle 1 \rangle p$ is valid on precisely those frames for the language that satisfy the condition $\forall xy\, (R_2 xy \to R_1 yx)$. What sort of frames does $p \to [1]\langle 1 \rangle p$ define?

3.1.2 Consider a language with three diamonds $\langle 1 \rangle$, $\langle 2 \rangle$, and $\langle 3 \rangle$. Show that the modal formula $\langle 3 \rangle p \leftrightarrow \langle 1 \rangle \langle 2 \rangle p$ is valid on a frame for this language if and only if the frame satisfies the condition $\forall xy\, (R_3 xy \leftrightarrow \exists z\, (R_1 xz \wedge R_2 zy))$.

3.2 Frame Definability and Second-Order Logic

In this section we show that modal languages can get to grips with notions that exceed the expressive power of first-order logic, and explain why. We start by presenting three well-known examples of modal formulas that define frame properties which cannot be expressed in first-order logic. Then, drawing on our discussion of the standard translation in Section 2.4, we show that such results are to be expected: as we will see, modal formulas standardly correspond to *second-order* frame conditions. Indeed, the real mystery is not why they do so (this turns out to be rather obvious), but why they sometimes correspond to simple *first-order* conditions such as reflexivity or transitivity (we discuss this more difficult issue in Sections 3.5–3.7).

Example 3.9 Consider the Löb formula $\Box(\Box p \to p) \to \Box p$, which we will call L for brevity. This formula plays an essential role in *provability* logic, a branch of modal logic where $\Box\phi$ is read as ‘it is *provable* (in some formal system) that ϕ’. The formula L is named after Löb, who proved L as a theorem of the provability logic of Peano Arithmetic. We’ll first show that L defines the class of frames (W, R) such that R is transitive and R’s converse is well-founded. (A relation R is *well-founded* if there is no infinite sequence $\ldots R w_2 R w_1 R w_0$; hence, R’s converse is well-founded if there is no infinite R-path emanating from any state. In particular, this excludes cycles and loops.)

We'll then show that this is a class of frames that first-order frame languages *cannot* define; that is, we'll show that this class is not *elementary*.

To see that L defines the stated property, assume that $\mathfrak{F} = (W, R)$ is a frame with a transitive and conversely well-founded relation R, and then suppose for the sake of a contradiction that L is not valid in $\mathfrak{F}$. This means that there is a valuation V and a state w such that $(\mathfrak{F}, V), w \nVdash \Box(\Box p \to p) \to \Box p$. In other words, $w \Vdash \Box(\Box p \to p)$, but $w \nVdash \Box p$. Then w must have a successor w_1 such that $w_1 \nVdash p$, and as $\Box p \to p$ holds at all successors of w, we have that $w_1 \nVdash \Box p$. This in turn implies that w_1 must have a successor w_2 where p is false; note that by the transitivity of R, w_2 is also a successor of w. But now, simply by repeating our argument, we see that w_2 must have a p-falsifying successor w_3 (which by transitivity must be a successor of w_1), that w_3 has a successor w_4 (which by transitivity must be a successor of w_1), and so on. In short, we have found an infinite path $wRw_1Rw_2Rw_3R\ldots$, contradicting the converse well-foundedness of R. (Note that the points w_1, w_2, $\ldots$ need not all be distinct.)

$$\underset{w}{\bullet} \longrightarrow \underset{w_1}{\bullet} \longrightarrow \underset{w_2}{\bullet} \longrightarrow \underset{w_3}{\bullet} \longrightarrow \quad \cdots$$

For the other direction, we use contraposition. That is, we assume that either R is not transitive or its converse is not well-founded; in both cases we have to find a valuation V and a state w such that $(\mathfrak{F}, V), w \nVdash L$. We leave the case where R is not transitive to the reader (hint: instead of L, consider the frame equivalent formula $\Diamond p \to \Diamond(p \wedge \neg\Diamond p)$) and only consider the second case. So assume that R is transitive, but not conversely well-founded. In other words, suppose we have a transitive frame containing an infinite sequence $w_0Rw_1Rw_2R\ldots$. We exploit the presence of this sequence by defining the following valuation V:

$$V(p) = W \setminus \{x \in W \mid \text{there is an infinite path starting from } x\}.$$

We leave it to the reader to verify that under this valuation, $\Box p \to p$ is true *everywhere* in the model, whence certainly, $(\mathfrak{F}, V), w_0 \Vdash \Box(\Box p \to p)$. The claim then follows from the fact that $(\mathfrak{F}, V), w_0 \nVdash \Box p$.

Finally, to show that the class of frames defined by L is not elementary, an easy compactness argument suffices. Suppose for the sake of a contradiction that there is a first-order formula equivalent to L; call this formula λ. As λ is equivalent to L, any model making λ true must be transitive. Let $\sigma_n(x_0, \ldots, x_n)$ be the first-order formula stating that there is an R-path of length n through $x_0, \ldots, x_n$:

$$\sigma_n(x_0, \ldots, x_n) = \bigwedge_{0 \leq i < n} Rx_ix_{i+1}.$$

Obviously, every *finite* subset of

$$\Sigma = \{\lambda\} \cup \{\forall xyz\,((Rxy \land Ryz) \to Rxz)\} \cup \{\sigma_n \mid n \in \omega\}$$

is satisfiable in a finite linear order, and hence in the class of transitive, conversely well-founded frames. Thus by the Compactness Theorem, Σ itself must have a model. But it is clear that Σ is *not* satisfiable in any conversely well-founded frame — and λ, being equivalent to L, is supposed to define the class of transitive, conversely well-founded frames. From this contradiction we conclude that L cannot be equivalent to any first-order formula.

Could L then perhaps be equivalent to an (infinite) *set* of first-order formulas? No — we already mentioned (right after Definition 3.5 that this kind of correspondence never occurs. $\dashv$

Our next example concerns *propositional dynamic logic* (PDL). Recall that this language contains a family of diamonds $\{\langle\pi\rangle \mid \pi \in \Pi\}$ (where Π is a collection of programs) and the program constructors $\cup$, ; and *. In the intended frames for this language (that is, the *regular* frames; see Example 1.26) we want the accessibility relations for diamonds built using these constructors to reflect choice, composition, and iteration of programs, respectively. Now, to reflect iteration we demanded that the relation R_{π^*} used for the program π^* be the reflexive, transitive closure of the relation R_π used for π. But it is well-known that this constraint *cannot* be expressed in first-order logic (as with the Löb example, this can be shown using a compactness argument, and the reader was asked to do this in Exercise 2.4.5). Because of this, when we discussed PDL at the level of models in Section 2.4 we used the *infinitary* language $\mathcal{L}_{\omega_1\omega}$ as the correspondence language for PDL; using infinite disjunctions enabled us to capture the 'keep looking!' force of $*$ that eludes first-order logic. But although first-order logic cannot get to grips with $*$, PDL itself can — via the concept of frame definability.

Example 3.10 PDL can be interpreted on any transition system of the form $\mathfrak{F} = (W, R_\pi)_{\pi\in\Pi}$. Let us call such a frame $*$-*proper* if the transition relation R_{π^*} of each program π^* is the reflexive and transitive closure of the transition relation R_π of π. Can we single out, by modal means, the $*$-proper frames within the class of all transition systems of the form $(W, R_\pi)_{\pi\in\Pi}$? And can we then go on to single out the class of all regular frames?

The answer to both questions is *yes*. Consider the following set of formulas

$$\Delta = \{[\pi^*](p \to [\pi]p) \to (p \to [\pi^*]p),\ \langle\pi^*\rangle p \leftrightarrow (p \lor \langle\pi\rangle\langle\pi^*\rangle p) \ \mid\ \pi \in \Pi\}.$$

As we mentioned in Example 1.15, $[\pi^*](p \to [\pi]p) \to (p \to [\pi^*]p)$ is called *Segerberg's axiom*, or the *induction axiom*. We claim that for any PDL-frame $\mathfrak{F}$:

$$\mathfrak{F} \Vdash \Delta \text{ iff } \mathfrak{F} \text{ is } *\text{-proper}. \tag{3.5}$$

The reader is asked to supply a proof of this in Exercise 3.2.1.

A straightforward consequence is that PDL is strong enough to define the class of regular frames. The constraints on the relations interpreting $\cup$ and ; are simple first-order conditions, and

$$\Gamma = \{\langle\pi_1;\pi_2\rangle p \leftrightarrow \langle\pi_1\rangle\langle\pi_2\rangle p, \langle\pi_1 \cup \pi_2\rangle p \leftrightarrow \langle\pi_1\rangle p \vee \langle\pi_2\rangle p \mid \pi \in \Pi\}.$$

pins down down what is required. So $\Delta \cup \Gamma$ defines the regular frames. $\dashv$

In the previous two examples we encountered modal formulas that expressed frame properties that were, although not elementary, still relatively easy to understand. (Note however that in order to formally express (converse) well-foundedness in a classical language, one needs heavy machinery — the infinitary language $\mathcal{L}_{\omega_1\omega}$ does not suffice!) The next example shows that extremely simple modal formulas can define second-order frame conditions that are not easy to understand at all.

Example 3.11 We will show that the McKinsey formula $\Box\Diamond p \to \Diamond\Box p$ does not correspond to a first-order condition by showing that it violates the Löwenheim-Skolem theorem.

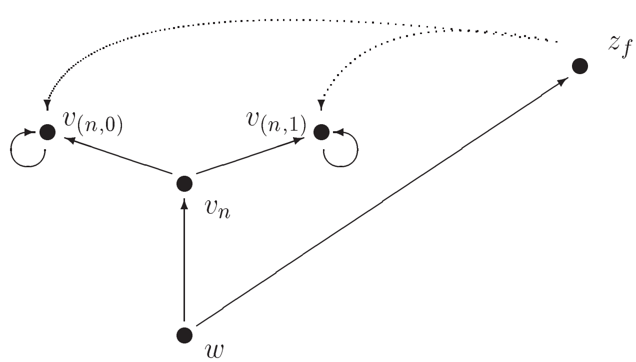
Consider the frame $\mathfrak{F} = (W, R)$, where

$$W = \{w\} \cup \{v_n, v_{(n,i)} \mid n \in \mathbb{N}, i \in \{0,1\}\} \cup \{z_f \mid f : \mathbb{N} \to \{0,1\}\},$$

and

$$\begin{aligned} R \;=\; & \{(w, v_n), (v_n, v_{(n,i)}), (v_{(n,i)}, v_{(n,i)}) \mid n \in \mathbb{N}, i \in \{0,1\}\} \cup \\ & \{(w, z_f), (z_f, v_{(n,f(n))}) \mid n \in \mathbb{N}, f : \mathbb{N} \to \{0,1\}\}. \end{aligned}$$

In a picture:



Note that W contains uncountably many points, for the set of functions indexing the z points is uncountable.

Our first observation is that $\mathfrak{F} \Vdash \Box\Diamond p \to \Diamond\Box p$. We leave it to the reader to verify that for all u different from w, $\mathfrak{F}, u \Vdash \Box\Diamond p \to \Diamond\Box p$. As to showing that

$\mathfrak{F}, w \Vdash \Box\Diamond p \to \Diamond\Box p$, suppose that $(\mathfrak{F}, V), w \Vdash \Box\Diamond p$. Then, for each $n \in \mathbb{N}$, $(\mathfrak{F}, V), v_n \Vdash \Diamond p$. From this we get either $(\mathfrak{F}, V), v_{(n,0)} \Vdash p$ or $(\mathfrak{F}, V), v_{(n,1)} \Vdash p$. Choose $f : \mathbb{N} \to \{0, 1\}$ such that $(\mathfrak{F}, V), v_{(n,f(n))} \Vdash p$, for each $n \in \mathbb{N}$. Then clearly, $(\mathfrak{F}, V), z_f \Vdash \Box p$, and so $(\mathfrak{F}, V), w \Vdash \Diamond\Box p$.

In order to show that $\Box\Diamond p \to \Diamond\Box p$ does not define a first-order frame condition, let's view the frame $\mathfrak{F}$ as a first-order model with domain W. By the downward Löwenheim-Skolem Theorem (here we need the strong version of Theorem A.11) there must be a countable elementary submodel $\mathfrak{F}'$ of $\mathfrak{F}$ whose domain W' contains w, and each v_n, $v_{(n,0)}$ and $v_{(n,1)}$. As W is uncountable and W' countable, there must be a mapping $f : \mathbb{N} \to \{0, 1\}$ such that z_f does not belong to W'. Now, if the McKinsey formula was equivalent to a first-order formula it would be valid on $\mathfrak{F}'$ (the Löwenheim-Skolem Theorem tells us that $\mathfrak{F}$ and $\mathfrak{F}'$ are elementarily equivalent). But we will show that the McKinsey formula is *not* valid on $\mathfrak{F}'$, hence it cannot be equivalent to a first-order formula.

Let V' be a valuation on $\mathfrak{F}'$ such that $V'(p) = \{v_{(n,f(n))} \mid n \in \mathbb{N}\}$; here f is a mapping such that z_f does not belong to W'. We will show that under V', $\Box\Diamond p$ is true at w, but $\Diamond\Box p$ is not.

It is easy to see that $(\mathfrak{F}', V'), w \nVdash \Diamond\Box p$. For a start, since p holds at exactly one of $v_{(n,0)}$ and $v_{(n,1)}$, $\Box p$ is false at each v_n. Now consider an arbitrary element z_g in W'. Then g is distinct from f, so there must be an element $n \in \mathbb{N}$ such that $g(n) \neq f(n)$. Observe that p is thus true at $v_{(n,f(n))}$, and, more interestingly, false at $v_{(n,g(n))}$; this means that z_g has a successor where p is false, so $(\mathfrak{F}', V'), z_g \nVdash \Box p$. Hence, we have not been able to find a successor for w where $\Box p$ holds, so $(\mathfrak{F}', V'), w \nVdash \Diamond\Box p$.

In order to show that $(\mathfrak{F}', V'), w \Vdash \Box\Diamond p$ we reason as follows. Note first that $(\mathfrak{F}', V'), v_n \Vdash \Diamond p$, for each state v_n. Now consider an arbitrary element z_g of W'. Call two states z_h and z_k of $\mathfrak{F}$ *complementary* if for all n, $h(n) = 1 - k(n)$; the reader should verify that this relation can be expressed in first-order logic. Now suppose that z_g is complementary to z_f; since complementary states are unique, the fact that $\mathfrak{F}'$ is an elementary submodel of $\mathfrak{F}$ would imply that z_f exists in $\mathfrak{F}'$ as well. Clearly then, we may conclude that z_g is *not* complementary to z_f. Hence, there exists some $n \in \mathbb{N}$ such that $g(n) = f(n)$. Therefore, $(\mathfrak{F}', V'), z_g \Vdash \Diamond p$. But then $\Diamond p$ holds at every successor of w. $\dashv$

Clearly then, modal languages can express many highly complex properties via the notion of frame validity. In fact, as was shown by S.K. Thomason for the basic modal similarity type, the consequence relation for the entire second-order language $\mathcal{L}^2_\tau$ can be reduced in a certain sense to the (global) consequence relation over frames. More precisely, Thomason showed that there is a computable translation f taking $\mathcal{L}^2$ sentences α to modal formulas $f(\alpha)$, and a special fixed modal formula δ, such that for all sets of of $\mathcal{L}^2$ sentences Σ, we have that

$$\Sigma \models \alpha \ \text{ iff } \ \{\delta\} \cup \{f(\sigma) \mid \sigma \in \Sigma\} \Vdash^g f(\alpha).$$

On the frame level, propositional modal logic must be understood as a rather strong fragment of classical monadic second-order logic. We now face the question: *why*?

The answer turns out to be surprisingly simple. Recall from Definition 3.1 that validity is defined by quantifying over all states of the universe and all possible valuations. But a valuation assigns a *subset* of a frame to each proposition letter, and this means that when we quantify across all valuations we are implicitly quantifying across all subsets of the frame. In short, monadic second-order quantification is hard-wired into the very definition of validity; it is hardly surprising that frame-definability is such a powerful concept.

Let's make this answer more precise. In the previous chapter, we saw that at the level of models, the modal language $ML(\tau, \Phi)$ can be translated in a truth-preserving way into the first-order language $\mathcal{L}^1_\tau(\Phi)$ (see Proposition 2.47). Let us adopt a slightly different perspective:

> *View the predicate symbol P that corresponds to the propositional letter p as a monadic second-order variable that we can quantify over.*

If we do this, we are in effect viewing the standard translation as a way of translating into the second-order frame language $\mathcal{L}^2_\tau(\Phi)$ introduced in Definition 3.4. And if we view the standard translation this way we are lead, virtually immediately, to the following result.

Proposition 3.12 *Let τ be a modal similarity type, and ϕ a τ-formula. Then for any τ-frame $\mathfrak{F}$ and any state w in $\mathfrak{F}$:*

$$\begin{aligned} \mathfrak{F}, w \Vdash \phi \quad &\textit{iff} \quad \mathfrak{F} \models \forall P_1 \ldots \forall P_n \, ST_x(\phi)[w], \\ \mathfrak{F} \Vdash \phi \quad &\textit{iff} \quad \mathfrak{F} \models \forall P_1 \ldots \forall P_n \forall x \, ST_x(\phi). \end{aligned}$$

Here, the second-order quantifiers bind second-order variables P_i corresponding to the proposition letters p_i occurring in ϕ.

Proof. Let $\mathfrak{M} = (\mathfrak{F}, V)$ be any model based on $\mathfrak{F}$, and let w be any state in $\mathfrak{F}$. Then we have that

$$(\mathfrak{F}, V), w \Vdash \phi \ \text{ iff } \ \mathfrak{F} \models ST_x(\phi)[w, P_1, \ldots, P_n],$$

where the notation $[w, P_1, \ldots, P_n]$ means 'assign w to the free first-order variable x in $ST_x(\phi)$, and $V(p_1), \ldots, V(p_n)$ to the free monadic second-order variables'. Note that this equivalence is nothing new; it's simply a restatement of Proposition 2.47 in second-order terms. But then we obtain the first part of the Theorem simply by universally quantifying over the free variables $P_1,\ldots,P_n$. The second part follows from the first by universally quantifying over the states of the frame (as in Proposition 3.30). $\dashv$

It is fairly common to refer to the $\mathcal{L}^2_\tau(\Phi)$ formula $\forall P_1 \ldots \forall P_n \forall x\, ST_x(\phi)$ as the standard translation of ϕ, since it is usually clear whether we are working at the level of models or the level of frames. Nonetheless, we will try and reserve the term standard translation to mean the $\mathcal{L}^1_\tau(\Phi)$ formula produced by the translation process, and refer to $\forall P_1 \ldots \forall P_n \forall x\, ST_x(\phi)$ as the *second-order translation* of ϕ.

Let's sum up what we have learned. That modal formulas can define second-order properties of frames is neither mysterious nor surprising: because modal validity is defined in terms of quantification over subsets of frames, it is intrinsically second-order, hence so is the notion of frame definability. Indeed, the real mystery lies not with such honest, hard-working, formulas as Löb and McKinsey, but with such lazy formulas as T, 4 and 5 discussed in the previous section. For example, if we apply the second-order translation to T (that is, $p \to \Diamond p$) we obtain

$$\forall P \forall x\, (Px \to \exists y(Rxy \wedge Py)).$$

We already know that T defines reflexivity, so this must be a (somewhat baroque) second-order way of expressing reflexivity — and it's fairly easy to see that this is so. But this sort of thing happens a lot: 4 and 5 give rise to (fairly complex) second-order expressions, yet the complexity melts away leaving a simple first-order equivalent behind. The contrast with the McKinsey formula is striking: what *is* going on? This is an interesting question, and we discuss it in detail in Sections 3.5–3.7.

Another point is worth making: our discussion throws light on the somewhat mysterious *general frames* introduced in Section 1.4. Recall that a general frame is a frame together with a collection of valuations A satisfying certain modally natural closure conditions. We claimed that general frames combined the key advantage of frames (namely, that they support the key logical notion of validity) with the advantage of models (namely, that they are concrete and easy to work with). The work of this section helps explain why.

The key point is this. A general frame can be viewed as a *generalized model* for (monadic) second-order logic. A generalized model for second-order logic is a model in which the second-order quantifiers are viewed as ranging not over *all* subsets, but only over a pre-selected sub-collection of subsets. And of course, the collection of valuations A in a general frame is essentially such a sub-collection of subsets. This means that the following equivalence holds:

$$(\mathfrak{F}, A) \Vdash \phi \text{ iff } (\mathfrak{F}, A) \models \forall P_1 \ldots \forall P_n \forall x\, ST_x(\phi)$$

Here the block of quantifiers $\forall P_1 \ldots \forall P_n$ denotes not genuine second-order quantification, but generalized second-order quantification (that is, quantification over the subsets in A). Generalized second-order quantification is essentially a first-order 'approximation' of second-order quantification that possesses many properties that genuine second-order quantification lacks (such as Completeness, Compactness, and Löwenheim-Skolem). In short, one of the reasons general frames are so useful is that they offer a first-order perspective (via generalized models) on what is essentially a second-order phenomenon (frame validity). This isn't the full story — the algebraic perspective on general frames is vital to modal logic — but it should make clear that these unusual looking structures fill an important logical niche.

Exercises for Section 3.2

3.2.1 (a) Consider a modal language with two diamonds $\langle 1\rangle$ and $\langle 2\rangle$. Prove that the class of frames in which R_1 is the reflexive transitive closure of R_2 is defined by the conjunction of the formulas $\langle 1\rangle p \to (p \vee \langle 1\rangle(\neg p \wedge \langle 2\rangle p)$ and $\langle 1\rangle p \leftrightarrow (p \vee \langle 2\rangle\langle 1\rangle p)$.

(b) Conclude that in the similarity type of PDL, the set Δ as defined in Example 3.10 defines the class of $*$-proper frames.

(c) Consider the example of multi-agent epistemic logic; let $\{1, \ldots, n\}$ be the set of agents. Suppose that one is interested in the operators E ($E\phi$ stands for 'everybody knows ϕ') and C ($C\phi$ meaning that 'it is common knowledge that ϕ'). The intended relations modeling E and C are given by:

$$\begin{array}{lll} R_E uv & \text{iff} & \bigwedge_{1\leq i\leq n} R_i uv \\ R_C uv & \text{iff} & \text{there is a path } u = x_0 R_E x_1 R_E \ldots x_{n-1} R_E x_n = v. \end{array}$$

Write down a set of (epistemic) formulas that characterizes the class of epistemic frames where these conditions are met.

3.2.2 Show that Grzegorczyk's formula, $\Box((p \to \Box p) \to p) \to p$ characterizes the class of frames $\mathfrak{F} = (W, R)$ satisfying (i) R is reflexive, (ii) R is transitive and (iii) there are no infinite paths $x_0 R x_1 R x_2 R \ldots$ such that for all i, $x_i \neq x_{i+1}$.

3.2.3 Consider the basic temporal language (see Example 1.24). Recall that a frame $\mathfrak{F} = (W, R_F, R_P)$ for this language is called *bidirectional* if R_P is the converse of R_F.

(a) Prove that among the finite bidirectional frames, the formula $G(Gp \to p) \to Gp$ together with its converse, $H(Hp \to p) \to Hp$ defines the transitive and irreflexive frames.

(b) Prove that among the bidirectional frames that are transitive, irreflexive, and satisfy $\forall xy\,(R_F xy \vee x = y \vee R_P xy)$, this same set defines the finite frames.

(c) Is there a finite set of formulas in the *basic* modal language that has these same definability properties?

3.2.4 Consider the following formula in the basic similarity type:

$$\psi := \Diamond\Box p \to \Diamond(\Box(p \wedge q) \vee \Box(p \wedge \neg q)).$$

The aim of this exercise is to show that ψ does not define a first-order condition on frames.

(a) To obtain some intuitions about the meaning of ψ, let us first give a relatively simple first-order condition *implying* the validity of ψ:

$$\alpha := \forall xy\,(Rxy \to \exists z\,(Rxz \wedge \forall uv\,((Rzu \wedge Rzv) \to (u = v \wedge Ryu)))),$$

stating (in words) that for every pair (x, y) in R, x has a successor z which itself has at most one successor, this point being also a successor of y.
Show that ψ is valid in any frame satisfying α.

(b) Consider the frame $\mathfrak{F} = (W, R)$ which we define as follows. Let u be a non-principal ultrafilter over the set $\mathbb{N}$ of the natural numbers. Then $W := \{u\} \cup u \cup \mathbb{N}$, that is, the states of W are u itself, each subset of $\mathbb{N}$ that is a member of u and each natural number. The relation R is the converse of the membership relation, that is, Rst iff $t \in s$. Show that $\mathfrak{F} \not\Vdash \alpha$ and $\mathfrak{F} \Vdash \psi$.

(c) Prove that ψ does not have a first-order correspondent by showing that ψ is invalid on all *countable* structures that are elementarily equivalent to $\mathfrak{F}$ (that is, all countable structures satisfying the same first-order formulas as $\mathfrak{F}$).

3.3 Definable and Undefinable Properties

We have seen that modal languages are a powerful tool for defining frames: we have seen examples of modally definable frame classes that are not first-order definable, and it is clear that validity is an inherently second-order concept. But what are the limits of modal definability? For example, can modal languages define all first-order frame classes (the answer is *no*, as we will shortly see)? And anyway, how should we go about showing that a class of frames is *not* modally definable? After all, we can't try out all possible formulas; something more sophisticated is needed.

In this section we will answer these question by introducing four fundamental frame constructions: *disjoint unions*, *generated subframes*, *bounded morphic images*, and *ultrafilter extensions*. The names should be familiar: these are the frame theoretic analogs of the model-theoretic constructions studied in the previous chapter, and they are going to do a lot of work for us, both here and in later chapters. For a start, it is a more-or-less immediate consequence of the previous chapter's work that the first three constructions preserve modal validity, while the fourth anti-preserves it. But this means that these constructions provide powerful tests for modal definability: by showing that some class of frames is *not* closed under one of these constructions, we will be able to show that it *cannot* be modally definable.

Definition 3.13 The definitions of the disjoint union of a family of frames, a generated subframe of a frame, and a bounded morphism from one frame to another, are obtained by deleting the clauses concerning valuations from Examples 2.2, 2.5 and 2.10.

That is, for disjoint τ-frames $\mathfrak{F}_i = (W_i, R_{\triangle i})_{\triangle \in \tau}$ ($i \in I$), their *disjoint union* is the structure $\biguplus_i \mathfrak{F}_i = (W, R_\triangle)_{\triangle \in \tau}$ such that W is the union of the sets W_i and for each $\triangle \in \tau$, $R_\triangle$ is the union $\bigcup_{i \in I} R_{\triangle i}$.

We say that a τ-frame $\mathfrak{F}' = (W', R'_\triangle)_{\triangle\in\tau}$ is a *generated subframe* of the frame $\mathfrak{F} = (W, R_\triangle)_{\triangle\in\tau}$ (notation: $\mathfrak{F}' \rightarrowtail \mathfrak{F}$) whenever $\mathfrak{F}'$ is a subframe of $\mathfrak{F}$ (with respect to $R_\triangle$ for all $\triangle \in \tau$), and the following heredity condition is fulfilled for all $\triangle \in \tau$

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

Let X be a subset of the universe of a frame $\mathfrak{F}$; we denote by $\mathfrak{F}_X$ *the subframe generated by* X, that is, the generated subframe of $\mathfrak{F}$ that is based on the smallest set W' that contains X and satisfies the above heredity condition. If X is a singleton $\{w\}$, we write $\mathfrak{F}_w$ for the *subframe generated by* w; if a frame $\mathfrak{F}$ is generated by a singleton subset of its universe, we call it *rooted* or *point-generated*.

And finally, a bounded morphism from a τ-frame $\mathfrak{F} = (W, R_\triangle)_{\triangle\in\tau}$ to a τ-frame $\mathfrak{F}' = (W', R'_\triangle)_{\triangle\in\tau}$ is a function from W to W' satisfying the following two conditions:

(forth) For all $\triangle \in \tau$, $R_\triangle w v_1 \dots v_n$ implies $R'_\triangle f(w) f(v_1) \dots f(v_n)$.
(back) 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 $1 \leq i \leq n$).

We say that $\mathfrak{F}'$ is a bounded morphic image of $\mathfrak{F}$, notation: $\mathfrak{F} \twoheadrightarrow \mathfrak{F}'$, if there is a surjective bounded morphism from $\mathfrak{F}$ onto $\mathfrak{F}'$. $\dashv$

It is an essential characteristic of modal formulas that their validity is preserved under the structural operations just defined:

Theorem 3.14 *Let τ be a modal similarity type, and ϕ a τ-formula.*

(i) *Let $\{\mathfrak{F}_i \mid i \in I\}$ be a family of frames. Then $\biguplus \mathfrak{F}_i \Vdash \phi$ if $\mathfrak{F}_i \Vdash \phi$ for every i in I.*
(ii) *Assume that $\mathfrak{F}' \rightarrowtail \mathfrak{F}$. Then $\mathfrak{F}' \Vdash \phi$ if $\mathfrak{F} \Vdash \phi$.*
(iii) *Assume that $\mathfrak{F} \twoheadrightarrow \mathfrak{F}'$. Then $\mathfrak{F}' \Vdash \phi$ if $\mathfrak{F} \Vdash \phi$.*

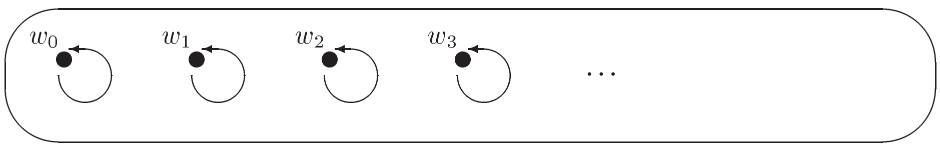
Proof. We only prove (iii), the preservation result for taking bounded morphic images, and leave the other cases to the reader as Exercise 3.3.1. So, assume that f is a surjective bounded morphism from $\mathfrak{F}$ onto $\mathfrak{F}'$, and that $\mathfrak{F} \Vdash \phi$. We have to show that $\mathfrak{F}' \Vdash \phi$. So suppose that ϕ is *not* valid in $\mathfrak{F}'$. Then there must be a valuation V' and a state w' such that $(\mathfrak{F}', V'), w' \nVdash \phi$. Define the following valuation V on $\mathfrak{F}$:

$$V(p_i) = \{x \in W \mid f(x) \in V'(p_i)\}.$$

This definition is tailored to make f a bounded morphism between the models $(\mathfrak{F}, V)$ and $(\mathfrak{F}', V')$ — the reader is asked to verify the details. Now we use the fact that f is surjective to find a w such that $f(w) = w'$. It follows from Proposition 2.14 that $(\mathfrak{F}, V), w \nVdash \phi$. In other words, we have falsified ϕ in the frame $\mathfrak{F}$, and shown the contrapositive of the desired result. $\dashv$

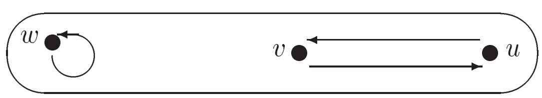
Think of these frame constructions as *test criteria* for the definability of frame properties: if a property is not preserved under one (or more) of these frame constructions, then it cannot be modally definable. Let's consider some examples of such testing.

Example 3.15 The class of finite frames is not modally definable. For suppose there was a set of formulas Δ (in the basic modal similarity type) characterizing the finite frames. Then Δ would be valid in every one-point frame $\mathfrak{F}_i = (\{w_i\}, \{(w_i, w_i)\})$ ($i < \omega$). By Theorem 3.14(1) this would imply that Δ was also valid in the disjoint union $\biguplus_i \mathfrak{F}_i$:



But clearly this cannot be the case, for $\biguplus_i \mathfrak{F}_i$ is infinite.

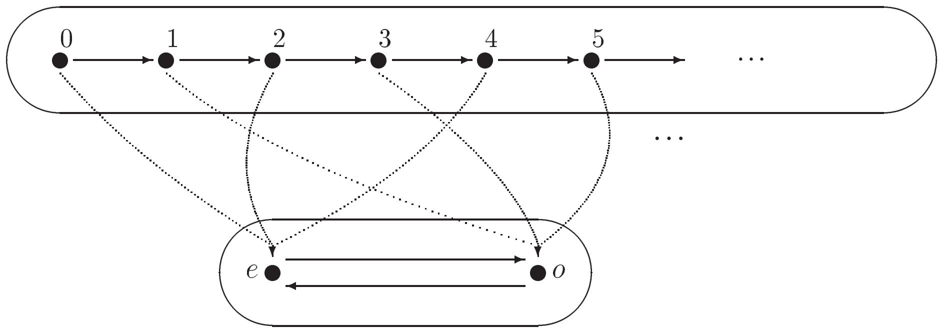
The class of frames having a reflexive point ($\exists x\, Rxx$) does not have a modal characterization either (again we work with the basic modal similarity type). For suppose that the set Δ characterized this class. Consider the following frame $\mathfrak{F}$:



As w is a reflexive state, $\mathfrak{F} \Vdash \Delta$. Now consider the generated subframe $\mathfrak{F}_v$ of $\mathfrak{F}$. Clearly, Δ cannot be valid in $\mathfrak{F}_v$, since neither v nor u is reflexive. But this contradicts the fact that validity of modal formulas is preserved under taking generated subframes (Theorem 3.14(ii)).

The two final examples involve the use of bounded morphisms. First, irreflexivity is not definable. To see this, simply note that the function which collapses the set of natural numbers in their usual order to a single reflexive point is a surjective bounded morphism. As the former frame is irreflexive, while the latter is not, irreflexivity cannot be modally definable.

Actually, a more sophisticated variant of this example lets us prove even more. Consider the following two frames: $\mathfrak{F} = (\omega, S)$, the natural numbers with the successor relation (Smn iff $n = m + 1$), and $\mathfrak{G} = (\{e, o\}, \{(e, o), (o, e)\})$ as depicted below.



In Example 2.11 we saw that the map f sending even numbers to e and odd numbers to o is a surjective bounded morphism. By the same style of reasoning as in the earlier examples, it follows that no property P is modally definable if $\mathfrak{F}$ has P and $\mathfrak{G}$ lacks it. This shows, for example, that there is no set of formulas characterizing the asymmetric frames ($\forall xy\,(Rxy \to \neg Ryx)$). $\dashv$

Now for the fourth frame construction. Recall that in Section 2.5 we introduced the idea of *ultrafilter extensions*; see Definition 2.57 and Proposition 2.59. Once again, simply by ignoring the parts of the definition that deal with valuations, we can lift this concept to the level of frames, and this immediately provides us with the following *anti*-preservation result:

Corollary 3.16 *Let τ be a modal similarity type, $\mathfrak{F}$ a τ-frame, and ϕ a τ-formula. Then $\mathfrak{F} \Vdash \phi$ if $\mathfrak{ue}\,\mathfrak{F} \Vdash \phi$.*

Proof. Assume that ϕ is not valid in $\mathfrak{F}$. That is, there is a valuation V and a state w such that $(\mathfrak{F}, V), w \Vdash \neg\phi$. By Proposition 2.59, $\neg\phi$ is false at u_w in the ultrafilter extension of $\mathfrak{M}$. But then we have refuted ϕ in $\mathfrak{ue}\,\mathfrak{F}$. $\dashv$

Once again, we can use this result to show that frame properties are not modally definable. For example, working in the basic modal similarity type, consider the property that every state has a reflexive successor: $\forall x \exists y\,(Rxy \wedge Ryy)$. We claim that this property is *not* modally definable, even though it is preserved under taking disjoint unions, generated subframes and bounded morphic images. To verify our claim, the reader is asked to consider the frame in Example 2.58. It is easy to see that every state of $\mathfrak{ue}\,\mathfrak{F}$ has a reflexive successor — take any non-principal ultrafilter. But $\mathfrak{F}$ itself clearly does not satisfy the property, as $\mathfrak{F}$ has *no* reflexive states. Now suppose that the property were modally definable, say by the set of formulas Δ. Then we would have $\mathfrak{ue}\,\mathfrak{F} \Vdash \Delta$, but $\mathfrak{F} \not\Vdash \Delta$ — a clear violation of Corollary 3.16.

Note the direction of the preservation result in Corollary 3.16. It states that modal validity is *anti*-preserved under taking ultrafilter extensions. This naturally raises the question whether the other direction holds as well, that is, whether $\mathfrak{F} \Vdash \phi$

implies $\mathfrak{ue}\,\mathfrak{F} \Vdash \phi$. For a partial answer to this question, we need the following theorem:

Theorem 3.17 *Let τ be a modal similarity type, and $\mathfrak{F}$ a τ-frame. Then $\mathfrak{F}$ has an ultrapower $\prod_U \mathfrak{F}$ such that $\prod_U \mathfrak{F} \twoheadrightarrow \mathfrak{ue}\,\mathfrak{F}$. In a diagram:*

$$\begin{tikzcd}
\prod_U \mathfrak{F} \arrow[d, dash] \arrow[dr] & \\
\mathfrak{F} & \mathfrak{ue}\,\mathfrak{F}
\end{tikzcd}$$

Proof. Advanced track readers will be asked to supply a proof of this Theorem in Exercise 3.8.1 below. ⊣

And now we have the following partial converse to Corollary 3.16:

Corollary 3.18 *Let τ be a modal similarity type, and ϕ a τ-formula. If ϕ defines a first-order property of frames, then frame validity of ϕ is preserved under taking ultrafilter extensions.*

Proof. Let ϕ be a modal formula which defines a first-order property of frames, and let $\mathfrak{F}$ be a frame such that $\mathfrak{F} \Vdash \phi$. By the previous theorem, there is an ultrapower $\prod_U \mathfrak{F}$ of $\mathfrak{F}$ such that $\prod_U \mathfrak{F} \twoheadrightarrow \mathfrak{ue}\,\mathfrak{F}$. As first-order properties are preserved under taking ultrapowers, $\prod_U \mathfrak{F} \Vdash \phi$. But then $\mathfrak{ue}\,\mathfrak{F} \Vdash \phi$ by Theorem 3.14. ⊣

We are on the verge of one of the best-known results in modal logic: the *Goldblatt-Thomason Theorem*. This result tells us that — at least as far as first-order definable frame classes are concerned — the four frame constructions we have discussed constitute necessary *and sufficient* conditions for a class of frames to be modally definable. We are not going to prove this important result right away, but we will take this opportunity to state it precisely. We use the following terminology: a class of frames K *reflects* ultrafilter extensions if $\mathfrak{ue}\,\mathfrak{F} \in \mathsf{K}$ implies $\mathfrak{F} \in \mathsf{K}$.

Theorem 3.19 (Goldblatt-Thomason Theorem) *Let τ be a modal similarity type. A first-order definable class K of τ-frames is modally definable if and only if it is closed under taking bounded morphic images, generated subframes, disjoint unions and reflects ultrafilter extensions.*

Proof. A model-theoretic proof will be given in Section 3.8 below; this proof lies on the advanced track. An algebraic proof will be given in Chapter 5; this proof lies on the basic track. In addition, a simple special case which holds for finite transitive frames is proved in the following section. ⊣

In fact, we can weaken the condition of first-order definability to closure under ultrapowers, cf. Exercise 3.8.4 or Theorem 5.54.

Exercises for Section 3.3

3.3.1 (a) Prove that frame validity is preserved under taking generated subframes and disjoint unions.
(b) Which of the implications in Theorem 3.14 can be replaced with an equivalence?
(c) Is frame validity preserved under taking ultraproducts?

3.3.2 Consider the basic modal language. Show that the following properties of frames are not modally definable:

(a) antisymmetry ($\forall xy(Rxy \land Ryx \to x \neq y)$),
(b) $|W| > 23$,
(c) $|W| < 23$,
(d) acyclicity (there is no path from any x to itself),
(e) every state has at most one predecessor,
(f) every state has at least two successors.

3.3.3 Consider a language with three diamonds, $\diamond_1$, $\diamond_2$ and $\diamond_3$. For each of the frame conditions on the corresponding accessibility relations below, find out whether it is modally definable or not.

(a) R_1 is the union of R_2 and R_3,
(b) R_1 is the intersection of R_2 and R_3,
(c) R_1 is the complement of R_2,
(d) R_1 is the composition of R_2 and R_3,
(e) R_1 is the identity relation,
(f) R_1 is the complement of the identity relation.

3.3.4 Show that any frame is a bounded morphic image of the disjoint union of its rooted generated subframes.

3.4 Finite Frames

In this section we prove two simple results about finite frames. First we state and prove a version of the Goldblatt-Thomason Theorem for finite transitive frames. Next we introduce the finite frame property, and show that a normal modal logic has the finite frame property if and only if it has the finite model property.

Finite transitive frames

An elegant analog of the Goldblatt-Thomason Theorem holds for finite transitive frames: within this class, closure under the three structural operations of (finite) disjoint unions, generated submodels, and bounded morphisms is a necessary *and sufficient* condition for a class of frames to be modally definable. The proof is straightforward and makes use of *Jankov-Fine formulas*.

Let $\mathfrak{F} = (W, R)$ be a point-generated finite transitive frame for the basic modal similarity type, and let w be a root of $\mathfrak{F}$. The Jankov-Fine formula $\phi_{\mathfrak{F},w}$ is essentially a description of $\mathfrak{F}$ that has the following property: it is satisfiable on a frame $\mathfrak{G}$ if and only if $\mathfrak{F}$ is a bounded morphic image of a generated subframe of $\mathfrak{G}$.

We build Jankov-Fine formulas as follows. Enumerate the states of $\mathfrak{F}$ as $w_0, \ldots,$ w_n, where $w = w_0$. Associate each state w_i with a distinct proposition letter p_i. Let $\phi_{\mathfrak{F},w}$ be the conjunction of the following formulas:

(i) p_0
(ii) $\Box(p_0 \vee \cdots \vee p_n)$.
(iii) $(p_i \to \neg p_j) \wedge \Box(p_i \to \neg p_j)$, for each i, j with $i \neq j \leq n$
(iv) $(p_i \to \Diamond p_j) \wedge \Box(p_i \to \Diamond p_j)$, for each i, j with Rw_iw_j
(v) $(p_i \to \neg\Diamond p_j) \wedge \Box(p_i \to \neg\Diamond p_j)$, for each i, j with $\neg Rw_iw_j$

Note that as R is transitive, each node in $\mathfrak{F}$ is accessible in one step from w. It follows that when formulas of the form $\psi \wedge \Box\psi$ are satisfied at w, ψ is true throughout $\mathfrak{F}$. With this observed, the content of Jankov-Fine formulas should be clear: the first three conjuncts state that each node in $\mathfrak{F}$ is uniquely labeled by some p_i (with p_0 labelling w_0) while the last two conjuncts use this labeling to describe the frame structure.

Lemma 3.20 *Let $\mathfrak{F}$ be a transitive, finite, point-generated frame, let w be a root of $\mathfrak{F}$, and let $\phi_{\mathfrak{F},w}$ be the Jankov-Fine formula for $\mathfrak{F}$ and w. Then for any frame $\mathfrak{G}$ we have the following equivalence: there is a valuation V and a node v such that $(\mathfrak{G}, V), v \Vdash \phi_{\mathfrak{F},w}$ if and only if there exists a bounded morphism from $\mathfrak{G}_v$ onto $\mathfrak{F}$.*

Proof. Left to the reader as Exercise 3.4.1. $\dashv$

With the help of this lemma, it is easy to prove the following Goldblatt-Thomason analog:

Theorem 3.21 *Recall that τ_0 denotes the basic modal similarity type. Let K be a class of τ_0-frames. Then K is definable within the class of transitive finite τ-frames if and only if it is closed under taking (finite) disjoint unions, generated subframes, and bounded morphic images.*

Proof. The right to left direction is immediate: we know from the previous section that any modally definable frame class is closed under these operations. So let's consider the more interesting converse.

Assume that K satisfies the stated closure condition. Let Λ_{K} be the logic of K; that is, $\Lambda_{\mathsf{K}} = \{\phi \mid \mathfrak{F} \Vdash \phi$, for all $\mathfrak{F} \in \mathsf{K}\}$. We will show that Λ_{K} defines K. Clearly Λ_{K} is valid on every frame in K, so to complete the proof we need to show that if $\mathfrak{F} \Vdash \Lambda_{\mathsf{K}}$, where $\mathfrak{F}$ is finite and transitive, then $\mathfrak{F} \in \mathsf{K}$. We split the proof into two cases.

First suppose that $\mathfrak{F}$ is point-generated with root w. Consider the Jankov-Fine formula $\phi_{\mathfrak{F},w}$ for $\mathfrak{F}$ and w. Clearly $\phi_{\mathfrak{F},w}$ is satisfiable in $\mathfrak{F}$ at w, so $\neg\phi_{\mathfrak{F},w} \notin \Lambda_{\mathsf{K}}$. Hence there is some $\mathfrak{G} \in \mathsf{K}$ such that $\mathfrak{G} \not\Vdash \neg\phi_{\mathfrak{F},w}$; in other words, for some valuation V and state v we have $(\mathfrak{G}, V), v \Vdash \phi_{\mathfrak{F},w}$. Thus by the previous lemma,

$\mathfrak{F}$ is a bounded morphic image of the point-generated subframe $\mathfrak{G}_v$ of $\mathfrak{G}$. By the closure conditions on K, it follows that $\mathfrak{F} \in \mathsf{K}$.

So suppose that $\mathfrak{F}$ is *not* point-generated. But then as $\mathfrak{F} \Vdash \Lambda_{\mathsf{K}}$, so does each point-generated subframe of $\mathfrak{F}$, hence by the work of the previous paragraph all these subframes belong to K. But by Exercise 3.3.4, $\mathfrak{F}$ is a bounded morphic image of the disjoint union of its rooted generated subframes, so $\mathfrak{F}$ belongs to K too. ⊣

The finite frame property

Our next result deals not with frame definability, but with the relationship between normal modal logics and finite frames. Normal modal logics were introduced in Section 1.6 (see in particular Definition 1.42). Recall that normal modal logics are sets of formulas (containing certain axioms) that are closed under three simple conditions (modus ponens, uniform substitution, and generalization). They are the standard tool for capturing the notion of validity *syntactically*.

Now, in Section 2.3 we introduced the finite *model* property. We did not apply the concept to normal modal logics — but as a normal logic is simply a set of formulas, we can easily extend the definition to permit this:

Definition 3.22 A normal modal logic Λ has the finite model property with respect to some class of models M if $\mathsf{M} \Vdash \Lambda$ and every formula *not* in Λ is refuted in a *finite* model $\mathfrak{M}$ in M. Λ has the finite model property if it has the finite model property with respect to some class of models. ⊣

Informally, if a normal modal logic has the finite model property, it has a finite *semantic* characterization: *it is precisely the set of formulas that some collection of finite models makes globally true*. This is an attractive property, and as we'll see in Chapter 6 when we discuss the decidability of normal logics, a useful one too.

But something seems wrong. It is the level of *frames*, rather than the level of models, which supports the key logical concept of validity. It certainly seems sensible to try and semantically characterize normal logics in terms of finite structures — but it seems we should do so using finite *frames*, not finite models. That is, the following property seems more appropriate:

Definition 3.23 (Finite Frame Property) Let Λ be a normal modal logic and F a class of finite frames. We say Λ has the *finite frame property with respect to* F if and only if $\mathsf{F} \Vdash \Lambda$, and for every formula ϕ such that $\phi \notin \Lambda$ there is some $\mathfrak{F} \in \mathsf{F}$ such that ϕ is falsifiable on $\mathfrak{F}$. We say Λ has the *finite frame property* if and only if it has the finite frame property with respect to some class of finite frames. ⊣

Note that to establish the finite frame property of a normal modal logic Λ, it is not sufficient to prove that any formula $\phi \notin \Lambda$ can be refuted on a model where Λ is globally true: in addition one has to ensure that the underlying *frame* of the model validates Λ. If a logic has the finite frame property (and many important ones do, as we will learn in Chapter 6) then clearly there is no room for argument: it really can be characterized semantically in terms of finite structures.

But now for a surprising result. The finite frame property is *not* stronger than the finite model property: we will show that a normal modal logic has the finite frame property if and only if it has the finite model property. This result will prove useful at a number of places in Chapters 4 and 6. Moreover, while proving it we'll meet some other concepts, notably *definable variants* and *distinguishing models*, which will be useful when proving Bull's Theorem in Section 4.9.

Definition 3.24 (Definable Variant) Let $\mathfrak{M} = (W, R, V)$ be a model and $U \subseteq W$. We say U is *definable in* $\mathfrak{M}$ if and only if there is a formula ϕ_U such that for all states $w \in W$, $\mathfrak{M}, w \Vdash \phi_U$ iff $w \in U$.

Any model $\mathfrak{M}'$ based on the frame (W, R) is called a *variant* of $\mathfrak{M}$. A variant (W, R, V') of $\mathfrak{M}$ is *definable in* $\mathfrak{M}$ if and only if for all proposition symbols p, $V'(p)$ is definable in $\mathfrak{M}$. If $\mathfrak{M}'$ is a variant of $\mathfrak{M}$ that is definable in $\mathfrak{M}$, we call $\mathfrak{M}'$ a *definable variant* of $\mathfrak{M}$. $\dashv$

Recall that normal modal logics are closed under uniform substitution, the process of uniformly replacing propositional symbols with arbitrary formulas (see Section 1.6), and that a formula obtained from ϕ by uniform substitution is called a *substitution instance* of ϕ. Our intuitive understanding of uniform substitution suffices for most purposes, but in order to prove the following lemma we need to refer to the precise concepts of Definition 1.18.

Lemma 3.25 *Let $\mathfrak{M} = (\mathfrak{F}, V)$ be a model and $\mathfrak{M}' = (\mathfrak{F}, V')$ be a definable variant of $\mathfrak{M}$. For any formula ϕ, let ϕ' be the result of uniformly replacing each atomic symbol p in ϕ by $\phi_{V'(p)}$, where $\phi_{V'(p)}$ defines $V'(p)$ in $\mathfrak{M}$. Then for all formulas ϕ, and all normal modal logics Λ:*

- (i) $\mathfrak{M}', w \Vdash \phi$ *iff* $\mathfrak{M}, w \Vdash \phi'$.
- (ii) *If every substitution instance of ϕ is true in $\mathfrak{M}$, then every substitution instance of ϕ is true in $\mathfrak{M}'$.*
- (iii) *If $\mathfrak{M} \Vdash \Lambda$ then $\mathfrak{M}' \Vdash \Lambda$.*

Proof. Item (i) follows by induction on ϕ. For the base case we have $\mathfrak{M}', w \Vdash p$ iff $\mathfrak{M}, w \Vdash \phi_{V'(p)}$. As $(\neg\phi)' = \neg\phi'$, $(\phi \vee \psi)' = \phi' \vee \psi'$, and $(\Diamond\phi)' = \Diamond\phi'$ (cf. Definition 1.18) the inductive steps are immediate.

For item (ii), we show the contrapositive. Let ψ be a substitution instance of ϕ and suppose that $\mathfrak{M}' \nVdash \psi$. Thus there is some w in $\mathfrak{M}$ such that $\mathfrak{M}', w \nVdash \psi$. By item (i), $\mathfrak{M}, w \nVdash \psi'$, which means that $\mathfrak{M} \nVdash \psi'$. But as ψ' is a substitution instance of ψ, and ψ is a substitution instance of ϕ, we have that ψ' is a substitution instance of ϕ (see Exercise 1.2.5) and the result follows.

Item (iii) is an immediate consequence of item (ii), for normal modal logics are closed under uniform substitution. $\dashv$

We now isolate a type of model capable of defining *all* its variants:

Definition 3.26 (Distinguishing Model) A model $\mathfrak{M}$ is *distinguishing* if the relation $\leftrightsquigarrow$ of modal equivalence between states of $\mathfrak{M}$ is the identity relation. $\dashv$

In other words, a model $\mathfrak{M}$ is *distinguishing* if and only if for all states w and u in $\mathfrak{M}$, if $w \neq u$, then there is a formula ϕ such that $\mathfrak{M}, w \Vdash \phi$ and $\mathfrak{M}, u \nVdash \phi$. Many important models are distinguishing. For example, all filtrations (see Definition 2.36) are distinguishing. Moreover, the canonical models introduced in Section 4.2 are distinguishing too. And, and as we will now see, when a distinguishing model is *finite*, it can define all its variants.

Lemma 3.27 *Let $\mathfrak{M} = (\mathfrak{F}, V)$ be a finite distinguishing model. Then:*

(i) *For every state w in $\mathfrak{M}$ there is a formula ϕ_w that is true at, and only at, w.*
(ii) *$\mathfrak{M}$ can define any subset of $\mathfrak{F}$. Hence $\mathfrak{M}$ can define all its variants.*
(iii) *If $\mathfrak{M} \Vdash \phi$ then $\mathfrak{F} \Vdash \phi$.*

Proof. For item (i), suppose that $\mathfrak{F} = (W, R)$, and enumerate the states in W as $w_1, \ldots, w_n$. For all pairs (i, j) such that $1 \leq i, j \leq n$ and $i \neq j$, choose $\phi_{i,j}$ to be a formula such that $\mathfrak{M}, w_i \Vdash \phi_{i,j}$ and $\mathfrak{M}, w_j \nVdash \phi_{i,j}$ (such a formula exists, for $\mathfrak{M}$ is distinguishing) and define ϕ_{w_i} to be $\phi_{i,1} \wedge \cdots \wedge \phi_{i,n}$. Clearly ϕ_{w_i} is true at w_i and false everywhere else.

Item (ii) is an easy consequence. For let U be any subset of W. Then $\bigvee_{w \in U} \phi_w$ defines U. Hence as $\mathfrak{M}$ can define all subsets of W, it can define $V'(p)$, for any valuation V' on $\mathfrak{F}$ and propositional symbol p.

As for item (iii), suppose $\mathfrak{M} \Vdash \phi$. By item (iii) of the previous lemma we have that $\mathfrak{M}' \Vdash \phi$, where $\mathfrak{M}'$ is any definable variant of $\mathfrak{M}$. But we have just seen that $\mathfrak{M}$ can define all its variants, hence $\mathfrak{F} \Vdash \phi$. $\dashv$

Lemmas 3.25 and 3.27 will be important in their own right when we prove Bull's theorem in Section 4.9. And with the help of a neat filtration argument, they yield the main result:

Theorem 3.28 *A normal modal logic has the finite frame property iff it has the finite model property.*

Proof. The left to right direction is immediate. For the converse, suppose that Λ is a normal modal logic with the finite model property. Since we will need to take a filtration through Λ, we have to be explicit about the set of proposition letters of the formulas in Λ, so assume that $\Lambda \subseteq Form(\tau, \Phi)$.

Take a formula in the language $ML(\tau, \Phi)$ that does not belong to Λ. We will show that ϕ can be refuted on a finite frame $\mathfrak{F}$ such that $\mathfrak{F} \Vdash \Lambda$.

As Λ has the finite model property, there is a finite model $\mathfrak{M}$ such that $\mathfrak{M} \Vdash \Lambda$ and $\mathfrak{M}, w \nVdash \phi$ for some state w in $\mathfrak{M}$. Let Σ be the set of all subformulas of formulas in $\{\phi\} \cup \Lambda$, and let $\mathfrak{M}^f$ be any filtration of $\mathfrak{M}$ through Σ. As $\mathfrak{M}$ is finite, so is $\mathfrak{M}^f$. As $\mathfrak{M}^f$ is a filtration, it is a distinguishing model. By the Filtration Theorem (Theorem 2.39), $\mathfrak{M}^f, |w| \nVdash \phi$. Moreover $\mathfrak{M}^f \Vdash \Lambda$, for as every state in $\mathfrak{M}$ satisfies all formulas in Λ, so does every state in $\mathfrak{M}^f$ (again, this follows from the Filtration Theorem). Let $\mathfrak{F}$ be the (finite) frame underlying $\mathfrak{M}^f$. By Lemma 3.27 item (iii), $\mathfrak{F} \Vdash \Lambda$, and we have proved the theorem. $\dashv$

Note the somewhat unusual use of filtrations in this proof. Normally we filtrate infinite models through finite sets of formulas. Here we filtrated a *finite* model through an *infinite* sets of formulas to guarantee that an entire logic remained true.

This result shows that the concepts of normal modal logics and frame validity fit together well in the finite domain: if a normal logic has a finite semantic characterization in terms of models, then it is guaranteed to have a finite *frame-based* semantic characterization as well. But be warned: one of the most striking results of the following chapter is that logics and frame validity don't always fit together so neatly. In fact, the *frame incompleteness results* will eventually lead us (in Chapter 5) to the use of new semantic structures, namely modal algebras, to analyze normal modal logics. But this is jumping ahead. It's time to revert to our discussion of frame definability — but from a rather different perspective. So far, our approach has been firmly *semantical*. This has taught us a lot: in particular, the Goldblatt-Thomason theorem has given us a model-theoretic characterization of the elementary frame classes that are modally definable. Moreover, we will see in Chapter 5 that the semantic approach has an important algebraic dimension. But it is also possible to approach frame definability from a more *syntactic* perspective, and that's what we're going to do now. This will lead us to the other main result of the chapter: the Sahlqvist Correspondence Theorem.

Exercises for Section 3.4

3.4.1 Prove Lemma 3.20. That is, suppose that $\phi_{\mathfrak{F},w}$ is the Jankov-Fine formula for a transitive finite frame $\mathfrak{F}$ with root w. Show that for any frame $\mathfrak{G}$, $\phi_{\mathfrak{F},w}$ is satisfiable on $\mathfrak{G}$ at a node v iff $\mathfrak{F}$ is a bounded morphic image of $\mathfrak{G}_v$.

3.4.2 Let $\mathfrak{M}$ be a model, let $\mathfrak{M}^f$ be any filtration of $\mathfrak{M}$ through some finite set of formulas Σ, and let f be the natural map associated with the filtration. If u is a point in the filtration, show that $f^{-1}[u]$ is definable in $\mathfrak{M}$.

3.5 Automatic First-Order Correspondence

We have learned a lot about frame definability in the previous sections. In particular, we have learned that frame definability is a second-order notion, and that the second-order correspondent of any modal formula can be straightforwardly computed using the Second-Order Translation. Moreover, we know that many modal formulas have first-order correspondents, and that the Goldblatt-Thomason Theorem gives us a model-theoretic characterization of the frame classes they define.

Nonetheless, there remains a gap in our understanding: although many modal formulas define first-order conditions on frames, it is not really clear *why* they do so. To put it another way, in many cases the (often difficult to decipher) second-order condition yielded by the second-order translation is equivalent to a much simpler first-order condition. Is there any system to this? Better, are there algorithms that enable us to compute first-order correspondents automatically, and if so, how general are these algorithms? This section, and the two that follow, develop some answers.

A large part of this work centers on a beautiful positive result: there is a large class of formulas, the *Sahlqvist formulas*, each of which defines a first-order condition on frames which is effectively calculable using the *Sahlqvist-van Benthem algorithm*; this is the celebrated *Sahlqvist Correspondence Theorem*, which we will state and prove in the following section. The proof of this theorem sheds light on why so many second-order correspondents turn out to be equivalent to a first-order condition. Moreover each Sahlqvist formula is *complete* with respect to the class of first-order frames it defines; this is the *Sahlqvist Completeness Theorem*, which we will formulate more precisely in Theorem 4.42 and prove in Section 5.6. All in all, the Sahlqvist fragment is interesting from both theoretical and practical perspectives, and we devote a lot of attention to it.

In this section we lay the groundwork for the proof of the Sahlqvist Correspondence Theorem. We are going to introduce two simple classes of modal formulas, the *closed* formulas and the *uniform* formulas, and show that they define first-order conditions on frames. Along the way we are going to learn about positive and negative formulas, what they have to do with monotonicity, and how they can help us get rid of second-order quantifiers. These ideas will be put to work, in a more sophisticated way, in the following section.

One other thing: in what follows we are going to work with a stronger notion of correspondence. The concept of correspondence given in Definition 3.5 is *global*: a modal and a (first- or second-order) frame formula are called correspondents if they are valid on precisely the same frames. But it is natural to demand that validity matches *locally*:

Definition 3.29 (Local Frame Correspondence) Let ϕ be a modal formula in some similarity type, and $\alpha(x)$ a formula in the corresponding first- or second-order frame language (x is supposed to be the only free variable of α). Then we say that ϕ and $\alpha(x)$ are *local frame correspondents* of each other if the following holds, for any frame $\mathfrak{F}$ and any state w of $\mathfrak{F}$:

$$\mathfrak{F}, w \Vdash \phi \text{ iff } \mathfrak{F} \models \alpha[w]. \quad \dashv$$

In fact, we've been implicitly using local correspondence all along. In Example 3.6 we showed that $p \to \Diamond p$ corresponds to $\forall x Rxx$ — but inspection of the proof reveals we did so by showing that $p \to \Diamond p$ *locally* corresponds to Rxx. Similarly, in Example 3.7 we showed that $\Diamond p \to \Diamond\Diamond p$ corresponds to density by showing that $\Diamond p \to \Diamond\Diamond p$ *locally* corresponds to $\forall yz\,(Rxy \wedge Ryz \to Rxz)$. It should be clear from these examples that the local notion of correspondence is fundamental, and that the following connection holds between the local and global notions:

Proposition 3.30 *If $\alpha(x)$ is a local correspondent of the modal formula ϕ, then $\forall x\, \alpha(x)$ is a global correspondent of ϕ. So if ϕ has a first-order local correspondent, then it also has a first-order global correspondent.*

Proof. Trivial. $\dashv$

What about the converse? In particular, suppose that the modal formula ϕ has a *first order* global correspondent; will it also have a first-order local correspondent? Intriguingly, the answer to this question is negative, as we will see in Example 3.57.

But until we come to this result, we won't mention global correspondence much: it's simpler to state and prove results in terms of local correspondence, relying on the previous lemma to guarantee correspondence in the global sense. With this point settled, it's time to start thinking about correspondence theory systematically.

Closed formulas

There is one obvious class of modal formulas guaranteed to correspond to first-order frame conditions: formulas which contain no proposition letters.

Example 3.31 Consider the basic temporal language. The formula $P\top$ defines the property that there is no first point of time. More precisely, $P\top$ is valid on precisely those frames such that every point has a predecessor.

Now, obviously it is easy to prove this directly, but for present purposes the following argument is more interesting. By Proposition 3.12, for any bidirectional frame $\mathfrak{F}$ and any point w in $\mathfrak{F}$ we have that:

$$\mathfrak{F}, w \Vdash P\top \;\text{ iff }\; \mathfrak{F} \models \forall P_1 \ldots \forall P_n\, ST_x(P\top)[w],$$

where $P_1, \ldots, P_n$ are the unary predicate variables corresponding to the proposition letters $p_1, \ldots, p_n$ occurring in $P\top$. But $P\top$ contains *no* propositional variables, hence there are no second-order quantifiers, and hence:

$$\mathfrak{F}, w \Vdash P\top \text{ iff } \mathfrak{F} \models ST_x(P\top)[w].$$

But $ST_x(P\top)$ is $\exists y\,(Ryx \wedge y = y)$, which is equivalent to $\exists y\,Ryx$. So $P\top$ *locally* corresponds to $\exists y\,Ryx$ (and thus *globally* corresponds to $\forall x \exists y\,Ryx$). $\dashv$

The argument used in this example is extremely simple, and obviously generalizes. We'll state and prove the required generalization, and then move on to richer pastures.

Definition 3.32 A modal formula ϕ is *closed* if and only if it contains no proposition letters. Thus closed formulas are built up from $\top$, $\bot$, and any nullary modalities (or modal constants) the signature may contain. $\dashv$

Proposition 3.33 *Let ϕ be a closed formula. Then ϕ locally corresponds to a first-order formula $c_\phi(x)$ which is effectively computable from ϕ.*

Proof. By Proposition 3.12 and the fact that ϕ contains no propositional variables we have:

$$\mathfrak{F}, w \Vdash \phi \text{ iff } \mathfrak{F} \models ST_x(\phi)[w].$$

As it is easy to write a program that computes $ST_x(\phi)$, the claim follows immediately. $\dashv$

Closed formulas arise naturally in some applications (a noteworthy example is provability logic), thus the preceding result is quite useful in practice.

Uniform formulas

Although the previous proposition was extremely simple, it does point the way to the strategy followed in our approach to the Sahlqvist Correspondence Theorem: we are going to look for ways of stripping off the initial block of monadic second-order universal quantifiers in $\forall P_1 \ldots \forall P_n ST_x(\phi)$, thus reducing the translation to $ST_x(\phi)$. The obvious way of getting rid of universal quantifiers is to perform universal instantiation, and this is exactly what we will do. Both here, and in the work of the next section, we will look for simple instantiations for the $P_1, \ldots, P_n$, which result in first-order formulas equivalent to the original. We will be able to make this strategy work because of the syntactic restrictions placed on ϕ.

One of the restrictions imposed on Sahlqvist formulas invokes the idea of *positive* and *negative* occurrences of proposition letters. We now introduce this idea, study its semantic significance, and then, as an introduction to the techniques of the following section, use a simple instantiation argument to show that the second-order translations of *uniform* formulas are effectively reducible to first-order conditions on frames.

Definition 3.34 An occurrence of a proposition letter p is a *positive* occurrence if it is in the scope of an even number of negation signs; it is a *negative* occurrence if it is in the scope of an odd number of negation signs. (This is one of the few places in the book where it is important to think in terms of the primitive connectives. For example, the occurrence of p in $\Diamond(p \to q)$ is *negative*, for this formula is shorthand for $\Diamond(\neg p \vee q)$.) A modal formula ϕ is *positive in* p (*negative in* p) if all occurrences of p in ϕ are positive (negative). A formula is called *positive* (*negative*) if it is positive (negative) in all proposition letters occurring in it.

Analogous concepts are defined for the corresponding second-order language. That is, an occurrence of a unary predicate variable P in a second-order formula is *positive* (*negative*) if it is in the scope of an even (odd) number of negation signs. A second-order formula ϕ is *positive in* P (*negative in* P) if all occurrences of P in ϕ are positive (negative), and it is called *positive* (*negative*) if it is positive (negative) in all unary predicate variables occurring in it. $\dashv$

Lemma 3.35 *Let ϕ be a modal formula.*

(i) *ϕ is positive in p iff $ST_x(\phi)$ is positive in the corresponding unary predicate P.*
(ii) *If ϕ is positive (negative) in p, then $\neg\phi$ is negative (positive) in p.*

Proof. Virtually immediate. $\dashv$

Positive and negative formulas are important because of their special semantic properties. In particular, they exhibit a useful form of *monotonicity*.

Definition 3.36 Fix a modal language $ML(\tau, \Phi)$, and let $p \in \Phi$. A modal formula ϕ is *upward monotone in* p if its truth is preserved under extensions of the interpretation of p. More precisely, ϕ is upward monotone in p if for every model $(W, R_\triangle, V)_{\triangle\in\tau}$, every state $w \in W$, and every valuation V' such that $V(p) \subseteq V'(p)$ and for all $q \neq p$, $V(q) = V'(q)$, the following holds:

$$\text{if } (W, R_\triangle, V)_{\triangle\in\tau}, w \Vdash \phi, \text{ then } (W, R_\triangle, V')_{\triangle\in\tau}, w \Vdash \phi.$$

In short, extending $V(p)$ (while leaving the interpretation of any other propositional variable unchanged) has the effect of extending $V(\phi)$ (or keeping it the same).

Likewise, a formula ϕ is *downward monotone in* p if its truth is preserved under shrinkings of the interpretation of p. That is, for every model $(W, R_\triangle, V)_{\triangle\in\tau}$, every state $w \in W$, and every valuation V' such that $V'(p) \subseteq V(p)$ and for all $q \neq p$, $V(q) = V'(q)$, the following holds:

$$\text{if } (W, R_\triangle, V)_{\triangle\in\tau}, w \Vdash \phi, \text{ then } (W, R_\triangle, V')_{\triangle\in\tau}, w \Vdash \phi.$$

The notions of a second-order formula being *upward* and *downward monotone in* a unary predicate variable P are defined analogously; we leave this task to the reader. $\dashv$

Lemma 3.37 *Let ϕ be a modal formula.*

(i) *If ϕ is positive in p, then it is upward monotone in p.*
(ii) *If ϕ is negative in p, then it is downward monotone in p.*

Proof. Prove both parts simultaneously by induction on ϕ; see Exercise 3.5.3. $\dashv$

But what do upward and downward monotonicity have to do with frame definability? The following example is instructive.

Example 3.38 The formula $\Diamond\Box p$ locally corresponds to a first-order formula. For suppose $\mathfrak{F}, w \Vdash \Diamond\Box p$. Regardless of the valuation, the formula $\Diamond\Box p$ holds at w. So consider a *minimal* valuation (for p) on $\mathfrak{F}$; that is, choose any V_m such that $V_m(p) = \varnothing$. Then as $w \Vdash \Diamond\Box p$, there must be a successor v of w such that $\Box p$ holds at v. However, there are no p-states, so v must be blind (that is, without successors). In other words, we have shown that

$$(\mathfrak{F}, V_m), w \Vdash \Diamond\Box p \text{ only if } \mathfrak{F} \models \exists y\,(Rxy \wedge \neg\exists z\, Ryz)[w].$$

Now for the interesting direction: assume that the state w in the frame $\mathfrak{F}$ has a blind successor. It follows immediately that $(\mathfrak{F}, V_m), w \Vdash \Diamond\Box p$, where V_m is any minimal valuation (for p). We claim that the formula $\Diamond\Box p$ is valid at w. To see this, consider an arbitrary valuation V and a point w of $\mathfrak{F}$. By item (i) of Lemma 3.37, $\Diamond\Box p$ is upward monotone in p. Hence it follows from the fact that $V_m(p) \subseteq V(p)$ that $(\mathfrak{F}, V), w \Vdash \Diamond\Box p$. As V was arbitrary, $\Diamond\Box p$ is valid on $\mathfrak{F}$ at w. $\dashv$

The key point is the last part of the argument: the use of a minimal valuation followed by an appeal to monotonicity to establish a result about *all* valuations. But now think about this argument from the perspective of the second-order correspondence language: in effect, we *instantiated* the predicate variable corresponding to p with the smallest subset of the frame possible, and then used a monotonicity argument to establish a result about *all* assignments to P.

This simple idea lies behind much of our work on the Sahlqvist fragment. To illustrate the style of argumentation it leads to, we will now use an instantiation argument to show that all *uniform* modal formulas define first-order conditions on frames.

Definition 3.39 A proposition letter p occurs *uniformly* in a modal formula if it occurs only positively, or only negatively. A predicate variable P occurs uniformly in a second-order formula if it occurs only positively, or only negatively. A modal formula is *uniform* if all the propositional letters it contains occur uniformly. A second-order formula is uniform if all the unary predicate variables it contains occur uniformly. $\dashv$

Theorem 3.40 *If ϕ is a uniform modal formula, then ϕ locally corresponds to a first-order formula $c_\phi(x)$ on frames. Moreover, c_ϕ is effectively computable from ϕ.*

Proof. Consider the universally quantified second-order equivalent of ϕ:

$$\forall P_1 \ldots \forall P_n \, ST_x(\phi), \tag{3.6}$$

where $P_1, \ldots, P_n$ are second-order variables corresponding to the proposition letters in ϕ. Our aim is to show that (3.6) is equivalent to a first-order formula by performing appropriate instantiations for the universally quantified monadic second-order variables $P_1, \ldots P_n$.

As ϕ is uniform, by Lemma 3.35 so is $ST_x(\phi)$. We will instantiate the unary predicates that occur positively with a predicate denoting as small a set as possible (that is, the empty set), and the unary predicates that occur negatively with a predicate denoting as large a set as possible (that is, all the states in the frame). We will use Church's λ-notation for the required substitution instance providing the formulas that define these predicates. For every P occurring in $ST_x(\phi)$, define

$$\sigma(P) \equiv \begin{cases} \lambda u.\, u \neq u, & \text{if } ST_x(\phi) \text{ is positive in } P \\ \lambda u.\, u = u, & \text{if } ST_x(\phi) \text{ is negative in } P. \end{cases}$$

Of course, the idea is that instantiating a universal second-order formula according to this substitution σ simply means (i) removing the second-order quantifiers and (ii) replacing every atomic subformula Py with the formula $\sigma(P)(y)$, that is, with either $y \neq y$ or $y = y$ (as given by the definition).[1]

Now consider the following instance of (3.6) in which every unary predicate P has been replaced by $\sigma(P)$:

$$[\sigma(P_1)/P_1, \ldots, \sigma(P_n)/P_n] \, ST_x(\phi). \tag{3.7}$$

We will show that (3.7) is equivalent to (3.6). It is immediate that (3.6) implies (3.7), for the latter is an instantiation of the former. For the converse implication we assume that

$$\mathfrak{M} \models [\sigma(P_1)/P_1, \ldots, \sigma(P_n)/P_n] \, ST_x(\phi)[w], \tag{3.8}$$

[1] If you are unfamiliar with λ-notation, all you really need to know to follow the proof is that $\lambda u.\, u \neq u$ and $\lambda u.\, u = u$ are predicates denoting the empty set and the set of all states respectively. Some explanatory remarks on λ-notation are given following the proof.

and we have to show that

$$\mathfrak{M} \models \forall P_1 \ldots \forall P_n \, ST_x(\phi)[w].$$

By the choice of $\sigma(P)$, for predicates P that occur only positively in $ST_x(\phi)$ we have that $\mathfrak{M} \models \forall y\,(\sigma(P)(y) \to P(y))$, and for predicates P that occur only negatively in $ST_x(\phi)$, we have that $\mathfrak{M} \models \forall y\,(P(y) \to \sigma(P)(y))$. (Readers familiar with λ-notation will realize that we have implicitly appealed to β-conversion here. Readers unfamiliar with λ-notation should simply note that when $\sigma(P)$ is a predicate denoting the empty set, then $\sigma(P)(y)$ is false no matter what y denotes, while if $\sigma(P)$ denotes the set of all states, $\sigma(P)(y)$ is guaranteed to be true.) Hence, as $ST_x(\phi)$ is positive or negative in all unary predicates P occurring in it, (3.8) together with Lemma 3.37 imply that for *any* choice of $P_1, \ldots, P_n$,

$$(\mathfrak{M}, P_1, \ldots, P_n) \models ST_x(\phi)[w],$$

which means that $\mathfrak{M} \models \forall P_1 \ldots \forall P_n \, ST_x(\phi)$ as required. Finally, in any programming language with decent symbol manipulation facilities it is straightforward to write a program which, when given a uniform formula ϕ, produces $ST_x(\phi)$ and carries out the required instantiations. Hence the first-order correspondents of uniform formulas are computable. $\dashv$

On λ-notation

Although it is not essential to use λ-notation, it *is* convenient and we will apply it in the following section. For readers unfamiliar with it, here's a quick introduction to the fundamental ideas.

We have used Church's λ-notation as a way of writing predicates, that is, entities which denote subsets. But lambda expressions don't denote subsets directly; rather they denote their *characteristic functions*. Suppose we are working with a frame (W, R). Let $S \subseteq W$. Then the characteristic function of S (with respect to W) is the function χ_S with domain W and range $\{0, 1\}$ such that $\chi_S(s) = 1$ if $s \in S$ and $\chi_S(s) = 0$ otherwise. Reading 1 as true and 0 as false, χ_S is simply the function that says truthfully of each element of W whether it belongs to S or not.

Lambda expressions pick out characteristic functions in the obvious way. For example, when working with a frame (W, R), $\lambda u.\, u \neq u$ denotes the function from W to $\{0, 1\}$ that assigns 1 to every element $w \in W$ that satisfies $u \neq u$ and 0 to everything else. But for *no* choice of w is it the case that $w \neq w$; hence, as we stated in the previous proof, $\lambda u.\, u \neq u$ denotes the characteristic function of the empty set. Similarly, $\lambda u.\, u = u$ denotes the characteristic function of W, for $w = w$ for every $w \in W$.

Lambda expressions take the drudgery out of dealing with substitutions. Consider the second-order formula Px. This is satisfied in a model if and only if the element assigned to x belongs to the subset assigned to P. For example, if P is assigned the empty set, Px will be false no matter what x is assigned. Now suppose we substitute $(\lambda u.\, u \neq u)$ for P in Px. This yields the expression $(\lambda u.\, u \neq u)x$. Read this as 'apply the function denoted by $\lambda u.\, u \neq u$ to the state denoted by x'. Clearly this yields the value 0 (that is, *false*). The process of β-conversion mentioned in the proof is essentially a way of rewriting such functional applications to simpler but equivalent forms; for more details, consult one of the introductions cited in the Notes. Newcomers to λ-notation should try Exercise 3.5.1 right away.

Exercises for Section 3.5

3.5.1 Explain why we could have used the following predicate definitions in the proof of Theorem 3.38: for every P occurring in $ST_x(\phi)$, define

$$\sigma(P) \equiv \begin{cases} \lambda u.\, \bot, & \text{if } ST_x(\phi) \text{ is positive in } P \\ \lambda u.\, \top, & \text{if } ST_x(\phi) \text{ is negative in } P. \end{cases}$$

If you have difficulties with this, consult one of the introductions to λ-calculus cited in the notes before proceeding further.

3.5.2 Let ϕ be a modal formula which is positive in all propositional variables. Prove that ϕ can be rewritten into a normal form which is built up from proposition letters, using $\wedge$, $\vee$, $\Diamond$ and $\Box$ only.

3.5.3 Prove Lemma 3.37. That is, show that if a modal formula ϕ is positive in p, then it is upward monotone in p, and that if it is negative in p, then it is downward monotone in p.

3.6 Sahlqvist Formulas

In the proof of Theorem 3.40 we showed that uniform formulas correspond to first-order conditions by finding a suitable *instantiation* for the universally quantified monadic second-order variables in their second-order translation and appealing to *monotonicity*. This is an important idea, and the rest of this section is devoted to extending it: the Sahlqvist fragment is essentially a large class of formulas to which this style of argument can be applied.

Very simple Sahlqvist formulas

Roughly speaking, Sahlqvist formulas are built up from implications $\phi \to \psi$, where ψ is positive and ϕ is of a restricted form (to be specified below) from which the required instantiations can be read off. We now define a limited version of the Sahlqvist fragment for the basic modal language; generalizations and extensions will be discussed shortly.

Definition 3.41 We will work in the basic modal language. A *very simple Sahlqvist antecedent* over this language is a formula built up from $\top$, $\bot$ and proposition letters, using only $\wedge$ and $\Diamond$. A *very simple Sahlqvist formula* is an implication $\phi \to \psi$ in which ψ is positive and ϕ is a very simple Sahlqvist antecedent. $\dashv$

Examples of very simple Sahlqvist formulas include $p \to \Diamond p$ and $(p \wedge \Diamond\Diamond q) \to \Box\Diamond(p \wedge q)$.

The following theorem is central for understanding what Sahlqvist correspondence is all about. Its proof describes and justifies an algorithm for converting simple Sahlqvist formulas into first-order formulas; the algorithms given later for richer Sahlqvist fragments elaborate on ideas introduced here. Examples of the algorithm in action are given below; it is a good idea to refer to these while studying the proof.

Theorem 3.42 *Let $\chi = \phi \to \psi$ be a very simple Sahlqvist formula in the basic modal language $ML(\tau_0, \Phi)$. Then χ locally corresponds to a first-order formula $c_\chi(x)$ on frames. Moreover, c_χ is effectively computable from χ.*

Proof. Our starting point is the formula $\forall P_1 \ldots \forall P_n\, (ST_x(\phi) \to ST_x(\psi))$, which is the local second-order translation of χ. We assume that this translation has undergone a pre-processing step to ensure that no two quantifiers bind the same variable, and no quantifier binds x. Let us denote $ST_x(\psi)$ by POS; that is, we have a translation of the form:

$$\forall P_1 \ldots \forall P_n\, (ST_x(\phi) \to \text{POS}). \tag{3.9}$$

We will now rewrite (3.9) to a form from which we can read off the instantiations that will yield its first-order equivalent.

Step 1. Pull out diamonds.
Use equivalences of the form

$$(\exists x_i\, \alpha(x_i) \wedge \beta) \;\leftrightarrow\; \exists x_i\, (\alpha(x_i) \wedge \beta)$$

and

$$(\exists x_i\, \alpha(x_i) \to \beta) \;\leftrightarrow\; \forall x_i\, (\alpha(x_i) \to \beta)$$

(in that order) to move all existential quantifiers in the antecedent $ST_x(\phi)$ of (3.9) to the front of the implication. Note that by our definition of Sahlqvist antecedents, the existential quantifiers only have to cross conjunctions before they reach the main implication. Of course, the above equivalences are not valid if the variable x_i occurs freely in β, but by our assumption on the pre-processing of the formula, this problem does not arise.

Step 1 results in a formula of the form

$$\forall P_1 \ldots \forall P_n \forall x_1 \ldots \forall x_m\, (\text{REL} \wedge \text{AT} \to \text{POS}), \tag{3.10}$$

where REL is a conjunction of atomic first-order statements of the form Rx_ix_j corresponding to occurrences of diamonds, and AT is a conjunction of (translations of) proposition letters. It may be helpful at this point to look at the concrete examples given below.

Step 2. Read off instances.
We can assume that every unary predicate P that occurs in the consequent of the matrix of (3.10), also occurs in the antecedent of the matrix of (3.10): otherwise (3.10) is positive in P and we can substitute $\lambda u.\, u \neq u$ for P (that is, make use of the substitution used in the proof of Theorem 3.40) to obtain an equivalent formula without occurrences of P.

Let P_i be a unary predicate occurring in (3.10), and let $P_ix_{i_1}, \dots, P_ix_{i_k}$ be all the occurrences of the predicate P_i in the antecedent of (3.10). Define

$$\sigma(P_i) \equiv \lambda u.\, (u = x_{i_1} \vee \cdots \vee u = x_{i_k}).$$

Note that $\sigma(P_i)$ is the *minimal* instance making the antecedent REL $\wedge$ AT true; this lambda expression says that if a node u has property P_i, then u must be one of the nodes $x_{i_1}, x_{i_2}, \dots$ or x_{i_k} explicitly stated to have property P_i in the antecedent. But this is nothing else than saying that if some model $\mathfrak{M}$ makes the formula AT true under some assignment, then the interpretation of the predicate P must *extend* the set of points where $\sigma(P)$ holds:

$$\mathfrak{M} \models \mathrm{AT}[ww_1 \dots w_m] \text{ implies } \mathfrak{M} \models \forall y\, (\sigma(P_i)(y) \to P_iy)[ww_1 \dots w_m] \quad (3.11)$$

This observation, in combination with the positivity of the consequent of the Sahlqvist formula, forms the key to understanding why Sahlqvist formulas have first-order correspondents.

Step 3. Instantiating.
We now use the formulas of the form $\sigma(P_i)$ found in Step 2 as instantiations; we substitute $\sigma(P_i)$ for each occurrence of P_i in the first-order matrix of (3.10). This results in a formula of the form

$$[\sigma(P_1)/P_1, \dots, \sigma(P_n)/P_n]\forall x_1 \dots \forall x_m\, (\mathrm{REL} \wedge \mathrm{AT} \to \mathrm{POS}).$$

Now, there are no occurrences of monadic second-order variables in REL. Furthermore, observe that by our choice of the substitution instances $\sigma(P)$, the formula $[\sigma(P_1)/P_1, \dots, \sigma(P_n)/P_n]$AT will be trivially true. So after carrying out these substitutions we end up with a formula that is equivalent to one of the form

$$\forall x_1 \dots \forall x_m\, (\mathrm{REL} \to [\sigma(P_1)/P_1, \dots, \sigma(P_n)/P_n]\mathrm{POS}). \quad (3.12)$$

As we assumed that every unary predicate occurring in the consequent of (3.10) also occurs in its antecedent, (3.12) must be a first-order formula involving only $=$ and the relation symbol R. So, to complete the proof of the theorem it suffices to show that (3.12) is equivalent to (3.10). The implication from (3.10) to (3.12) is simply an instantiation. To prove the other implication, assume that (3.12) and the antecedent of (3.10) are true. That is, assume that

$$\mathfrak{M} \models \forall x_1 \ldots \forall x_m (\mathrm{REL} \to [\sigma(P_1)/P_1, \ldots, \sigma(P_n)/P_n]\mathrm{POS})$$

and

$$\mathfrak{M} \models \mathrm{REL} \wedge \mathrm{AT}[ww_1 \ldots w_m].$$

We need to show that $\mathfrak{M} \models \mathrm{POS}[ww_1 \ldots w_m]$. First of all, it follows from the above assumptions that

$$\mathfrak{M} \models [\sigma(P_1)/P_1, \ldots, \sigma(P_n)/P_n]\mathrm{POS}[ww_1 \ldots w_m].$$

As POS is positive, it is upwards monotone in all unary predicates occurring in it, so it suffices to show that $\mathfrak{M} \models \forall y\, (\sigma(P_i)(y) \to P_i y)[ww_1 \ldots w_m]$. But, by the essential observation (3.11) in Step 2, this is precisely what the assumption $\mathfrak{M} \models \mathrm{AT}[ww_1 \ldots w_m]$ amounts to. $\dashv$

Example 3.43 First consider the formula $p \to \Diamond p$. Its second-order translation is the formula

$$\forall P\, (\underbrace{Px}_{\mathrm{AT}} \to \exists z\, (Rxz \wedge Pz)).$$

There are no diamonds to be pulled out here, so we can read off the minimal instance $\sigma(P) \equiv \lambda u.\, u = x$ immediately. Instantiation gives

$$(\lambda u.\, u = x)x \to \exists z\, (Rxz \wedge \lambda u.\, u = x)z),$$

Which (either by β-conversion or semantic reasoning) yields the following first-order formula.

$$x = x \to \exists z\, (Rxz \wedge z = x).$$

Note that this is equivalent to Rxx.

Our second example is the density formula $\Diamond p \to \Diamond\Diamond p$, which has

$$\forall P\, (\exists x_1\, (Rxx_1 \wedge Px_1) \to \exists z_0\, (Rxz_0 \wedge \exists z_1\, (Rz_0 z_1 \wedge Pz_1))).$$

as its second-order translation. Here we can pull out the diamond $\exists x_1$:

$$\forall P \forall x_1\, (\underbrace{Rxx_1}_{\mathrm{REL}} \wedge \underbrace{Px_1}_{\mathrm{AT}} \to \exists z_0\, (Rxz_0 \wedge \exists z_1\, (Rz_0 z_1 \wedge Pz_1))).$$

Instantiating with $\sigma(P) \equiv \lambda u.\, u = x_1$ gives

$$\forall x_1\, (Rxx_1 \wedge x_1 = x_1 \to \exists z_0\, (Rxz_0 \wedge \exists z_1\, (Rz_0 z_1 \wedge z_1 = x_1))),$$

which can be simplified to $\forall x_1\, (Rxx_1 \to \exists z_0\, (Rxz_0 \wedge Rz_0 x_1))$.

Our last example of a very simple Sahlqvist formula is $(p \wedge \Diamond\Diamond p) \to \Diamond p$. Its second-order translation is

$$\forall P\,(Px \wedge \exists x_1\,(Rxx_1 \wedge \exists x_2\,(Rx_1x_2 \wedge Px_2)) \to \exists z_0\,(Rxz_0 \wedge Pz_0)).$$

Pulling out the diamonds $\exists x_1$ and $\exists x_2$ results in

$$\forall P \forall x_1 \forall x_2\,(\underbrace{Rxx_1 \wedge Rx_1x_2}_{\mathrm{REL}} \wedge \underbrace{Px \wedge Px_2}_{\mathrm{AT}} \to \exists z_0\,(Rxz_0 \wedge Pz_0)).$$

Our minimal instantiation here is: $\sigma(P) \equiv \lambda u.\,(u = x \vee u = x_2)$. After instantiating we obtain

$$\forall x_1 \forall x_2\,(Rxx_1 \wedge Rx_1x_2 \wedge (x = x \vee x = x_2) \wedge (x_2 = x \vee x_2 = x_2) \to \\ \exists z_0\,(Rxz_0 \wedge (z_0 = x \vee z_0 = x_2))).$$

This formula simplifies to $\forall x_1 \forall x_2\,(Rxx_1 \wedge Rx_1x_2 \to (Rxx \vee Rxx_2))$. $\dashv$

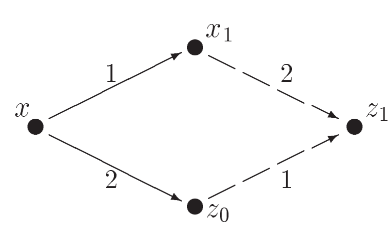
Simple Sahlqvist formulas

What is the crucial observation we need to make about the preceding proof? Simply this: the algorithm for very simple Sahlqvist formulas worked because we were able to find a minimal instantiation for their antecedents. We now show that minimal instantiations can be found for more complex Sahlqvist antecedents. First a motivating example.

Example 3.44 Consider the formula $\Diamond_1 \Box_2 p \to \Box_2 \Diamond_1 p$; we will show that this formula locally corresponds to a kind of local *confluence* (or *Church-Rosser*) property of R_1 and R_2:

$$\forall x_1 z_0\,(R_1xx_1 \wedge R_2xz_0 \to \exists z_1(R_2x_1z_1 \wedge R_1z_0z_1)).$$

The reason for the apparently unnatural choice of variable names will soon become clear, as will the somewhat roundabout approach to the proof that we take. The name 'confluence' is explained by the following picture:



Let $\mathfrak{F} = (W, R_1, R_2)$ be a frame and w a state in $\mathfrak{F}$ such that $\mathfrak{F}, w \Vdash \Diamond_1 \Box_2 p \to \Box_2 \Diamond_1 p$, and let v be a state in $\mathfrak{F}$ such that R_1wv. A sufficient condition for a valuation to make $\Diamond_1\Box_2 p$ true at w would be that p holds at all R_2-successors of v. So a *minimal* such valuation can be defined as

$$V_m(p) = \{x \in W \mid R_2 vx\}.$$

That is, V_m makes p true at *precisely* the R_2-successors of v. As $\mathfrak{F}, w \Vdash \Diamond_1\Box_2 p \to \Box_2\Diamond_1 p$, we have $(\mathfrak{F}, V_m), w \Vdash \Box_2\Diamond_1 p$, but what does this tell us about the (first-order) properties of $\mathfrak{F}$? The crucial observation is that by the choice of V_m:

$$\begin{aligned}&(\mathfrak{F}, V_m), w \Vdash \Box_2\Diamond_1 p \text{ iff}\\ &\qquad (\mathfrak{F}, V_m) \models \forall z_0\,(R_2 x z_0 \to \exists z_1\,(R_2 x_1 z_1 \wedge R_1 z_0 z_1))[wv],\end{aligned} \tag{3.13}$$

which yields that $\mathfrak{F} \models \forall x_1 z_0\,(R_1 x x_1 \wedge R_2 x z_0 \to \exists z_1\,(R_2 x_1 z_1 \wedge R_1 z_0 z_1))[w]$.

Conversely, assume that $\mathfrak{F}$ has the confluence property at w. In order to show that $\mathfrak{F}, w \Vdash \Diamond_1\Box_2 p \to \Box_2\Diamond_1 p$, let V be a valuation on $\mathfrak{F}$ such that $(\mathfrak{F}, V), w \Vdash \Diamond_1\Box_2 p$. We have to prove that $w \Vdash \Box_2\Diamond_1 p$. By the truth definition of $\Diamond_1$, w has an R_1-successor v satisfying $R_1 wv$ and $v \Vdash \Box_2 p$. Now we use the minimal valuation V_m again; first note that by the definition of V_m, we have $V_m(p) \subseteq V(p)$. Therefore, Lemma 3.37 ensures that it suffices to show that $\Box_2\Diamond_1 p$ holds at w *under the valuation* V_m. But this is immediate by the assumption that $\mathfrak{F}$ is confluent and (3.13). $\dashv$

This example inspires the following definitions.

Definition 3.45 Let τ be a modal similarity type. A *boxed atom* is a formula of the form $\Box_{i_1} \cdots \Box_{i_k} p$ ($k \geq 0$), where $\Box_{i_1}, \ldots, \Box_{i_k}$ are (not necessarily distinct) boxes of the language. In the case where $k = 0$, the boxed atom $\Box_{i_1} \cdots \Box_{i_k} p$ is just the proposition letter p. $\dashv$

Convention 3.46 In the sequel, it will be convenient to treat sequences of boxes as single boxes. We will therefore denote the formula $\Box_{i_1} \cdots \Box_{i_k} p$ by $\Box_\beta p$, where β is the sequence $i_1 \ldots i_k$ of indices. Analogously, we will pretend to have a corresponding binary relation symbol R_β in the frame language $\mathcal{L}^1_\tau$. Thus the expression $R_\beta xy$ abbreviates the formula

$$\exists y_1\,(R_{i_1} x y_1 \wedge \exists y_2\,(R_{i_2} y_1 y_2 \wedge \cdots \wedge \exists y_{k-1}\,(R_{i_{k-1}} y_{k-2} y_{k-1} \wedge R_{i_k} y_{k-1} y) \ldots)).$$

Note that this convention allows us to write the second-order translation of the boxed atom $\Box_\beta p$ as $\forall y\,(R_\beta xy \to Py)$.

If $k = 0$, β is the empty sequence ϵ; in this case the formula $R_\epsilon xy$ should be read as $x = y$. Note that the Second-Order Translation of $\Box_\epsilon p$ (that is, of the proposition letter p) can indeed be written as $\forall y\,(R_\epsilon xy \to Py)$.

Definition 3.47 Let τ be a modal similarity type. A *simple Sahlqvist antecedent* over this similarity type is a formula built up from $\top$, $\bot$ and boxed atoms, using only $\wedge$ and existential modal operators ($\Diamond$ and $\triangle$). A *simple Sahlqvist formula* is an implication $\phi \to \psi$ in which ψ is positive (as before) and ϕ is a simple Sahlqvist antecedent. $\dashv$

Example 3.48 Typical examples of simple Sahlqvist formulas are $\Diamond p \to \Diamond\Diamond p$, $\Box p \to \Box\Box p$, $\Box_1\Box_2 p \to \Box_3 p$, $\Diamond_1\Box_2 p \to \Box_2\Diamond_1 p$ and $(\Box_1\Box_2 p)\triangle(\Diamond_3 p \wedge \Box_2\Box_1 q) \to \Diamond_3(q\triangle p)$.

Typically *forbidden* in a simple Sahlqvist antecedent are:

(i) boxes over disjunctions, as in $H(r \vee Fq) \to G(Pr \wedge Pq)$,
(ii) boxes over diamonds, as in $\Box\Diamond p \to \Diamond\Box p$,
(iii) dual-triangled atoms, as in $p \,\triangledown\, p \to p$. $\dashv$

Theorem 3.49 *Let τ be a modal similarity type, and let $\chi = \phi \to \psi$ be a simple Sahlqvist formula over τ. Then χ locally corresponds to a first-order formula $c_\chi(x)$ on frames. Moreover, c_χ is effectively computable from χ.*

Proof. The proof of this theorem is an adaptation of the proof of Theorem 3.42. Consider the universally quantified second-order transcription of χ:

$$\forall P_1 \ldots \forall P_n \, (ST_x(\phi) \to ST_x(\psi)). \tag{3.14}$$

Again, we first make sure that no two quantifiers bind the same variable, and that no quantifier binds x. As before, the idea of the algorithm is to rewrite (3.14) to a formula from which we can easily read off instantiations which yield a first-order equivalent of (3.14).

Step 1. Pull out diamonds.
This is the same as before. This process results in a formula of the form

$$\forall P_1 \ldots \forall P_n \forall x_1 \ldots \forall x_m \, (\text{REL} \wedge \text{BOX-AT} \to ST_x(\psi)), \tag{3.15}$$

where REL is a conjunction of atomic first-order statements of the form Rx_ix_j corresponding to occurrences of diamonds, and BOX-AT is a conjunction of (translations of) boxed atoms, that is, formulas of the form $\forall y\, (R_\beta x_i y \to Py)$.

Step 2. Read off instances.
Let P be a unary predicate occurring in (3.15), and let $\pi_1(x_{i_1}), \ldots, \pi_k(x_{i_k})$ be all the (translations of the) boxed atoms in the antecedent of (3.10) in which the predicate P occurs. Observe that every π_j is of the form $\forall y\, (R_{\beta_j} x_{i_j} y \to Py)$, where β_j is a sequence of diamond indices (recall Convention 3.46). Define

$$\sigma(P) \equiv \lambda u.\, (R_{\beta_1} x_{i_1} u \vee \cdots \vee R_{\beta_k} x_{i_k} u).$$

Again, $\sigma(P_1), \ldots, \sigma(P_n)$ form the *minimal* instances making the antecedent $\text{REL} \wedge$ BOX-AT true.

The remainder of the proof is the same as the proof of Theorem 3.42, with the proviso that all occurrences of 'AT' should be replaced by 'BOX-AT'. $\dashv$

As in the case of very simple Sahlqvist formulas, the algorithm is best understood by inspecting some examples:

Example 3.50 Let us investigate some of the formulas given in Example 3.48. The simple Sahlqvist formula $\Box_1\Box_2 p \to \Box_3 p$ has the following second-order translation:

$$\forall P\,(\underbrace{\forall y\,(R_{12}xy \to Py)}_{\text{BOX-AT}} \to \forall z\,(R_3xz \to Pz)).$$

There are no diamonds to be pulled out here, so we can read off the required substitution instance $\sigma(P) \equiv \lambda u.\, R_{12}xu$ immediately. Carrying out the substitution we obtain

$$\forall y\,(R_{12}xy \to R_{12}xy) \to \forall z\,(R_3xz \to R_{12}xz),$$

which is equivalent to $\forall z\,(R_3xz \to R_{12}xz)$.

Next we consider the confluence formula $\Diamond_1\Box_2 p \to \Box_2\Diamond_1 p$, whose second-order translation is

$$\forall P(\exists x_1\,(R_1xx_1 \wedge \forall y\,(R_2x_1y \to Py)) \to \forall z_0\,(R_2xz_0 \to \exists z_1\,(R_1z_0z_1 \wedge Pz_1))).$$

Pulling out the existential quantification $\exists x_1$ yields

$$\forall P\forall x_1\,(\underbrace{R_1xx_1}_{\text{REL}} \wedge \underbrace{\forall y\,(R_2x_1y \to Py)}_{\text{BOX-AT}} \to \forall z_0\,(R_2xz_0 \to \exists z_1\,(R_1z_0z_1 \wedge Pz_1))).$$

The minimal instance making BOX-AT true is $\sigma(P) \equiv \lambda u.\, R_2x_1u$. After instantiating we obtain

$$\forall x_1\,(R_1xx_1 \wedge \forall y\,(R_2x_1y \to R_2x_1y) \to \forall z_0\,(R_2xz_0 \to \exists z_1\,(R_1z_0z_1 \wedge R_2x_1z_1))),$$

which can be simplified to

$$\forall x_1\forall z_0\,(R_1xx_1 \wedge R_2xz_0 \to \exists z_1\,(R_1z_0z_1 \wedge R_2x_1z_1)).$$

As our final example, let us treat a formula using a dyadic modality $\triangle$:

$$(\Box_1\Box_2 p)\triangle(\Diamond_3 p \wedge \Box_2\Box_1 q) \to \Diamond_3(q\triangle p).$$

We use a ternary relation symbol T for the triangle $\triangle$. Its second-order translation is the rather formidable looking

$$\begin{aligned}\forall P\forall Q\,(\exists x_1x_2\,(&Txx_1x_2 \wedge \forall y\,(R_{12}x_1y \to Py) \wedge \\ &\exists x_3\,(R_3x_2x_3 \wedge Px_3) \wedge \forall y\,(R_{21}x_2y \to Qy)) \\ &\to \exists z\,(R_3xz \wedge \exists z_1z_2\,(Tzz_1z_2 \wedge Qz_1 \wedge Pz_2))),\end{aligned}$$

from which we can pull out the diamonds $\exists x_1$, $\exists x_2$ and $\exists x_3$. This leads to

$$\forall P \forall Q \forall x_1 x_2 \forall x_3 \, (\overbrace{Txx_1x_2 \wedge R_3x_2x_3}^{\mathrm{REL}} \wedge \\ \overbrace{\forall y\, (R_{12}x_1y \to Py) \wedge Px_3 \wedge \forall y\, (R_{21}x_2y \to Qy)}^{\mathrm{BOX-AT}} \to \\ \exists z\, (R_3xz \wedge \exists z_1z_2\, (Tzz_1z_2 \wedge Qz_1 \wedge Pz_2))).$$

Now we can easily read off the required instantiations:

$$\begin{array}{rcl} \sigma(P) & \equiv & \lambda u.\, (R_{12}x_1u \vee u = x_3) \\ \sigma(Q) & \equiv & \lambda u.\, (R_{21}x_2u). \end{array}$$

Performing the substitution $[\sigma(P)/P, \sigma(Q)/Q]$ and deleting the tautological parts from the antecedent gives

$$\forall x_1x_2 \forall x_3 (Txx_1x_2 \wedge R_3x_2x_3 \to \\ \exists z\, (R_3xz \wedge \exists z_1z_2\, (Tzz_1z_2 \wedge R_{21}x_2z_1 \wedge (R_{12}x_1z_2 \vee z_2 = x_3)))). \quad \dashv$$

Sahlqvist formulas

We are now ready to introduce the full Sahlqvist fragment and the full version of the Sahlqvist-van Benthem algorithm.

Definition 3.51 Let τ be a modal similarity type. A *Sahlqvist antecedent* over τ is a formula built up from $\top$, $\bot$, boxed atoms, and negative formulas, using $\wedge$, $\vee$ and existential modal operators ($\Diamond$ and $\triangle$). A *Sahlqvist implication* is an implication $\phi \to \psi$ in which ψ is positive and ϕ is a Sahlqvist antecedent.

A *Sahlqvist* formula is a formula that is built up from Sahlqvist implications by freely applying boxes and conjunctions, and by applying disjunctions only between formulas that do not share any proposition letters. $\dashv$

Example 3.52 Both simple and very simple Sahlqvist formulas are examples of Sahlqvist formulas, as are $\Box(p \to \Diamond p)$, $p \wedge \Diamond\neg p \to \Diamond p$, and $\Box(\Diamond_1\Box_2 p \to \Box_2\Diamond_1 p) \wedge \Box_1(p \to \Diamond_2 p)$. As with simple Sahlqvist formulas, typically forbidden combinations in Sahlqvist antecedent are 'boxes over disjunctions,' 'boxes over diamonds,' and 'dual-triangled atoms' as in $p \bigtriangledown p \to p$ (see Example 3.48). $\dashv$

The following lemma is instrumental in reducing the correspondence problem for arbitrary Sahlqvist formulas, first to that of Sahlqvist implications, and then to to that of simple Sahlqvist formulas.

Lemma 3.53 *Let τ be a modal similarity type, and let ϕ and ψ be τ-formulas.*

(i) *If ϕ and $\alpha(x)$ are local correspondents, then so are $\Box_\beta \phi$ and $\forall y\,(R_\beta xy \to [y/x]\alpha)$.*
(ii) *If ϕ (locally) corresponds to α, and ψ (locally) corresponds to β, then $\phi \wedge \psi$ (locally) corresponds to $\alpha \wedge \beta$.*
(iii) *If ϕ locally corresponds to α, ψ locally corresponds to β, and ϕ and ψ have no proposition letters in common, then $\phi \vee \psi$ locally corresponds to $\alpha \vee \beta$.*

Proof. Left as Exercise 3.6.3. $\dashv$

The local perspective in part one and three of the Lemma is essential. For instance, one can find a modal formula ϕ that globally corresponds to a first-order condition $\forall x\, \alpha(x)$ without $\Box\phi$ globally corresponding to the formula $\forall x \forall y\,(Rxy \to \alpha(y))$; see Exercise 3.6.3.

Theorem 3.54 *Let τ be a modal similarity type, and let χ be a Sahlqvist formula over τ. Then χ locally corresponds to a first-order formula $c_\chi(x)$ on frames. Moreover, c_χ is effectively computable from χ.*

Proof. The proof of the theorem is virtually the same as the proof of Theorem 3.49, with the exception of the use of Lemma 3.53 and of the fact that we have to do some pre-processing of the formula χ.

By Lemma 3.53 it suffices to show that the theorem holds for all Sahlqvist implications. So assume that χ has the form $\phi \to \psi$ where ϕ is a Sahlqvist antecedent and ψ a positive formula. Proceed as follows.

Step 1. Pull out diamonds and pre-process.
Using the same strategy as in the proof of Theorem 3.49 together with equivalences of the form

$$((\alpha \vee \beta) \to \gamma) \;\leftrightarrow\; (\,(\alpha \to \gamma) \wedge (\beta \to \gamma)\,)$$

and

$$\forall \ldots (\alpha \wedge \beta) \;\leftrightarrow\; (\forall \ldots \alpha \wedge \forall \ldots \beta\,),$$

we can rewrite the second-order translation of $\phi \to \psi$ into a conjunction of formulas of the form

$$\forall P_1 \ldots \forall P_n \forall x_1 \ldots \forall x_m\,(\text{REL} \wedge \text{BOX-AT} \wedge \text{NEG} \to ST_x(\psi)), \qquad (3.16)$$

where REL is a conjunction of atomic first-order statements of the form $R_\triangle \vec{x}$ corresponding to occurrences of diamonds and triangles, BOX-AT is a conjunction of (translations of) boxed atoms, and NEG is a conjunction of (translations of) negative formulas. By Lemma 3.53(ii) it suffices to show that each formula of the form displayed in (3.16) has a first-order equivalent. This is done by using the equivalence

$$(\alpha \wedge \text{NEG} \to \beta) \;\leftrightarrow\; (\alpha \to \beta \vee \neg\text{NEG}),$$

where $\neg$NEG is the *positive* formula that arises by negating the negative formula NEG. Using this equivalence we can rewrite (3.16) to obtain a formula of the form

$$\forall P_1 \ldots \forall P_n \forall x_1 \ldots \forall x_m \, (\text{REL} \wedge \text{BOX-AT} \to \text{POS}),$$

and from here on we can proceed as in Step 2 of the proof of Theorem 3.49. $\dashv$

Example 3.55 By way of example we determine the local first-order correspondents of two of the modal formulas given in Example 3.52. To determine the first-order correspondent of the Sahlqvist formula $\Box(p \to \Diamond p)$ we first recall that the local first-order correspondent of $p \to \Diamond p$ is Rxx. So, by Lemma 3.53(i) $\Box(p \to \Diamond p)$ locally corresponds to $\forall y \, (Rxy \to Ryy)$.

Next we consider the Sahlqvist formula $(p \wedge \Diamond\neg p) \to \Diamond p$. Its translation is

$$\forall P \, (Px \wedge \exists y \, (Rxy \wedge \neg Py) \to \exists z \, (Rxz \wedge Pz)).$$

Pulling out the diamond produces

$$\forall P \forall y \, (\underbrace{Px}_{\text{BOX-AT}} \wedge \underbrace{Rxy}_{\text{REL}} \wedge \underbrace{\neg Py}_{\text{NEG}} \to \underbrace{\exists z \, (Rxz \wedge Pz)}),$$

and moving the negative part $\neg Py$ to the consequent we get

$$\forall P \forall y \, (\underbrace{Px}_{\text{BOX-AT}} \wedge \underbrace{Rxy}_{\text{REL}} \to \underbrace{Py \vee \exists z \, (Rxz \wedge Pz)}_{\text{POS}}).$$

The minimal instantiation to make Px true is $\lambda u.\, u = x$. After instantiation we obtain

$$\forall y \, (Rxy \to y = x \vee \exists z \, (Rxz \wedge z = x)),$$

which can be simplified to $\forall y \, (Rxy \wedge x \neq y \to Rxx)$. $\dashv$

Exercises for Section 3.6

3.6.1 Compute the first-order formulas locally corresponding to the following Sahlqvist formulas:

(a) $\Diamond_1 \Diamond_2 p \to \Diamond_2 \Diamond_1 p$,
(b) $(p \wedge \Box p \wedge \Box\Box p) \to \Diamond p$,
(c) $\Diamond^k \Box^l p \to \Box^m \Diamond^n p$, for arbitrary natural numbers k, l, m and n,
(d) $(\Box p) \triangle (\Box p) \to p \nabla p$,
(e) $\Diamond(\neg p \wedge \Diamond(p \wedge q)) \to \Diamond(p \wedge q)$,
(f) $\Box((p \wedge \Box\neg p \wedge q) \to \Diamond q))$.

3.6.2 (a) Show that the formula $\Box(p \vee q) \to \Diamond(\Box p \vee \Box q)$ does not locally correspond to a first-order formula on frames. (Hint: modify the frame of Example 3.11.)
(b) Use this example to show that dual-*triangled* atoms cannot be allowed in Sahlqvist antecedents.

3.6.3 Prove Lemma 3.53:

(a) Show that if ϕ and $\alpha(x)$ locally correspond, so do $\Box_{\beta}\phi$ and $\forall y\,(R_{\beta}xy \to \alpha(y))$.
(b) Prove that if ϕ (locally) corresponds to $\alpha(x)$, and ψ (locally) corresponds to $\beta(x)$, then $\phi \wedge \psi$ (locally) corresponds to $\alpha(x) \wedge \beta(x)$.
(c) Show that if ϕ locally corresponds to α, ψ locally corresponds to $\beta(x)$, and ϕ and ψ have no proposition letters in common, then $\phi \vee \psi$ locally corresponds to $\alpha(x) \vee \beta(x)$.
(d) Prove that (a) and (c) do not hold for global correspondence, and that the condition on the proposition letters in (c) is necessary as well. (Hint: for (a), think of the modal formula $\Diamond\Diamond p \to \Diamond p$ and the first-order formula $\forall xyz\,(Ryz \wedge Rzx \to Ryx)$.)

3.7 More about Sahlqvist Formulas

It is time to step back and think more systematically about the Sahlqvist fragment, for a number of questions need addressing. For a start, does this fragment contain *all* modal formulas with first-order correspondents? And why did we forbid disjunctions in the scope of boxes, and occurrences of nested duals of triangles in Sahlqvist antecedents, while we allowed boxed atoms? Most interesting of all, *which* first-order conditions are expressible by means of Sahlqvist formulas? That is, is it possible to prove some sort of converse to the Sahlqvist Correspondence Theorem?

Limitative results

To set the stage for our discussion, we first state (without proof) the principal limitative result in this area: Chagrova's Theorem. Good presentations of the proof are available in the literature; see the Notes for references.

Theorem 3.56 (Chagrova's Theorem) *It is undecidable whether an arbitrary basic modal formula has a first-order correspondent.*

This implies that, even for the basic modal language, it is *not* possible to write a computer program which when presented with an arbitrary modal formula as input, will terminate after finitely many steps, returning the required first-order correspondent (if there is one) or saying 'No!' (if there isn't).

Quite apart from its intrinsic interest, this result immediately tells us that the Sahlqvist fragment cannot possibly contain *all* modal formulas with first-order correspondents. For it is straightforward to decide whether a modal formula is a Sahlqvist formula, and to compute the first-order correspondents of Sahlqvist formulas. Hence if all modal formulas with first-order correspondents were Sahlqvist, this would contradict Chagrova's Theorem.

But a further question immediately presents itself: is every modal formula with a first-order correspondent *equivalent* to a Sahlqvist formula? (The preceding argument does not rule this out.) The answer is *no*: there are modal formulas corresponding to first-order frame conditions which are not equivalent to any Sahlqvist formula.

Example 3.57 Consider the conjunction of the following two formulas:

(M) $\Box\Diamond p \to \Diamond\Box p$
(4) $\Diamond\Diamond q \to \Diamond q$.

(M) is the McKinsey formula we discussed in Example 3.11, and (4) is the transitivity axiom. It is obvious that M itself is not a Sahlqvist axiom, and by Example 3.11 it does not express a first-order condition.

It requires a little argument to show that the *conjunction* $\mathrm{M} \wedge 4$ is not *equivalent* to a Sahlqvist formula. One way to do so is by proving that $\mathrm{M} \wedge 4$ does not have a *local* first-order correspondent (cf. Exercise 3.7.1).

Nevertheless, the conjunction $\mathrm{M} \wedge 4$ does have a first-order correspondent, as we can prove the following equivalence for all transitive frames $\mathfrak{F}$:

$$\mathfrak{F} \Vdash \mathrm{M} \text{ iff } \mathfrak{F} \models \forall x \exists y\, (Rxy \wedge \forall z\, (Ryz \to z = y)). \tag{3.17}$$

We leave the right to left direction as an exercise to the reader. To prove the other direction, we reason by contraposition. That is, we assume that there is a transitive frame $\mathfrak{F} = (W, R)$ on which the McKinsey formula is valid, but which does *not* satisfy the first-order formula given in (3.17). Let r be a state witnessing that the first-order formula in (3.17) does not hold in $\mathfrak{F}$. That is, assume that each successor s of r has a successor distinct from it. We may assume that the frame is generated from r, so that $\mathfrak{F} \models \forall y \exists z\, (Ryz \wedge y \neq z)$.

In order to derive a contradiction from this, we need to introduce some terminology. Call a subset X of W *cofinal in* W if for all $w \in W$ there is an $x \in X$ such that Rwx. We now claim that

$$W \text{ has a subset } X \text{ such that both } X \text{ and } W \setminus X \text{ are cofinal in } W. \tag{3.18}$$

From (3.18) we can immediately derive a contradiction by considering the valuation V given by $V(p) = X$. For, cofinality of X implies that $(\mathfrak{F}, V), r \Vdash \Box\Diamond p$, while cofinality of $W \setminus X$ likewise gives $(\mathfrak{F}, V), r \Vdash \Box\Diamond\neg p$. But then $(\mathfrak{F}, V), r \nVdash \mathrm{M}$.

To prove (3.18), consider the collection C of all pairs of disjoint subsets Y, $Z \subset W$ satisfying $\forall y \in Y \exists z \in Z\, Ryz$ and $\forall z \in Z \exists y \in Y\, Rzy$. This set is nonempty because $\mathfrak{F} \models \forall y \exists z\, (Ryz \wedge y \neq z)$; order it under coordinate-wise inclusion. It is obvious that every chain in this partial ordering is bounded above; hence, we may apply Zorn's Lemma and obtain a *maximal* such pair Y, Z. We claim that

$$Y \cup Z = W \tag{3.19}$$

Since Y and Z are disjoint, this implies that $Z = W \setminus Y$ and thus proves (3.18).

Suppose that (3.19) does *not* hold. Then there is an element $w \in W$ which belongs neither to Y nor to Z. If there were some $z \in Z$ with Rwz then the pair $(Y \cup \{w\}, Z)$ would belong to C, contradicting the maximality of (Y, Z). Likewise, there is no $y \in Y$ with Rwy. Even so, we will define non-empty sets Y', Z' such that $(Y \cup Y', Z \cup Z') \in C$, again contradicting the maximality of (Y, Z). First put w in Y'. Now choose an element z_1 of W such that Rwz_1 and $w \neq z_1$ and put z_1 in Z' — remember that $z_1 \notin Y \cup Z$. Then choose an element y_1 of W such that Rz_1y_1 and $z_1 \neq y_1$ and put y_1 into Y'. Continue this process and observe that none of the y_n, z_n will belong to $Y \cup Z$; this is by transitivity of R and our assumption on w.

The process will finish if, for instance, some u has just been put in Z', but all of its successors have already been put in $Y' \cup Z'$ at some earlier state. In such a case we break off the process; at this moment it is obvious that each $y \in Y \cup Y'$ has a successor in $Z \cup Z'$, and that each $z \in Z \cup Z'$ distinct from u has a successor in $Y \cup Y'$. To show that u itself has a successor in Y', let v be the *first* element in the sequence $wRz_1Ry_1Rz_2 \ldots$ such that Ruv. If v itself does not belong to Y', it must belong to Z'; but since we did not break off the process at this stage, this means that we could put a successor y_i of v in Y'; by transitivity, Ruy_i. The only other case in which the process may finish is symmetric to the case described.

Finally, if the process does *not* finish in this way we are dealing with an infinite sequence $wRz_1Ry_1Rz_2 \ldots$. But then the pair $(Y \cup Y', Z \cup Z')$ belongs to C. ⊣

Obviously, the example begs the question whether there is a modal formula that *locally* corresponds to a first-order formula without being equivalent to a Sahlqvist formula. The answer to this question is affirmative: the formula $\Box M \wedge 4$ is a counterexample. In Exercise 3.7.1 the reader is asked to show that it has a local first-order correspondent; in Chapter 5 we will develop the techniques needed to prove that the formula is not equivalent to a Sahlqvist formula, see Exercise 5.6.2.

Thus the Sahlqvist fragment does not contain all modal formulas with first-order correspondents. So the next question is: can the Sahlqvist fragment be further extended? The answer is *yes* — but we should reflect a little on what we hope to achieve through such extensions. The Sahlqvist fragment is essentially a good compromise between the demands of generality and simplicity. By adding further restrictions it is possible to extend it further, but it is not obvious that the resulting loss of simplicity is really worth it. Moreover, the Sahlqvist fragment also gives rise to a matching completeness theorem; we would like proposed extensions to do so as well. We don't know of simple generalizations of the Sahlqvist fragment which manage to do this. In short, while there is certainly room for experiment here, it is unclear whether anything interesting is likely to emerge.

However, one point is worth stressing once more: the Sahlqvist fragment *cannot* be further extended simply by dropping some of the restrictions in the definition of a Sahlqvist formula. We forbid disjunctions in the scope of boxes and nested duals of triangles in Sahlqvist antecedents for a very good reason: these forbidden combinations easily lead to modal formulas that have no first-order correspondent, as we have seen in Example 3.11 and Exercise 3.6.2.

Kracht's theorem

Let's turn to a nice positive result. As has already been mentioned, not only does each Sahlqvist formula define a first-order class of frames, but when we use one as an axiom in a normal modal logic, that logic is guaranteed to be complete with respect to the elementary class of frames the axiom defines. (This is the content of the Sahlqvist Completeness Theorem; see Theorem 4.42 for a precise statement.) So it would be very pleasant to know *which* first-order conditions are the correspondents of Sahlqvist formulas. Kracht's Theorem is a sort of converse to the Sahlqvist Correspondence Theorem which gives us this information.

Before we can define the fragment of first-order logic corresponding to Sahlqvist formulas we need some auxiliary definitions; we also introduce some helpful notation. For reasons of notational simplicity, we work in the basic modal similarity type. First of all, we will abbreviate the first-order formula $\forall y\,(Rxy \to \alpha(y))$ to $(\forall y {\rhd} x)\alpha(y)$, speaking of *restricted quantification* and calling x the *restrictor* of y. Likewise $\exists y\,(Rxy \wedge \alpha(y))$ is abbreviated to $(\exists y {\rhd} x)\alpha(y)$. We will call the constructs $(\forall y {\rhd} x)$ and $(\exists y {\rhd} x)$ *restricted quantifiers*. If we wish not to specify the restrictor of a restricted quantifier we will write $\forall^r y$ or $\exists^r y$. Moreover, if we don't wish to specify whether a quantifier is existential or universal we denote it by Q (Q^r in the restricted case). Second, for the duration of this subsection it will be convenient for us to consider formulas of the form $u \neq u$ as atomic. Third, in this subsection we will work exclusively with formulas in which no variable occurs both free and bound, and in which no two distinct (occurrences of) quantifiers bind the same variable; we will call such formulas *clean*.

Now we call a formula *restrictedly positive* if it is built up from atomic formulas, using $\wedge$, $\vee$ and restricted quantifiers only; observe that monadic predicates occur positively in restrictedly positive formulas. Finally, we assume that the reader knows how to rewrite an arbitrary positive propositional formula to a *disjunctive normal form* or DNF (that is, to an equivalent disjunction of conjunctions of atomic formulas) and to a *conjunctive normal form* or CNF (that is, to an equivalent conjunction of disjunctions of atomic formulas).

The crucial notion in this subsection is that of a variable occurring *inherently universally* in a first-order formula.

Definition 3.58 We say that an occurrence of the variable y in the (clean!) formula α is inherently universal if either y is free, or else y is bound by a restricted quantifier of the form $(\forall y \triangleright x)\beta$ which is not in the scope of an existential quantifier. A formula $\alpha(x)$ in the basic first-order frame language is called a *Kracht formula* if α is clean, restrictedly positive and furthermore, every atomic formula is either of the form $u = u$ or $u \neq u$, or else it contains at least one inherently universal variable. $\dashv$

Restricted quantification is obviously the modal face of quantification in first-order logic; indeed, we could have defined the standard translation of a modal formula using this notion. As for Kracht formulas, first observe that every *universal* restricted first-order formula satisfies the definition. A second example of a Kracht formula is $(\forall w \triangleright v)(\forall x \triangleright v)(\exists y \triangleright w)Rxy$: note that it does not matter that the 'x' in Rxy falls within the scope of an existential quantifier; what matters is that the universal *quantifier* that *binds* x does not occur within the scope of any existential quantification. On the other hand, the formula $(\exists w \triangleright v)(\forall x \triangleright v)w = x$ is not a Kracht formula since the occurrence of neither w nor x in $w = x$ is inherently universal: w is disqualified because it is bound by an existential quantifier and x because it is bound within the scope of the existential quantifier $(\exists w \triangleright v)$.

The following result states that Kracht formulas are the first-order counterparts of Sahlqvist formulas — but not only that. As will become apparent from its proof, from a given Kracht formula we can *compute* a Sahlqvist formula locally corresponding to it. The reader is advised to glance at the examples provided below while reading the proof.

Theorem 3.59 *Any Sahlqvist formula locally corresponds to a Kracht formula; and conversely, every Kracht formula is a local first-order correspondent of some Sahlqvist formula which can be effectively obtained from the Kracht formula.*

Proof. For the left to right direction, we leave it as an exercise to the reader to show that the algorithm discussed in the sections 3.5 and 3.6 in fact produces, given a Sahlqvist formula, a first-order correspondent *within* the Kracht fragment. We'll give the proof of the other direction: we'll show how rewrite a given Kracht formula to an equivalent Sahlqvist formula.

Our first step is to provide special prenex formulas as normal forms for Kracht formulas. Define a *type 1* formula to be of the form

$$\forall^r x_1 \dots \forall^r x_n \mathsf{Q}^r_1 y_1 \dots \mathsf{Q}^r_m y_m \, \beta(x_0, \dots, x_n, y_1, \dots, y_m)$$

such that $n, m \geq 0$ and each variable is restricted by an earlier variable (that is, the restrictor of any x_i is some x_j with $j < i$ and the restrictor of any y_i is either some x_k or some y_j with $j < i$. Furthermore we require that β is a DNF of formulas $u = u$, $u \neq u$, Rux, $u = x$ and Rxu (that is, we allow all atomic formulas that are *not* of the form Ryy' or $y = y'$). Here and in the remainder of this proof we use the convention that u and z denote arbitrary variables in $\{x_0, \ldots, x_n, y_1, \ldots, y_m\}$ and x an arbitrary variable in $\{x_0, \ldots, x_n\}$.

Clearly then, type 1 formulas form a special class of Kracht formulas. This inclusion is not proper (modulo equivalence), since we can prove the following claim.

Claim 1 *Every Kracht formula can be effectively rewritten into an equivalent type 1 formula.*

Proof of Claim. Let $\alpha(x_0)$ be a Kracht formula. By definition it is built up from atomic formulas using $\wedge$, $\vee$ and restricted quantifiers. Furthermore, since $\alpha(x_0)$ is clean, in a subformula of the form $\mathsf{Q}^r v\, \beta$ the variable v may not occur outside of β. Hence, we may use the equivalences

$$(\mathsf{Q}^r v\, \beta) \heartsuit \gamma \;\leftrightarrow\; \mathsf{Q}^r v\, (\beta \heartsuit \gamma) \tag{3.20}$$

(where $\heartsuit$ uniformly denotes either $\wedge$ or $\vee$) to pull out quantifiers to the front. However, if we want to remain within the Kracht fragment we have to take care about the *order* in which we pull out quantifiers.

Without loss of generality we may assume that each inherently universal variable is named x_i for some i, while each of the remaining variables is named y_j for some j. This ensures that no atomic subformula of $\alpha(x_0)$ is of the form Ryy' or $y = y'$ (with distinct variables y and y').

Observe also that in every subformula of the form $((\forall x \triangleright u)\beta)\heartsuit\gamma$, the variable u occurs free. If this u is not the variable x_0 then it is a bound variable of α; hence, the mentioned subformula must occur in the scope of a quantifier $(\mathsf{Q}^r u \triangleright x')$. This quantification must have been universal, for otherwise, the variable x could not have been among the inherently universal ones. But this means that the variable u itself must be inherently universal as well, so u is some x_i. This shows that by successively pulling out restricted universal quantifiers $\forall^r x$ we end up with a Kracht formula of the form

$$\forall^r x_1 \ldots \forall^r x_n\, \alpha'(x_0, \ldots, x_n, y_1, \ldots, y_m),$$

such that each atomic formula of α' is of the form $u = u$ or $u \neq u$, or else it contains some occurrence of a variable x_i. Furthermore, the restrictor of each x_i is some x_j with $j < i$.

It remains to pull out the other restricted quantifiers from α'. But this can easily be done using the equivalences of (3.20), since we do not have to worry anymore about the order in which we pull out the quantifiers. In the end, we arrive at a formula of the form

$$\forall^r x_1 \dots \forall^r x_n \, \mathsf{Q}^r_1 y_1 \dots \mathsf{Q}^r_m y_m \, \alpha''(x_0, \dots, x_n, y_1, \dots, y_m)$$

such that the atomic subformulas of α'' satisfy the same condition of those in α' (in fact, they are the very same formulas), while in addition, α'' is quantifier free. Hence, if we rewrite α'' into disjunctive normal form, we are finished. $\dashv$

Enter diamonds and boxes. A *type 2* formula is a formula in the second-order frame language of the form

$$\tilde{\forall} P_0 \dots \tilde{\forall} P_n \tilde{\forall} Q_0 \dots \tilde{\forall} Q_n \forall^r x_1 \dots \forall^r x_n \left(\bigwedge_{0 \leq i \leq n} ST_{x_i}(\sigma_i) \to \beta \right)$$

such that each σ_i is a conjunction of boxed atoms in p_i and q_i, whereas β is a DNF of formulas $ST_x(\psi)$, with ψ some modal formula which is positive in each p_i, q_j.

Claim 2 *Every type 1 formula can be can be effectively rewritten into an equivalent type 2 formula.*

Proof of Claim. Now the prominent role of the inherently universal formulas will come out: they determine the propositional variables of the Sahlqvist formula and the 'BOX-AT' part of its antecedent. Consider the type 1 formula

$$\forall^r x_1 \dots \forall^r x_n \mathsf{Q}^r_1 y_1 \dots \mathsf{Q}^r_m y_m \, \beta(x_0, \dots, x_n, y_1, \dots, y_m).$$

We abbreviate the sequence $\forall^r x_1 \dots \forall^r x_n$ by $\forall^r \bar{x}$, and use similar abbreviations for other sequences of quantifiers. Recall that β is a DNF of formulas $u = u$, $u \neq u$, $u = x_i$, Rux_i and Rx_iu. Our first move is to replace such subformulas with the formulas $ST_u(\top)$, $ST_u(\bot)$, $ST_u(p_i)$, $ST_u(\Diamond p_i)$ and $ST_u(q_i)$, respectively; call the resulting formula β'.

Our first claim is that

$$\forall^r \bar{x} \, \mathsf{Q}^r \bar{y} \, \beta \quad \text{is equivalent to}$$
$$\tilde{\forall} \bar{P} \bar{Q} \, \forall^r \bar{x} \left(\bigwedge_{0 \leq i \leq n} ST_{x_i}(p_i \wedge \Box q_i) \to \mathsf{Q}^r \bar{y} \, \beta' \right). \tag{3.21}$$

Forbidding as (3.21) may look, its proof is completely analogous to proofs in Sections 3.5 and 3.6: the direction from right to left is immediate by instantiation, while the other direction simply follows from the fact that β' is monotone in each predicate symbol P_i and Q_i.

Two remarks are in order here. First, since β may contain atomic formulas of the form Rx_ix_j and $x_i = x_j$ (that is, with *both* variables being inherently universal), there is some *choice* here. For instance, the formula Rx_ix_j may be replaced with either $ST_{x_i}(\Diamond p_j)$ or with $ST_{x_j}(q_j)$. Having this choice can sometimes be of use if one wants to find Sahlqvist correspondents satisfying some additional constraints.

Related to this is our second remark: we don't need to introduce *both* propositional variables p_i and q_i for *each* x_i. We can do with any supply of variables that is sufficient to replace all atomic formulas of β with the standard translation of either $ST_u(p_i)$, $ST_u(\Diamond p_i)$ or $ST_u(q_i)$. A glance at the examples below will make this point clear.

We are now halfway through the proof of Claim 2: observe that β' is already a DNF of formulas $ST_u(\psi)$ with ψ positive in each p_i, q_j. It remains to eliminate the quantifier sequence $\mathsf{Q}^r\bar{y}$. This will be done step by step, using the following procedure.

Consider the formula

$$(\exists y_{i+1} \rhd z) \left(\bigvee_{k \leq K} \bigwedge_{l \leq L_k} ST_{u_{kl}}(\psi_{kl}) \right), \tag{3.22}$$

where each modal formula ψ_{kl} is positive in all variables p_i, q_j; z is either an x or a y_j with $j \leq i$; and each u is either an x or a y_j with $j \leq i+1$. We first distribute the existential quantifier over the disjunction, yielding a disjunction of formulas

$$(\exists y_{i+1} \rhd z) \bigwedge_{l \leq L_k} ST_{u_{kl}}(\psi_{kl}). \tag{3.23}$$

We may assume all these variables u to be distinct (otherwise, replace $ST_u(\psi') \wedge ST_u(\psi'')$ with $ST_u(\psi' \wedge \psi'')$); we may also assume that y_{i+1} is the variable u_{lL_k} (if y_{i+1} does not occur among the u's, add a conjunct $ST_{y_{i+1}}(\top)$). But then (3.23) is equivalent to the formula

$$ST_z(\Diamond \psi_{kL}) \ \wedge \ \bigwedge_{l < L_k} ST_{u_{kl}}(\psi_{kl}),$$

whence (3.22) is equivalent to a disjunction of such formulas. Observe further that y_{i+1} does not occur in these formulas.

This shows how to get rid of an existential innermost restricted quantifier of the prenex $K^r\bar{y}$. A universal innermost restricted quantifier can be removed dually, by first converting the matrix β' into a *conjunctive* normal form; details are left to the reader. In any case, it will be clear that by this procedure we can rewrite any type 1 formula into an equivalent type 2 formula. $\dashv$

We are now almost through with the proof of Theorem 3.59. All we have to do now is show how to massage arbitrary type 2 formulas into Sahlqvist shape.

Claim 3 *Any type 2 formula can be can be effectively rewritten into an equivalent Sahlqvist formula.*

Proof of Claim. Let

$$\tilde{\forall}\bar{P}\bar{Q}\,\forall^r\bar{x}\ \left(\bigwedge_{0\leq i\leq n} ST_{x_i}(\sigma_i)\ \rightarrow\ \beta\right) \tag{3.24}$$

be an arbitrary type 2 formula.

First we rewrite β into conjunctive normal form, and we distribute the implication and the prenex of universal quantifiers over the conjunctions. Thus we obtain a conjunction of formulas of the form

$$\tilde{\forall}\bar{P}\bar{Q}\forall^r\bar{x}\ \left(\bigwedge_{0\leq i\leq n} ST_{x_i}(\sigma_i)\ \rightarrow\ \beta'\right), \tag{3.25}$$

where β' is a disjunction of formulas of the form $ST_x(\psi)$ with each ψ positive in all p_i and q_j. As before, we may assume that each x_i occurs in exactly one disjunct of β', so (3.25) is equivalent to a formula

$$\tilde{\forall}\bar{P}\bar{Q}\,\forall^r\bar{x}\ \left(\bigwedge_{0\leq i\leq n} ST_{x_i}(\sigma_i) \rightarrow \bigvee_{0\leq i\leq n} ST_{x_i}(\psi_i)\right),$$

where each σ_i is a Sahlqvist antecedent and each ψ_i is positive. But clearly then, (3.25) is equivalent to the formula

$$\tilde{\forall}\bar{P}\bar{Q}\neg\exists^r\bar{x}\ \bigwedge_{0\leq i\leq n} ST_{x_i}(\sigma_i \wedge \neg\psi_i).$$

Observe that each modal formula $\sigma_i \wedge \neg\psi_i$ is a Sahlqvist antecedent.

But now, as before, working inside out we may eliminate all remaining restricted quantifiers, step by step. For, observe that the formula

$$\exists^r x_1 \ldots \exists^r x_{k-1}(\exists x_k \rhd x_j)\ \bigwedge_{0\leq i\leq k} ST_{x_i}(\chi_i)$$

is equivalent to

$$\exists^r x_1 \ldots \exists^r x_{k-1}\ \left(ST_{x_j}(\chi_j \wedge \Diamond\chi_{k+1})\ \wedge\ \bigwedge_{0\leq i<k, i\neq j} ST_{x_i}(\chi_i)\right).$$

Note that $\chi_j \wedge \Diamond\chi_{k+1}$ is a Sahlqvist antecedent if χ_j and χ_{k+1} are.

It turns out that for some Sahlqvist antecedent ϕ, (3.25) is equivalent to the second-order formula

$$\tilde{\forall}\bar{P}\bar{Q}\,\neg ST_{x_0}(\phi).$$

But then (3.24) is equivalent to a conjunction of such formulas, and thus equivalent to a formula

$$\tilde{\forall}\bar{P}\bar{Q}\, ST_{x_0}\left(\bigvee_l \phi_l \to \bot\right),$$

which is the local second-order frame correspondent of the formula $\bigvee_l \phi_l \to \bot$, which is obviously in Sahlqvist form. $\dashv$

This completes the proof of the third claim, and hence of the theorem. $\dashv$

Example 3.60 Consider the formula

$$\alpha(x_0) \;\equiv\; (\forall x_1 \rhd x_0)(\exists y_1 \rhd x_0)(\exists y_2 \rhd y_1)\, Rx_1y_2.$$

This is already a type 2 Kracht formula, so we proceed by the procedure described in the proof of Claim 2 in the proof of Theorem 3.59. According to (3.21), $\alpha(x_0)$ is equivalent to the second order formula

$$\tilde{\forall}Q_1(\forall x_1 \rhd x_0)\,(ST_{x_1}(\Box q_1) \to (\exists y_1 \rhd x_0)(\exists y_2 \rhd y_1)ST_{y_2}(q_1)).$$

Then, using the equivalences described further on in the proof of Claim 2 we obtain the following sequences of formulas that are equivalent to $\alpha(x_0)$:

$$\begin{aligned}
&\tilde{\forall}Q_1(\forall x_1 \rhd x_0)\,(ST_{x_1}(\Box q_1) \to (\exists y_1 \rhd x_0)(\exists y_2 \rhd y_1)ST_{y_2}(q_1))\\
&\quad\Leftrightarrow\quad \tilde{\forall}Q_1(\forall x_1 \rhd x_0)\,(ST_{x_1}(\Box q_1) \to (\exists y_1 \rhd x_0)ST_{y_1}(\Diamond q_1)),\\
&\quad\Leftrightarrow\quad \tilde{\forall}Q_1(\forall x_1 \rhd x_0)\,(ST_{x_1}(\Box q_1) \to ST_{x_0}(\Diamond\Diamond q_1)).
\end{aligned}$$

The last formula is a type 2 formula. Hence, the only thing left to do is to rewrite it to an equivalent Sahlqvist formula; this we do via the sequence of equivalent formulas below, following the pattern of the proof of Claim 3.

$$\begin{aligned}
&\tilde{\forall}Q_1(\,(\forall x_1 \rhd x_0)\,(ST_{x_1}(\Box q_1) \to ST_{x_0}(\Diamond\Diamond q_1))\,)\\
&\quad\Leftrightarrow\quad \tilde{\forall}Q_1(\,(\forall x_1 \rhd x_0)\,\neg(ST_{x_1}(\Box q_1) \wedge \neg ST_{x_0}(\Diamond\Diamond q_1))\,)\\
&\quad\Leftrightarrow\quad \tilde{\forall}Q_1(\,(\forall x_1 \rhd x_0)\,\neg(ST_{x_1}(\Box q_1) \wedge ST_{x_0}(\neg\Diamond\Diamond q_1))\,)\\
&\quad\Leftrightarrow\quad \tilde{\forall}Q_1(\,\neg(\exists x_1 \rhd x_0)\,(ST_{x_1}(\Box q_1) \wedge ST_{x_0}(\neg\Diamond\Diamond q_1))\,)\\
&\quad\Leftrightarrow\quad \tilde{\forall}Q_1(\,\neg((\exists x_1 \rhd x_0)ST_{x_1}(\Box q_1) \wedge ST_{x_0}(\neg\Diamond\Diamond q_1))\,)\\
&\quad\Leftrightarrow\quad \tilde{\forall}Q_1(\,\neg(ST_{x_0}(\Diamond\Box q_1) \wedge ST_{x_0}(\neg\Diamond\Diamond q_1))\,)\\
&\quad\Leftrightarrow\quad \tilde{\forall}Q_1(\,\neg ST_{x_0}(\Diamond\Box q_1 \wedge \neg\Diamond\Diamond q_1)\,)\\
&\quad\Leftrightarrow\quad \tilde{\forall}Q_1(\,ST_{x_0}((\Diamond\Box q_1 \wedge \neg\Diamond\Diamond q_1) \to \bot)\,).
\end{aligned}$$

This means that $\alpha(x_0)$ locally corresponds to the Sahlqvist formula $(\Diamond\Box q_1 \wedge \neg\Diamond\Diamond q_1) \to \bot$, or to the equivalent formula $\Diamond\Box q_1 \to \Diamond\Diamond q_1$. $\dashv$

Example 3.61 Consider the Kracht formula

$$\alpha(x_0) \;\equiv\; (\forall x_1 \rhd x_0)(\forall x_2 \rhd x_0)\,(Rx_1x_2 \vee Rx_2x_1 \vee x_1 = x_2).$$

According to (3.21), $\alpha(x_0)$ is equivalent to

$$\begin{aligned}&\tilde{\forall} P_1 \tilde{\forall} Q_1 (\forall x_1 \rhd x_0)(\forall x_2 \rhd x_0)\,(ST_{x_1}(p_1 \wedge \Box q_1)\\ &\qquad\qquad \rightarrow (ST_{x_2}(q_1) \vee ST_{x_2}(\Diamond p_1) \vee ST_{x_2}(p_1)))\end{aligned}$$

and to

$$\tilde{\forall} P_1 \tilde{\forall} Q_1 (\forall x_1 \rhd x_0)(\forall x_2 \rhd x_0)\;(ST_{x_1}(p_1 \wedge \Box q_1) \rightarrow ST_{x_2}(q_1 \vee \Diamond p_1 \vee p_1)).$$

The latter is a type 2 formula; in order to find a Sahlqvist equivalent for it, we proceed as follows:

$$\begin{aligned}
&\tilde{\forall} P_1 \tilde{\forall} Q_1 (\forall x_1 \rhd x_0)(\forall x_2 \rhd x_0)\,(ST_{x_1}(p_1 \wedge \Box q_1) \rightarrow ST_{x_2}(q_1 \vee \Diamond p_1 \vee p_1))\\
\Leftrightarrow\quad & \tilde{\forall} P_1 \tilde{\forall} Q_1 (\forall x_1 \rhd x_0)(\forall x_2 \rhd x_0)\,\neg(ST_{x_1}(p_1 \wedge \Box q_1) \wedge\\
&\qquad\qquad \neg ST_{x_2}(q_1 \vee \Diamond p_1 \vee p_1))\\
\Leftrightarrow\quad & \tilde{\forall} P_1 \tilde{\forall} Q_1 (\forall x_1 \rhd x_0)(\forall x_2 \rhd x_0)\,\neg(ST_{x_1}(p_1 \wedge \Box q_1) \wedge\\
&\qquad\qquad ST_{x_2}(\neg(q_1 \vee \Diamond p_1 \vee p_1)))\\
\Leftrightarrow\quad & \tilde{\forall} P_1 \tilde{\forall} Q_1\, \neg(\exists x_1 \rhd x_0)(\exists x_2 \rhd x_0)\,(ST_{x_1}(p_1 \wedge \Box q_1) \wedge\\
&\qquad\qquad ST_{x_2}(\neg(q_1 \vee \Diamond p_1 \vee p_1)))\\
\Leftrightarrow\quad & \tilde{\forall} P_1 \tilde{\forall} Q_1\, \neg(\exists x_1 \rhd x_0)\,(ST_{x_1}(p_1 \wedge \Box q_1) \wedge\\
&\qquad\qquad (\exists x_2 \rhd x_0) ST_{x_2}(\neg(q_1 \vee \Diamond p_1 \vee p_1)))\\
\Leftrightarrow\quad & \tilde{\forall} P_1 \tilde{\forall} Q_1\, \neg(\exists x_1 \rhd x_0)\,(ST_{x_1}(p_1 \wedge \Box q_1) \;\wedge\; ST_{x_0}(\Diamond\neg(q_1 \vee \Diamond p_1 \vee p_1)))\\
\Leftrightarrow\quad & \tilde{\forall} P_1 \tilde{\forall} Q_1\, \neg((\exists x_1 \rhd x_0) ST_{x_1}(p_1 \wedge \Box q_1) \;\wedge\; ST_{x_0}(\Diamond\neg(q_1 \vee \Diamond p_1 \vee p_1)))\\
\Leftrightarrow\quad & \tilde{\forall} P_1 \tilde{\forall} Q_1\, \neg(ST_{x_0}(\Diamond(p_1 \wedge \Box q_1)) \;\wedge\; ST_{x_0}(\Diamond\neg(q_1 \vee \Diamond p_1 \vee p_1)))\\
\Leftrightarrow\quad & \tilde{\forall} P_1 \tilde{\forall} Q_1\, \neg(ST_{x_0}(\Diamond(p_1 \wedge \Box q_1) \;\wedge\; \Diamond\neg(q_1 \vee \Diamond p_1 \vee p_1)))
\end{aligned}$$

From this, the fastest way to proceed is by observing that the last formula is equivalent to

$$\tilde{\forall} P_1 \tilde{\forall} Q_1\,(ST_{x_0}(\Diamond(p_1 \wedge \Box q_1) \;\rightarrow\; \neg\Diamond\neg(q_1 \vee \Diamond p_1 \vee p_1))),$$

and hence, to the Sahlqvist formula

$$\Diamond(p_1 \wedge \Box q_1) \;\rightarrow\; \Box(q_1 \vee \Diamond p_1 \vee p_1). \quad \dashv$$

Example 3.62 Consider the type 1 Kracht formula

$$\alpha(x_0) \;\equiv\; (\forall x_1 \rhd x_0)(\exists y_1 \rhd x_1)\, y_1 \neq y_1.$$

According to (3.21), we can rewrite $\alpha(x_0)$ into the equivalent

$$\tilde{\forall} P_0 (\forall x_1 \rhd x_0)\;(ST_{x_0}(p_0) \rightarrow (\exists y_1 \rhd x_1) ST_{y_1}(\bot))$$

and, hence, to

$$\tilde{\forall} P_0 (\forall x_1 \rhd x_0)\ (ST_{x_0}(p_0) \to ST_{x_1}(\Diamond\bot))$$

This is a type 2 formula for which we can find a Sahlqvist equivalent as follows:

$$\begin{aligned}
&\tilde{\forall} P_0 (\forall x_1 \rhd x_0)\ (ST_{x_0}(p_0) \to ST_{x_1}(\Diamond\bot)) \\
\Leftrightarrow\quad & \tilde{\forall} P_0 (\forall x_1 \rhd x_0)\ \neg(ST_{x_0}(p_0) \wedge \neg ST_{x_1}(\Diamond\bot)) \\
\Leftrightarrow\quad & \tilde{\forall} P_0 \neg(\exists x_1 \rhd x_0)\ (ST_{x_0}(p_0) \wedge ST_{x_1}(\neg\Diamond\bot)) \\
\Leftrightarrow\quad & \tilde{\forall} P_0 \neg\ (ST_{x_0}(p_0) \wedge (\exists x_1 \rhd x_0) ST_{x_1}(\neg\Diamond\bot)) \\
\Leftrightarrow\quad & \tilde{\forall} P_0 \neg\ (ST_{x_0}(p_0) \wedge ST_{x_0}(\Diamond\neg\Diamond\bot)) \\
\Leftrightarrow\quad & \tilde{\forall} P_0\ (ST_{x_0}(\neg(p_0 \wedge \Diamond\neg\Diamond\bot))
\end{aligned}$$

The latter formula is equivalent to the Sahlqvist formula $p_0 \to \Box\Diamond\bot$. (Obviously, the latter formula is equivalent to $\Box\Diamond\bot$ and, hence, to $\Box\bot$. Our algorithm will not always provide the simplest correspondents!) $\dashv$

This finishes our discussion of Sahlqvist correspondence. In the next chapter we will see that Sahlqvist formulas also have very nice completeness properties, in that any modal logic axiomatized by Sahlqvist formulas is complete with respect to the class of frames defined by (the global first-order correspondents of) the formulas. Here Kracht's theorem can be useful: if we want to axiomatize a class of frames defined by formulas of the form $\forall x\, \alpha(x)$ with $\alpha(x)$ a Kracht formula, then it suffices to compute the Sahlqvist correspondents of these formulas and add these as axioms to the basic modal logic.

Exercises for Section 3.7

3.7.1 (a) Prove that the conjunction $\mathrm{M} \wedge 4$ of McKinsey's formula $\Box\Diamond p \to \Diamond\Box p$ and the transitivity formula $\Diamond p \to \Diamond\Diamond p$ does not have a *local* first-order correspondent. Conclude that this conjunction is not equivalent to a Sahlqvist formula.

(b) Show that on the other hand, the formula $\Box M \wedge 4$ *does* have a local first-order correspondent.

3.7.2 Prove that the local correspondent of a Sahlqvist formula is a Kracht formula.

3.7.3 Find Sahlqvist formulas that locally correspond to the following formulas:

(a) $(\forall y \rhd x)\, Ryy$,
(b) $(\forall y_1 \rhd x)(\forall y_2 \rhd x)(\forall y_3 \rhd x)\, (y_1 = y_2 \vee y_1 = y_3 \vee y_2 = y_3)$
(c) $(\forall y_1 \rhd x)(\forall y_2 \rhd y_1)\, (y_1 = y_2 \vee \exists z\, (Rxz \vee (Ry_1 z \wedge Ry_2 z)))$.
(d) $(\forall x_1 \rhd x)(\exists y_1 \rhd x)(\forall y_2 \rhd y_1)\, (Ry_1 x_1 \vee (Rxy_2 \wedge Ry_x x_1))$

3.7.4 Prove that if $\phi \to \psi$ is a simple Sahlqvist formula, then $\Box(\phi \to \psi)$ is equivalent to a simple Sahlqvist formula.

3.7.5 Let π be the basic temporal similarity type. Show that over the class of bidirectional frames, every simple Sahlqvist formula is equivalent to a very simple Sahlqvist formula. (Hint: first find a very simple Sahlqvist formula that is equivalent to the formula $FGp \to GFp$.)

3.8 Advanced Frame Theory

The main aim of this section is to prove Theorem 3.19, the Goldblatt-Thomason, characterizing the elementary frame classes that are modally definable. We'll also prove a rather technical result needed in our later work on algebras. We'll start by proving the Goldblatt-Thomason Theorem.

Theorem 3.19 *Let τ be a modal similarity type. A first-order definable class K of τ-frames is modally definable if and only if it is closed under taking bounded morphic images, generated subframes, disjoint unions and reflects ultrafilter extensions.*

Proof. The preservation direction follows from earlier results. For the other direction let K be a class of frames which is elementary (hence, closed under taking ultraproducts), closed under taking bounded morphic images, generated subframes and disjoint unions, and reflecting ultrafilter extensions. Let Λ_{K} be the logic of K; that is, $\Lambda_{\mathsf{K}} = \{\phi \mid \mathfrak{F} \Vdash \phi$, for all $\mathfrak{F} \in \mathsf{K}\}$. We will show that Λ_{K} defines K. In order to avoid cumbersome notation we restrict ourselves to the basic modal similarity type.

Let $\mathfrak{F} = (W, R)$ be a frame such that $\mathfrak{F} \Vdash \Lambda_{\mathsf{K}}$. We need to show that $\mathfrak{F}$ is a member of K. This we will do by moving around lots of structures; here's a map of where we are heading for in the proof:

$$\begin{tikzcd}
\{\mathfrak{F}_a \mid a \in W\} \arrow[d, dash, "1"] \\
\mathfrak{F}
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathfrak{ue}\,\mathfrak{F} \arrow[d, dash, "6"] \arrow[dr, dash, "5"] & \mathfrak{G} \arrow[d, dash, "4"] \arrow[dr, dash, "3"] & \{\mathfrak{G}_\delta \mid \delta \subseteq_{fin} \Delta\} \arrow[d, dash, "2"] \\
\mathfrak{F} & \mathfrak{G}' & \prod_U \mathfrak{G}_\delta
\end{tikzcd}$$

First, we can assume without loss of generality that $\mathfrak{F}$ is point-generated. For if $\mathfrak{F}$ validates Λ_{K}, then each of its point-generated does so as well. And if we can prove that each point-generated subframe of $\mathfrak{F}$ is in K, then the membership in K of $\mathfrak{F}$ itself follows immediately from the closure properties of K and the fact that any frame is a bounded morphic image of the disjoint union of its point-generated subframes (as the reader was asked to show in Exercise 3.3.4). So from now on we assume that $\mathfrak{F}$ is generated by the point w.

Now for (one of) the main idea(s) of the proof. Let Φ be a set of propositional variables containing a propositional variable p_A for each subset A of W. This may be a huge language: if W is infinite, then Φ will be uncountable. We will look at the model $\mathfrak{M} = (\mathfrak{F}, V)$ where V is the natural valuation given by $V(p_A) = A$.

Now let Δ be the modal type of w; that is, $\Delta = \{\phi \in ML(\tau, \Phi) \mid \mathfrak{M}, w \Vdash \phi\}$. We claim that

$$\Delta \text{ is satisfiable in } \mathsf{K}. \tag{3.26}$$

In order to prove this, we first show that Δ is finitely satisfiable in K. Let δ be a finite subset of Δ. It is easy to see that δ is satisfiable in K: if it were not, then $\neg \bigwedge \delta$ would belong to Λ_{K} whence we would have $\mathfrak{F} \Vdash \neg \bigwedge \delta$. (Note that whereas Δ is written in a *particular* language, namely, the one having a proposition letter for each subset of W, when we are talking about Λ_{K} we are not really interested in a specific language. This is why we simply assume that '$\neg \bigwedge \delta$ would belong to Λ_{K}' even though we have not verified that this formula uses only proposition letters that occur in Λ_{K}.) But $\mathfrak{F} \Vdash \neg \bigwedge \delta$ would contradict that $\mathfrak{M}, w \Vdash \bigwedge \delta$. But if each finite $\delta \subset \Delta$ is finitely satisfiable in some frame $\mathfrak{G}_\delta$ in K then Δ is satisfiable in some ultraproduct of these frames (the reader is asked to supply a proof of this in Exercise 3.8.2 below). Since K is closed under ultraproducts by assumption, this proves (3.26).

But to say that Δ is satisfiable in K amounts to the following. There is a model $\mathfrak{N} = (X, S, U)$ and a point b in X such that the underlying frame $\mathfrak{G} = (X, S)$ is in K and $\mathfrak{N}, b \Vdash \Delta$. Since K is closed under (point-)generated subframes and modal truth is preserved under taking generated subframes, we may assume that the frame $\mathfrak{G}$ is generated from b.

The only thing left to do is to link up $\mathfrak{G}$ with our original frame $\mathfrak{F}$. This link is as follows.

$$\mathfrak{ue}\,\mathfrak{F} \text{ is a bounded morphic image of some ultrapower of } \mathfrak{G}. \tag{3.27}$$

We first ensure the existence of an m-saturated ultrapower of $\mathfrak{N}$. Note that we may view $\mathfrak{N}$ as a first-order structure for the language $\mathcal{L}^1_\tau(\Phi)$, analogous to the perspective in the previous chapter. Now consider a countably saturated ultrapower of this first-order structure, which we see again as a modal model $\mathfrak{N}' = (X', S', U')$. Note that the existence of such an ultrapower is not guaranteed by Lemma 2.73, since the first-order language $\mathcal{L}^1_{\tau,\mathfrak{F}}$ may not be countable. We need some heavier model-theoretic equipment here; the reader is referred to Theorems 6.1.4 and 6.1.8 in [89]. In any case, $\mathfrak{N}'$ is m-saturated and also has the property that every set Σ that is finitely satisfiable in $\mathfrak{N}'$ is satisfiable in $\mathfrak{N}'$.

How are we going to define the bounded morphism? That is, given an element s of X', which ultrafilter over W (the universe of our original frame $\mathfrak{F}$) are we going to assign to it? Recall that an ultrafilter over W is some collection of subsets of W; this means that given s, we have to decide for each subset of W whether to put it in $f(s)$ or not. But now it will become clear that there is only one natural choice for $f(s)$: simply put a subset A of W in $f(s)$ if p_A is true at s in the model $\mathfrak{N}'$:

$$f(s) = \{A \subseteq W \mid \mathfrak{N}', s \Vdash p_A\}.$$

We will now show that f indeed maps points in $\mathfrak{N}'$ to ultrafilters over W, that f is a bounded morphism, and that f is onto $\mathfrak{ue}\,\mathfrak{F}$. In these proofs, the following equivalence comes in handy:

$$\text{for all formulas } \phi \in ML(\tau, \Phi),\, \mathfrak{M} \Vdash \phi \text{ iff } \mathfrak{N}' \Vdash \phi. \tag{3.28}$$

The proof of (3.28) is by the following chain of equivalences:

$$\begin{array}{lcll}
\mathfrak{M} \Vdash \phi & \Leftrightarrow & \mathfrak{M}, w \Vdash \Box^n \phi \text{ for all } n \in \mathbb{N} & (\mathfrak{M} \text{ is generated from } w) \\
& \Leftrightarrow & \Box^n \phi \in \Delta \text{ for all } n \in \mathbb{N} & (\text{definition of } \Delta) \\
& \Leftrightarrow & \mathfrak{N}, b \Vdash \Box^n \phi \text{ for all } n \in \mathbb{N} & (\text{definition of } \mathfrak{N} \text{ and } b) \\
& \Leftrightarrow & \mathfrak{N} \Vdash \phi & (\mathfrak{N} \text{ is generated from } b) \\
& \Leftrightarrow & \mathfrak{N}' \Vdash \phi & (\mathfrak{N}' \text{ is an ultrapower of } \mathfrak{N})
\end{array}$$

This proves (3.28).

Let us now first check that for all $s \in X'$, $f(s)$ is indeed an ultrafilter over W. We will only check the condition that $f(s)$ is closed under intersection, leaving the other conditions as exercises for the reader. Suppose that A and B are subsets of W that both belong to $f(s)$. Hence, by the definition of $f(s)$ we have that $\mathfrak{N}', s \Vdash p_A$ and $\mathfrak{N}', s \Vdash p_B$. It is easy to see that the formula $p_A \wedge p_B \leftrightarrow p_{A\cap B}$ holds throughout the original model $\mathfrak{M}$. It then follows from (3.28) that $\mathfrak{N}' \Vdash p_A \wedge p_B \leftrightarrow p_{A\cap B}$. In particular, this formula is true at s, so we find that $\mathfrak{N}', s \Vdash p_{A\cap B}$. Hence, by the definition of f, $A \cap B$ belongs to $f(s)$.

In order to show that f is a bounded morphism, we will prove that for all ultrafilters u over W and all points s in X', we have that $u = f(s)$ if and only if u (in $\mathfrak{ue}\,\mathfrak{M}$) and s (in $\mathfrak{N}'$) satisfy the same formulas. This suffices, by Proposition 2.54 and the m-saturation of $\mathfrak{ue}\,\mathfrak{M}$ and $\mathfrak{N}'$. The right to left direction of the equivalence is easy to prove. If the same formulas hold in s and u, then in particular we have for each $A \subseteq W$ that $\mathfrak{N}, s \Vdash p_A$ iff $\mathfrak{ue}\,\mathfrak{M}, u \Vdash p_A$. But by definition of the valuation on $\mathfrak{ue}\,\mathfrak{M}$ we have that $\mathfrak{ue}\,\mathfrak{M}, u \Vdash p_A$ iff $A = V(p_A) \in u$. Hence, we find that $\mathfrak{N}, s \Vdash p_A$ iff $A \in u$. This immediately yields $u = f(s)$.

For the other direction, it suffices to show that for each formula $\phi \in ML(\tau, \Phi)$ and each point s in $\mathfrak{N}'$, $\mathfrak{ue}\,\mathfrak{M}, f(s) \Vdash \phi$ only if $\mathfrak{N}', s \Vdash \phi$. Suppose that ϕ holds at $f(s)$ in $\mathfrak{ue}\,\mathfrak{M}$. By Proposition 2.59 we have that $V(\phi) \in f(s)$. Thus by definition of f we obtain that $\mathfrak{N}', s \Vdash p_{V(\phi)}$. It follows easily from the definition of V that $\mathfrak{M} \Vdash \phi \leftrightarrow p_{V(\phi)}$, so by (3.28) we have that $\mathfrak{N}' \Vdash \phi \leftrightarrow p_{V(\phi)}$. But then we may immediately infer that $\mathfrak{N}', s \Vdash \phi$.

Finally, we have to show that f is surjective; that is, each ultrafilter over W should belong to its range. Let u be such an ultrafilter; we claim that the set $\Sigma = \{p_A \mid A \in u\}$ is finitely satisfiable in $\mathfrak{N}'$. Let σ be a finite subset of Σ. To start with, σ is satisfiable in $\mathfrak{M}$. Since $\mathfrak{M}$ is generated from w, this shows that $\mathfrak{M}, w \Vdash \Diamond^n \bigwedge \sigma$ for some natural number n. From the definition of $\mathfrak{N}$ and b it follows that $\mathfrak{N}, b \Vdash \Diamond^n \bigwedge \sigma$, so from the fact that $\mathfrak{N}$ is point-generated from b we obtain that $\bigwedge \sigma$ is satisfiable in $\mathfrak{N}$. Now $\mathfrak{N}'$ is an ultrapower of $\mathfrak{N}$, so we have that $\bigwedge \sigma$ is also satisfiable in $\mathfrak{N}'$. But $\mathfrak{N}'$ is countably saturated; so Σ, being finitely satisfiable in $\mathfrak{N}'$, is satisfiable in some point s of $\mathfrak{N}'$. It is then immediate that $f(s) = u$.

This proves (3.27), but why does that mean that $\mathfrak{F}$ belongs to K? Here we use the closure properties of K. Recall that $\mathfrak{G}$ is the underlying frame of the model $\mathfrak{N}$ in which we assumed that the set Δ is satisfiable. Since $\mathfrak{G}$ is in K by assumption, $\mathfrak{G}'$ belongs to K by closure under ultraproducts; $\mathfrak{ue}\,\mathfrak{F}$ is in K as it is a bounded morphic image of $\mathfrak{G}'$; and finally, $\mathfrak{F}$ is in K since K reflects ultrafilter extensions. $\dashv$

The following proposition, which is of a rather technical nature, will be put to good use in Chapter 5.

Proposition 3.63 *Let τ be a modal similarity type, and K a class of τ-frames. Suppose that $\mathfrak{G}$ is an ultrapower of the disjoint union $\biguplus_{i \in I} \mathfrak{F}_i$, where $\{\mathfrak{F}_i \mid i \in I\}$ is a family of frames in K. Then $\mathfrak{G}$ is a bounded morphic image of a disjoint union of ultraproducts of frames in K.*

Proof. Let $\mathfrak{F} = (W, R)$ denote the disjoint union $\biguplus_{i \in I} \mathfrak{F}_i$, and assume that $\mathfrak{G}$ is some ultrapower of $\mathfrak{F}$, say $\mathfrak{G} = \prod_U \mathfrak{F}$, where U is an ultrafilter over some index set J. We assume that τ contains only one operator $\triangle$, of arity n. This allows us to write $\mathfrak{F} = (W, R)$ and $\mathfrak{F}_i = (W_i, R_i)$ (that is, the subscript i refers to an index element of I, not to an operator from the similarity type).

Consider an arbitrary state t of $\mathfrak{G}$. By the definition of ultrapowers, there exists a sequence $f_t \in \prod_{j \in J} W$ such that

$$t = (f_t)_U = \{g \in \textstyle\prod_{j \in J} W \mid f_t \sim_U g\}.$$

As W is the disjoint union of the universes W_i, for each $j \in J$ there exists an element $i_j \in I$ such that $f_t(i_j)$ is an element of W_{i_j}. Form the ultraproduct

$$\mathfrak{F}_t := \textstyle\prod_U \mathfrak{F}_{i_j}.$$

Clearly this frame is an ultraproduct of frames in K.

We will now define a map θ_t sending states of the frame $\mathfrak{F}_t$ to states of the frame $\mathfrak{G}$, and show that θ_t is a bounded morphism with t in its range. From this it easily follows that $\mathfrak{G}$ is a bounded morphic image of the disjoint union $\biguplus_{t \in X} \mathfrak{F}_t$, where X is the universe of $\mathfrak{G}$. Observe that a typical element of $\mathfrak{F}_t$ has the form

$$g^U := \{h \in \textstyle\prod_{j \in J} W_{i_j} \mid g \sim_U h\}$$

for some $g \in \prod_{j \in J} W_{i_j}$. Since $\prod_{j \in J} W_{i_j} \subseteq \prod_{j \in J} W$, we have that $g^U \subseteq g_U$. Note that in general these two equivalence classes will not be identical, since g_U may contain elements h for which $h(j') \in W \setminus W_{i_{j'}}$ for some index j'. However, it is evident that *if* both g and h are in $\prod_{j\in J} W_{i_j}$, then we find that $g^U = h^U$ iff $g \sim_U h$ iff $g_U = h_U$. This means that if we put

$$\theta_t(g^U) := g_U,$$

we have found a well-defined map from the universe of $\mathfrak{F}_t$ to the universe X of $\mathfrak{G}$ (in fact, this map is injective).

Now consider the element $f_t \in \prod_{j\in J} W$. By definition of the indices i_j, we must have $f_t \in \prod_{j\in J} W_{i_j}$. It follows that f_t^U is in the domain of θ_t. Now

$$\theta_t(f_t^U) = (f_t)_U = t.$$

It remains to be proved that θ_t is a bounded morphism. However, this follows by a straightforward argument using standard properties of ultrafilters. $\dashv$

Exercises for Section 3.8

3.8.1 Let τ be an arbitrary modal similarity type and $\mathfrak{F}$ a τ-frame. Prove that the ultrafilter extension of $\mathfrak{F}$ is the bounded morphic image of some ω-saturated ultrapower of $\mathfrak{F}$; in other words, supply a proof for Theorem 3.17. (Hint: use an argument analogous to one in the proof of Theorem 3.19. That is, consider a language having a propositional variable p_A for each subset A of the universe of $\mathfrak{F}$, and take a countably saturated ultrapower of the model $\mathfrak{M} = (\mathfrak{F}, V)$, where V is the natural valuation mapping p_A to A for each variable p_A.)

3.8.2 Let K be some class of frames, and Δ a set of formulas which is finitely satisfiable in K. Show that Δ is satisfiable in an ultraproduct of frames in K.

3.8.3 (a) Show that the complement of a modally definable class is closed under taking ultrapowers.

Now suppose that the class K of frames is definable by a *single* formula ϕ.

(b) Show that the complement of K is closed under taking ultraproducts..

Let $\Gamma(\phi)$ be the set of first-order sentences that are semantic consequences of ϕ, in the sense that for any frame $\mathfrak{F}$ we have that $\mathfrak{F} \Vdash \phi$ only if $\mathfrak{F} \models \Gamma(\phi)$. In other words, $\Gamma(\phi)$ is the first-order theory of K.

(c) Prove that ϕ is a semantic consequence of $\Gamma(\phi)$. Hint: reason by contraposition and use (b).
(d) Prove that ϕ is a semantic consequence of a finite subset of $\Gamma(\phi)$. Hint: prove that $\Gamma(\phi) \models \forall x\, ST_x(\phi)$, and use compactness.
(e) Conclude that if a modal formula ϕ defines an elementary frame class, then ϕ corresponds to a (single) first-order formula.

3.8.4 Prove the strong version of the Goldblatt-Thomason Theorem which applies to any frame class that is closed under taking ultrapowers.
(Hint: strengthen the result of Exercise 3.8.2 by showing that any set of modal formulas that is finitely satisfiable in a frame class K is itself satisfiable in an ultrapower of a disjoint union of frames in K.)

3.8.5 Point out where, in the picture summarizing the proof of Theorem 3.19, we use which closure conditions on K. (For instance: in step 2 we need the fact that K is closed under taking ultraproducts.)

3.9 Summary of Chapter 3

- *Frame Definability*: A modal formula is valid on a frame if and only if it is satisfied at every point in the frame, no matter which valuation is used. A modal formula defines a class of frames if and only if it is valid on precisely the frames in that class.
- *Frame Definability is Second-Order*: Because the definition of validity quantifies across all possible valuations, and because valuations are assignments of *subsets* of frames, the concept of validity, and hence frame definability, is intrinsically second-order.
- *Frame Languages*: Every modal formula can be translated into the appropriate second-order frame language. Such languages have an $n+1$-place relation symbol for every n-place modality. Proposition letters correspond to unary predicate variables. The required translation is called the Second-Order Translation. This is simply the standard translation modified to send proposition letters to (unary) predicate *variables* rather than predicate *constants*.
- *Correspondence*: Sometimes the second-order formulas obtained using this translation are equivalent to first-order formulas. But often they correspond to genuinely second-order formulas. This can sometimes be shown by exhibiting a failure of Compactness or the Löwenhein-Skolem property.
- *Frame Constructions*: The four fundamental model constructions discussed in the previous chapter have obvious frame-theoretic counterparts. Moreover, *validity* is preserved under the formation of disjoint unions, generated subframes and bounded morphic images, and anti-preserved under ultrafilter extensions.
- *Goldblatt-Thomason Theorem*: A first-order definable frame class is modally definable if and only if it is closed under disjoint unions, generated subframes and bounded morphic images, and reflects ultrafilter extensions.
- *Modal Definability on Finite Transitive Frames*: A class of finite transitive frames is modally definable if and only if it is preserved under (finite) disjoint unions, generated subframes and bounded morphic images.
- *The Finite Frame Property*: A normal modal logic Λ has the finite frame property if and only if any formula that does not belong to Λ can be falsified on a finite frame that validates all the formulas in Λ. A normal logic has the finite frame property if and only if it has the finite model property.
- *The Sahlqvist Fragment*: Formulas in the Sahlqvist fragment have the property that the second-order formula obtained via the Second-Order Translation can

be reduced to an equivalent first-order formula. The Sahlqvist-Van Benthem algorithm is an effective procedure for carrying out such reductions.

- ▶ *Why Sahlqvist Formulas have First-Order Correspondents*: Syntactically, the Sahlqvist fragment forbids universal operators to take scope over existential or disjunctive connectives in the antecedent. Semantically, this guarantees that we will always be able to find a unique minimal valuation that makes the antecedent true. This ensures that Sahlqvist formulas have first-order correspondents.
- ▶ *Negative Results*: There are non-Sahlqvist formulas that define first-order conditions. Moreover, Chagrova's Theorem tells us that it is undecidable whether a modal formula has a first-order equivalent.
- ▶ *Kracht's Theorem*: Kracht's Theorem takes us back from first-order languages to modal languages. It identifies a class of first-order formulas that are the first-order correspondents of Sahlqvist formulas.
- ▶ *Frames and their Ultrafilter Extensions*: The ultrafilter extension of a frame may be obtained as a bounded morphic image of an ultrapower of the frame.
- ▶ *Ultrapowers of Disjoint Unions*: Ultrapowers of a disjoint union may be obtained as bounded morphic images of disjoint unions of ultraproducts.

Notes

The study of frames has been central to modal logic since the dawn of the classical era (see the Historical Overview in Chapter 1), but the way frames have been studied has changed dramatically over this period. The insight that gave birth to the classical era was that simple properties of frames (such as transitivity and reflexivity) could be used to characterize normal modal logics, and most of the 1960s were devoted to exploring this topic. It is certainly an important topic. For example, in the first half of the following chapter we will see that most commonly encountered modal logics can be given simple, intuitively appealing, frame-based characterizations. But the very success of this line of work meant that for a decade modal logicians paid little attention to modal languages as tools for *describing* frame structure. Frames were simply tools for analyzing normal logics. The notion of frame definability, and the systematic study of modal expressivity over frames, only emerged as a research theme after the frame incompleteness results showed that not all normal logics could be given frame-based characterizations. The first incompleteness result (shown for the basic temporal language) was published in 1972 by S.K. Thomason [426]. The first incompleteness results for the basic modal language were published in 1974 by S.K. Thomason [427] and Kit Fine [137].

The frame incompleteness theorems and the results which accompanied them decisively changed the research agenda of modal logic, essentially because they made it clear that the modal perspective on frames was intrinsically second-order. We've seen ample evidence for this in this chapter: as we saw in Example 3.11

a formula as innocuous looking as McKinsey's $\Box\Diamond p \to \Diamond\Box p$ defines a non-elementary class of frames. This was proved independently by Goldblatt [189] and van Benthem [34]. The proof given in the text is from Theorem 10.2 of van Benthem [41]. It was shown by S.K. Thomason [428] that on the level of frames, modal logic is expressive enough to capture the semantic consequence relation for $\mathcal{L}^2$. Moreover, in unpublished work, Doets showed showed that modal formulas can act as a reduction class for the the theory of finite types; see Benthem [41, 23–24] for further discussion.

So by the mid 1970s it was clear that modal logic embodied a substantial fragment of second-order logic, and a radically different research program was well under way. One strand of this program was algebraic: these years saw the (re)-emergence of algebraic semantics together with a belated appreciation of the work of Jónsson and Tarski [260, 261]; this line of work is treated in Chapter 5. The other strand was the emergence of correspondence theory.

Given that modal logic over frames is essentially second-order logic in disguise, it may seem that the most obvious way to develop correspondence theory would be to chart the second-order powers of modal logic. In fact, examples of modal formulas that define second-order classes of frames were known by the early 1970s (for example, Johan van Benthem proved that the Löb formula defined the class of transitive and converse well-founded frames using the argument given in Example 3.9). And there is interesting work on more general results on second-order frame definability, much of which may be found in Chapters XVII–XIX of van Benthem [41]. Nonetheless, most work on correspondence theory for frames has concentrated on its *first-order* aspects. There are two main reasons for this. First, second-order model theory is less well understood than first-order model theory, so investigations of second-order correspondences have fewer useful results to draw on. Second, there is a clear sense that it is the first-order aspects of frame definability which are truly mysterious (this has long been emphasized by Johan van Benthem). With the benefit of hindsight, the second-order nature of validity is obvious; understanding when — and why — it's sometimes first-order is far harder.

In this chapter we examined the two main strands in first-order correspondence theory (for frames): the *semantic*, exemplified by the Goldblatt-Thomason Theorem, and the *syntactic* exemplified by the Sahlqvist Correspondence Theorem. (Incidentally, as we will learn in Chapter 5, both results have a substantial algebraic dimension.)

What we call the Goldblatt-Thomason Theorem was actually proved by Goldblatt. His result was in fact stronger than our Theorem 3.19, applying to any frame class that is closed under elementary equivalence. This theorem was published in a joint paper [194] with S.K. Thomason, who added a more general result which applies to all definable frame classes but has a less appealing frame construction. The model-theoretic proof of the theorem that we supplied in this chapter is due to van Benthem [45], who also proved the finite transitive version we recorded as Theorem 3.21. Barwise and Moss [27] obtain correspondence results for *models* as opposed to frames; their main result is that if a modal formula ϕ has a first-order frame correspondent c_ϕ, then for all models $\mathfrak{M}$, $\mathfrak{M}$ satisfies all substitution instances of ϕ in infinitary modal logic iff a certain frame underlying $\mathfrak{M}$ satisfies c_ϕ.

Concerning the identification of syntactic classes of modal formulas that correspond to first-order formulas, Sahlqvist's result was not the first. As early as in the Jónsson-Tarski papers [260, 261] particular examples such as reflexivity and transitivity were known. And an article by Fitch [144] was a stimulus for van Benthem's investigations in this area, which lead to van Benthem (unaware of Sahlqvist's earlier work) proving what is now known as Sahlqvist's theorem. But Sahlqvist's paper [388] (essentially a presentation of results contained in his Master's thesis) remains the classic reference in the area. It greatly generalized all previous known results in the area and drew a beautiful link between definability and completeness.

Kracht isolated the first-order formulas that are the correspondents of Sahlqvist formulas in [282], as an application of his so-called calculus of internal describability. This calculus relates modal and first-order formulas on the level of general frames; see also [286].

During the 1990s a number of alternative correspondence languages have been considered for the basic modal language. In the so-called functional translation the accessibility relations are replaced by certain terms which can be seen as functions mapping worlds to accessible worlds. From a certain point of view this functional language is more expressive than the relational language, and that certain second-order frame properties can be mapped to formulas expressed in the functional language — but this is not too surprising: in the functional language one can quantify over functions; this additional expressive power allows one to do without quantification over unary predicate variables; see Ohlbach et al. [350, 349] and Simmons [407].

As with finite model theory, the theory of finite frames is rather underdeveloped. However some of the basic results have been known a long time. We showed in Theorem 3.28 that a normal logic has the finite model property if and only if it has the finite frame property. This result is due to Segerberg [396, Corollary 3.8, page 33]. For some interesting results concerning frame correspondence theory over the class of finite frames the reader should consult the dissertation of Doets [118].

To conclude these Notes, we'll tidy up a few loose ends. Example 3.6.2 is due to van Benthem [41, Theorem 10.4]. Exercise 3.2.4 is based on a result in Fine [140]. Second, we mentioned Chagrova's theorem [87] that it is undecidable whether a modal formula has a first-order equivalent. For pointers to, and a brief discussion of, extensions of this line of work, see Chagrov and Zakharyaschev [86, Chapter 17]. At the end of Section 3.2 we remarked that general frames can be seen as a model version of the *generalized models* or *Henkin models* for second-order logic. Henkin [222] introduced such models, and good discussions of them can be found in Doets and van Benthem [120] or Manzano [320]. Finally, for more on the lambda calculus see Barendregt [23] or Hindley and Seldin [229].