Fragmentary Memories

Carlos Areces

Based on work done in collaboration
with Patrick Blackburn and Maarten Marx
who should actually be considered co-authors of this note.

Abstract

The modern approach to modal logics views them as particular fragmens of first-order logic, which enjoy nice properties like decidability, low complexity, interpolation, etc. If we take this idea at the heels, we can actually use modal logics to capture interesting fragments. In this note we will study one of such fragments, the set of formulas which are invariant under generated submodels. It will turn out to coincide precisely with a beautiful modal logic wich goes under the strange name of $\mathcal{H}(@, \downarrow)$.

Contents

Those who know me will say that indeed my memory is fragmentary. But the title is not referring to that. It is just that this paper is about fragments and memories. Let's start with the latter.

1 Memories

Now I am here, where 'now' is around March, 1999 and 'here' is ILLC, Amsterdam, The Netherlands. But less than three years ago I was finishing my first degree at the Computer Science Department in the University of Buenos Aires, Argentina. The topic of my thesis: Modal Logic and Software Engineering, in which I was making a case for the use of modal languages as specification languages not only in the very late stages of software development (Verification and Testing) but since the very begining (Design).

That was the culminating point of about two or three years of studying and working with modal logics[1]. Again it's not clear for me how I came to know the field. I think that I was already hooked into logic ("good old classical logic"), when somebody came to give a talk about software verification and I first heard about boxes and diamonds (was it Professor Ugo Montanari?).

In any case, the very appealing idea of taking simple propositional logic, adding only two new operators $\Box$ and $\Diamond$, and obtain such a beautiful landscape of possibilities shocked me.

> Let me digress for a second to tell you some details about the Computer Science Department in Buenos Aires. First of all it is quite young, second it is quite small, third (at my time at least and I hope for many more years) it had a great director and excellent professors. Back to modal logics now.

So I was shocked with modal logics, and I wanted to know more about them. Well, I talked to friends. And me and my friends talked to the professors. And me and my friends and the professors talked to the director. And a course on Modal Logic was scheduled for the next semester. Because the department was young we could do these *crazy things*. Because it was small, organizing these things was *easier*. Because we had excellent people teaching us this was *possible*.

The professor in charge was Francisco Naishtat, the course was simply called "Modal Logic" and we were following Hughes and Cresswell's book *An Introduction to Modal Logic* (yes, the 1968 edition).

2 Fragments

And we dived into modal systems: **K**, **T**, **S4**, ... , the whole lot of them. Francisco Naishtat was an invited professor from the Philosophy Department (we had to "import" him because nobody at the Computer Science Department

[1] There seems to be a lost fragment in my memory there, because even though I've been trying to, I cannot remember the exact year when I took my first course on modal logics.

had ever worked with modal logics), and his point of view on Modal Logic was hence tuned to the classical meanings of Necessity and Possibility (only after we asked him we covered also some Temporal Systems, following Prior's books mostly).

We studied the deductive systems and we did (plenty of) completeness proofs. I still remember quite clearly how puzzled I was about some modal formulas "defining second order conditions" like Löb's axiom. How comes that by just "modifying slightly" propositional logic you get to that!

After the first course was over I kept working on modal logics by myself. An then I came across Johan van Benthem's "Correspondence Theory" chapter in the Handbook of Philosophical Logic [7] (a *photocopy* of the chapter I mean, I don't think the library had a copy of the Handbook). And already in the third page I read

> "After all, the clauses of the basic Kripke truth definition amount to a *translation* from modal formulas into classical ones ... "

That was the first time I thought of modal logics as *fragments* of first-order logic instead of *extensions* of propositional logic. The question about second order definability was explained away quickly in the following few pages of [5] by differentiating between model and frame validity.

Of course, this put a completely new map in my mind. For example, together with the modal logics course I've been studying Computability Theory where I learned about the undecidability of first order satisfiability; under that light modal logics appeared as ways to restrict the set of first-order theorems to achieve decidability.

After this adjustment, it didn't take too long for me to get acquainted with things like the standard translation and the characterization of the modal fragment in terms of bisimulations. I simply got and read every article from van Benthem (and the Dutch group in general) I could lay my hands on. It was really like changing your old black and white TV for a new color one. Things were the same, but everything looked so different!

And FRAGMENTS was the name of the game: modal logics ARE ways of defining attractive subsets of full first-order logic (or even extensions like first-order plus fix points).

Going through the pages of van Benthem's "Modal Logic and Classical Logic" (cf. [4]) is to find the fragment-theme once and again. In Chapter XV a very modal like fragment of first-order logic is introduced: the set of formulas which are invariant under generated subframes.

In the next section I will introduce hybrid logics, which are modal logics with the ability to "name" worlds. The name "hybrid logics" is motivated because things that might look like terms (specifically, constants and variables) are used instead as formulas. They will turn to be the modal counterpart of the generated subframe invariant fragment of first-order logic. Things will be clearer soon (I hope!).

3 Hybrid logics

In the following sections things will start getting technical (continue only at your own risk!). If you want a quick resume of it all, we'll just introduce a new logic and identify the first-order fragment it corresponds to. That's it, and you are free to skip the next bunch of pages. But I had to tell you that it is a *beautiful* fragment! It will characterize a "very modal" fragment of first-order logic: The set of first-order formulas which are invariant under generated submodels.

The following sections draw heavily on work I had (and I'm having) the pleasure to share with Patrick Blackburn and Maarten Marx (see [2] for the complete article). If you find mistakes below they are mine; if you find wonderful ideas they are probably theirs.

The first step to obtain a hybrid language is the following: extend the set of propositional letters PROP with a set $\mathsf{NOM} = \{i_1, i_2, \ldots\}$ of *nominals* and a set $\mathsf{WVAR} = \{x_1, x_2, \ldots\}$ of *world variables*. Then the well-formed formulas of the hybrid language are

$$\varphi := a \mid \neg\varphi \mid \varphi \wedge \varphi' \mid \Box\varphi \mid \ \downarrow x_j.\varphi \mid @_s\varphi$$

where $a \in \mathsf{ATOM} = \mathsf{PROP} \cup \mathsf{NOM} \cup \mathsf{WVAR}$, $x_j \in \mathsf{WVAR}$ and $s \in \mathsf{WSYM} = \mathsf{NOM} \cup \mathsf{WVAR}$.

Why can we treat terms as formulas? Because modal formulas are evaluated *locally* (i.e., evaluation takes place at a particular point in the model). Then we can interpreted terms as the question "Does this term represent the actual world?" This is exactly how nominals and world variables will be understood. Furthermore, the $\downarrow$ binder lets us give a generic name to the actual world (compare to the English phrase "Let's say that somebody, say John ... ") while the @ operator let us come back to a named world ("Well, this John, he ... ").

Formally, a hybrid *model* $\mathfrak{M}$ is a Kripke model $\mathfrak{M} = \langle M, R, V\rangle$ where $V : \mathsf{PROP} \cup \mathsf{NOM} \to Pow(M)$ is such that for all nominals $i \in \mathsf{NOM}$, $V(i)$ is a singleton subset of M. We also need the notion of an *assignment* $g : \mathsf{WVAR} \to M$ to handle variables. The x_i-*variant* g^i_m of g is defined as $g^i_m(x_j) = g(x_j)$ for $j \neq i$ and $g^i_m(x_i) = m$.

Let $\mathfrak{M} = \langle M, R, V\rangle$ be a model, $m \in M$, and g an assignment. For any atom a, let $[V, g](a) = \{g(a)\}$ if a is a world variable, and $V(a)$ otherwise. Then the *forcing relation* is defined as follows

$$\begin{array}{lll}
\mathfrak{M}, g, m \Vdash a & \text{iff} & m \in [V,g](a),\ a \in \mathsf{ATOM} \\
\mathfrak{M}, g, m \Vdash \downarrow x_i.\varphi & \text{iff} & \mathfrak{M}, g^i_m, m \Vdash \varphi \\
\mathfrak{M}, g, m \Vdash @_s\varphi & \text{iff} & \mathfrak{M}, g, m' \Vdash \varphi,\ \text{where } [V,g](s) = \{m'\},\ s \in \mathsf{WSYM}.
\end{array}$$

For the other operators the standard definitions apply.

It is perhaps interesting to play a little with the new operators. $\downarrow$ is specially powerful (and it is the main culprit of the undecidability of $\mathcal{H}(\downarrow, @)$). As we hinted above, using $\downarrow$ we plant a flag in the actual word which we can then use as a reference point "further on" in the formula. A wonderful tiny example is

$\downarrow x.\Box\neg x$, which when evaluated on frames forces the accessibility relation to be irreflexive. The intuitive reading is: "Say that this world is called x then every accessible worlds should not be x", banning all reflexive arrows.

Axiomatizations, soundness and completeness for this language which is called $\mathcal{H}(\downarrow, @)$ has already been established in [6]. But we are now interested in discovering which is the first-order fragment corresponding to $\mathcal{H}(\downarrow, @)$.

We begin by providing a *syntactic* characterization. In particular, we shall first extend the *standard translation* (cf. [4]) to $\mathcal{H}(\downarrow, @)$. The range of this translation will lie in certain *bounded fragment* of the first-order language, and we will define a reverse translation HT which maps this bounded fragment back into the hybrid language (in both cases preserving truth). This established a first full correspondence.

But how are these languages characterized *semantically*? $\mathcal{H}(\downarrow, @)$ is a genuine hybrid also at this level as there are two obvious ways to proceed. The first is essentially first-order: we could look for a weaker notion of Ehrenfeucht-Fraïssé game [8]. The second is essentially modal: we could try defining a stronger notion of bisimulation. Both yield natural notions of equivalence between models, and by relating them (and drawing on our syntactic characterization) we can provide a detailed picture of what $\mathcal{H}(\downarrow, @)$ offers.

3.1 Translations

We focus on two kinds of signature for first-order logic with equality. First, we have *modal* signatures (familiar from modal correspondence theory [4]) which consist of one binary predicate R, countably many unary predicates, and no constant symbols. It will be convenient to make the set of first-order variables at our disposal explicit in the signature thus, a modal signature has the form $\langle\{R\}\cup \mathsf{UN\text{-}REL}, \{\}, \mathsf{VAR}\rangle$. A *hybrid* signature is an expansion of the modal signature with countably many constant symbols, thus hybrid signatures have the form $\langle\{R\}\cup \mathsf{UN\text{-}REL}, \mathsf{CONS}, \mathsf{VAR}\rangle$. Note that any hybrid model $\mathfrak{M} = \langle M, R, V\rangle$ can be regarded as a first-order model with domain M, for the accessibility relation R can be used to interpret the binary predicate R, unary predicates can be interpreted by the subsets V assigns to propositional variables, and constants (if any) can be interpreted by the worlds that nominals name. So we let the context determine whether we are working with first-order or hybrid models, and continue to use the notation $\mathfrak{M} = \langle M, R, V\rangle$.

We first extend the standard translation to $\mathcal{H}(\downarrow, @)$. The function ST from the hybrid language over $\langle \mathsf{PROP}, \mathsf{NOM}, \mathsf{WVAR}\rangle$ into first-order logic over the signature $\langle\{R\} \cup \{P_j \mid p_j \in \mathsf{PROP}\}$, NOM, $\mathsf{WVAR} \cup \{x, y\}\rangle$ is defined by mutual recursion between two functions ST_x and ST_y. Recall that $\varphi[x/y]$ means "replace all free instances of x in φ by y."

$ST_x(p_j)$	$=$	$P_j(x)$, $p_j \in \mathsf{PROP}$.	$ST_y(p_j)$	$=$	$P_j(y)$, $p_j \in \mathsf{PROP}$.
$ST_x(i_j)$	$=$	$x = i_j$, $i_j \in \mathsf{NOM}$.	$ST_y(i_j)$	$=$	$y = i_j$, $i_j \in \mathsf{NOM}$.
$ST_x(x_j)$	$=$	$x = x_j$, $x_j \in \mathsf{WVAR}$.	$ST_y(x_j)$	$=$	$y = x_j$, $x_j \in \mathsf{WVAR}$.
$ST_x(\neg\varphi)$	$=$	$\neg ST_x(\varphi)$.	$ST_y(\neg\varphi)$	$=$	$\neg ST_y(\varphi)$.
$ST_x(\varphi \wedge \psi)$	$=$	$ST_x(\varphi) \wedge ST_x(\psi)$.	$ST_y(\varphi \wedge \psi)$	$=$	$ST_y(\varphi) \wedge ST_y(\psi)$.
$ST_x(\Diamond\varphi)$	$=$	$\exists y.(Rxy \wedge ST_y(\varphi))$.	$ST_y(\Diamond\varphi)$	$=$	$\exists x.(Ryx \wedge ST_x(\varphi))$.
$ST_x(\downarrow x_j.\varphi)$	$=$	$(ST_x(\varphi))[x_j/x]$.	$ST_y(\downarrow x_j.\varphi)$	$=$	$(ST_y(\varphi))[x_j/y]$.
$ST_x(@_s\varphi)$	$=$	$(ST_x(\varphi))[x/s]$.	$ST_y(@_s\varphi)$	$=$	$(ST_y(\varphi))[y/s]$.

For m an element in the domain of a given model $\mathfrak{M}$ we will often write $ST_m(\varphi)$ as shorthand for $ST_x(\varphi)[m]$.

Of course, we took care that the translation preserve truth of formulas.

Proposition 3.1 (*ST* Preserves Truth) *Let φ be a hybrid formula, then for all hybrid models $\mathfrak{M}$, $m \in M$ and assignments g, $\mathfrak{M}, g, m \Vdash \varphi$ iff $\mathfrak{M} \models ST_m(\varphi)[g]$.*

PROOF. A straightforward extension of the induction familiar from ordinary modal logics. The only cases which are new are $ST_x(\downarrow x_j.\varphi)$ and $ST_x(@_s\varphi)$. $\mathfrak{M}, g, m \Vdash \downarrow x_j.\varphi$, iff $\mathfrak{M}, g^{x_j}_m, m \Vdash \varphi$. By inductive hypothesis this is the case if and only if $\mathfrak{M} \models ST_m(\varphi)[g^{x_j}_m]$, iff $\mathfrak{M} \models (ST_m(\varphi))[x_j/x]\ \ [g]$. The argument for $ST_x(@_s\varphi)$ is similar. QED

Now for the interesting question: what is the *range* of ST? In fact it belongs to a *bounded fragment* of first-order logic. It was Professor van Benthem who suggested this connection with the bounded fragment. I cannot avoid digressing again

> It was after Maarten Marx presented in a seminar our first ideas about the Hybrid Language and its semantics characterization in terms of bisimulations (next section!). Johan van Benthem was attending the talk, and he seemed busy going through a batch of articles while half listening to Maarten (or so I thought). When the time for questions and discussion arrived, I discovered that somehow he had managed to understand the key ideas without effort. Furthermore, he had already made the connection with the bounded fragment introduced in [4]. He invited us for a short meeting afterwards to put things in order. Just in case we missed any detail he sent us a mail the same night with the main points of the proof of the syntactic characterization. And I would bet, he also finished reading the big bunch of paper he was reviewing in the afternoon!

The bounded fragment now. Given the first-order signature $\langle \{R\} \cup \mathsf{UN\text{-}REL}, \mathsf{CONS}, \mathsf{VAR}\rangle$ we define the bounded fragment as the set of formulas obtained as:

$$\varphi = Rtt' \mid P_j t \mid t = t' \mid \neg\varphi \mid \varphi \wedge \varphi' \mid \exists x_i.(Rtx_i \wedge \varphi) \text{ (for } x_i \neq t)^2$$

[2]The side-condition on the generation of existentially quantified formulas is crucial: it prevents sentences like $\exists x.(Rxx \wedge x = x)$ (which are not preserved under generated submodels) from falling into the fragment.

where $x_i \in \mathsf{VAR}$ and $t, t' \in \mathsf{VAR} \cup \mathsf{CONS}$.

Clearly ST maps hybrid formulas into formulas in the bounded fragment. Crucially, however, we can also translate the bounded fragment into $\mathcal{H}(\downarrow, @)$. The translation HT from the bounded fragment over $\langle \{R\} \cup \mathsf{UN\text{-}REL}, \mathsf{CONS}, \mathsf{VAR}\rangle$ into the hybrid language over $\langle \mathsf{UN\text{-}REL}, \mathsf{CONS}, \mathsf{VAR}\rangle$ is defined as follows:

$$
\begin{array}{lcl}
HT(Rtt') & = & @_t \Diamond t'. \\
HT(P_j t) & = & @_t p_j. \\
HT(t = t') & = & @_t t'. \\
HT(\neg\varphi) & = & \neg HT(\varphi). \\
HT(\varphi \wedge \psi) & = & HT(\varphi) \wedge HT(\psi). \\
HT(\exists v.(Rtv \wedge \varphi)) & = & @_t \Diamond \downarrow v.HT(\varphi).
\end{array}
$$

By construction, $HT(\varphi)$ is a hybrid formula, but furthermore it is a boolean combination of @-formulas (formulas whose main operator is @). We can now prove the following strong truth preservation result. Recall that $\mathfrak{M}, g \Vdash \phi$, abbreviates "(for all $m \in M$) $\mathfrak{M}, g, m \Vdash \phi$".

Proposition 3.2 (HT Preserves Truth) *Let φ be a bounded formula. Then for every first-order model $\mathfrak{M}$ and for every assignment g, $\mathfrak{M} \models \varphi[g]$ iff $\mathfrak{M}, g \Vdash HT(\varphi)$.*

Proof. The proof uses the following fact about boolean combination of @-formulas. Let φ be such a formula, let $\mathfrak{M}$ and g be given and let $m \in M$. Then $\mathfrak{M}, g, m \Vdash \varphi$ iff $\mathfrak{M}, g \Vdash \varphi$.

The proof is by induction and there is only one interesting case: $HT(\exists v.(Rtv \wedge \varphi))$. Now $\mathfrak{M} \models \exists v.(Rtv \wedge \varphi)[g]$ iff $\mathfrak{M} \models (Rtv \wedge \varphi)[g^v_m]$ for some $m \in M$. Let m' be the interpretation of t in $\mathfrak{M}$ under g^v_m. Because of the restriction on variables in bounded quantification, $t \neq v$, whence m' is also the interpretation of t in $\mathfrak{M}$ under g. So $Rm'm$ holds in $\mathfrak{M}$ and $\mathfrak{M} \models \varphi[g^v_m]$. By the inductive hypothesis, $\mathfrak{M}, g^v_m \Vdash HT(\varphi)$ iff $\mathfrak{M}, g, m \Vdash \downarrow v.HT(\varphi)$, iff $\mathfrak{M}, g, m' \Vdash \Diamond \downarrow v.HT(\varphi)$ iff $\mathfrak{M}, g, m' \Vdash @_t \Diamond \downarrow v.HT(\varphi)$ iff $\mathfrak{M}, g \Vdash @_t \Diamond \downarrow v.HT(\varphi)$. QED

As simple corollaries we have:

Corollary 3.3 *Let $\varphi(x)$ be a bounded formula with only x free, then for all models $\mathfrak{M}$ and for all $m \in M$, $\mathfrak{M} \models \varphi(m)$ iff $\mathfrak{M}, m \Vdash \downarrow x.HT(\varphi)$.*

Corollary 3.4 *Let φ be a first-order formula in the hybrid signature. Then the following are equivalent.*

i. φ is equivalent to the standard translation of a hybrid formula.

ii. φ is equivalent to a formula in the bounded fragment.

Moreover, there are effective translations between $\mathcal{H}(\downarrow, @)$ and the bounded fragment.

3.2 Generated back-and-forth systems

We now turn to the problem of providing semantic characterizations of $\mathcal{H}(\downarrow, @)$ (or equivalently, of the bounded fragment). In this section we define *generated back-and-forth systems* —basically a restricted form of Ehrenfeucht-Fraïssé game— and link it to the concept of *generated submodels*.

Generated back-and-forth systems Let $\mathfrak{M}$ and $\mathfrak{N}$ be two first-order models in the hybrid signature. A generated back-and-forth system between $\mathfrak{M}$ and $\mathfrak{N}$ is a non-empty family F of finite partial isomorphisms between $\mathfrak{M}$ and $\mathfrak{N}$ satisfying the following two extension rules:

($\Diamond$-extension)

- (forth) if $h \in F$, x in its domain and $R^{\mathfrak{M}}xy$, then $h \cup \{\langle y, y' \rangle\} \in F$ for some $y' \in N$.
- (back) if $h \in F$, x in its range and $R^{\mathfrak{N}}xy$, then $h \cup \{\langle y', y \rangle\} \in F$ for some $y' \in M$.

(nominal extension)

- (forth) if $h \in F$ and there exists an $x \in M$ such that $V^{\mathfrak{M}}(i) = \{x\}$ for some nominal i, then there exists an $x' \in N$ such that $h \cup \{\langle x, x' \rangle\} \in F$.
- (back) a similar condition backwards.

Let $\bar{m}$ ($\bar{n}$) be a tuple in ${}^kM$ (${}^kN$). Then $(\mathfrak{M}, \bar{m}) \equiv_R (\mathfrak{N}, \bar{n})$ means that there exists a generated back-and-forth system between $\mathfrak{M}$ and $\mathfrak{N}$, containing a partial isomorphism which sends $m(i)$ to $n(i)$.

Note how closely this definition follows the familiar one from first-order logic (cf. e.g., [8]). In fact, if we think of such a system as describing an Ehrenfeucht-Fraïssé game, then the sole difference is that in the "generated back-and-forth game" the universal player must choose his moves from R-successors or worlds named by a nominal, whereas he can play whatever he likes in the full first-order game. Unsurprisingly, restricting the play to *accessible* worlds puts generated back-and-forth systems very close to *generated submodels.*

Definition 3.5 (Generated Submodel) Let $\mathfrak{M} = \langle M, R, V \rangle$ be a hybrid model and $S \subseteq M$. Let NOM denote the subset of M whose elements are in the interpretation of some nominal. The *submodel of* $\mathfrak{M}$ *generated by* S is the substructure of $\mathfrak{M}$ with domain $\{m \in M \mid \text{exists } s \in S \cup \mathsf{NOM} \text{ s.t. } R^*sm\}$, where R^* is the reflexive and transitive closure of R.

Note that if $\mathsf{NOM} = \emptyset$, we obtain the familiar modal notion of a generated submodel; and that if in addition S is a singleton set, we have the usual notion of a point-generated submodel.

We now define two kinds of invariance. The first is taken from [4]. A first-order formula $\varphi(\bar{x})$ in free variables $\bar{x}$ in a signature with one binary relation R, unary predicates and constants (and equality) is *invariant for generated submodels* if for all models $(\mathfrak{M}, \bar{m})$ and $(\mathfrak{M}', \bar{m})$ such that $\mathfrak{M}'$ is the $\bar{m}$-generated submodel of $\mathfrak{M}$, $\mathfrak{M} \models \varphi(\bar{m})$ iff $\mathfrak{M}' \models \varphi(\bar{m})$. In a similar spirit, we shall say that a first-order formula $\varphi(\bar{x})$ in the same signature is invariant for generated back-and-forth systems if for all models $(\mathfrak{M}, \bar{m})$ and $(\mathfrak{N}, \bar{n})$, $(\mathfrak{M}, \bar{m}) \equiv_R (\mathfrak{N}, \bar{n})$ implies ($\mathfrak{M} \models \varphi(\bar{m})$ iff $\mathfrak{N} \models \varphi(\bar{n})$).

Theorem 3.6 *Let $\varphi(\bar{x})$ be a first-order formula in the hybrid signature. Then the following are equivalent.*

i. $\varphi(\bar{x})$ is equivalent to a formula in the bounded fragment.

ii. $\varphi(\bar{x})$ is invariant for generated submodels.

iii. $\varphi(\bar{x})$ is invariant for generated back-and-forth systems.

PROOF.

i. $\Rightarrow$ *ii.* is obvious.

ii. $\Rightarrow$ *iii.* First note that $\varphi(\bar{x})$ is invariant for generated submodels if and only if $\neg\varphi(\bar{x})$ is. Now suppose $\varphi(\bar{x})$ is invariant for generated submodels but *not* preserved under generated back-and-forth systems. Then we have models $(\mathfrak{M}, \bar{m})$ and $(\mathfrak{N}, \bar{n})$, a generated back-and-forth system linking $\bar{m}$ and $\bar{n}$, and $\mathfrak{M} \models \varphi(\bar{m})$ while $\mathfrak{N} \models \neg\varphi(\bar{n})$.

Let $\mathfrak{M}'$ ($\mathfrak{N}'$) be the $\bar{m}$- ($\bar{n}$) generated submodel of $\mathfrak{M}$ ($\mathfrak{N}$). Then still $\mathfrak{M}' \models \varphi(\bar{m})$ and $\mathfrak{N}' \models \neg\varphi(\bar{n})$, by invariance. Clearly $(\mathfrak{M}', m) \equiv_R (\mathfrak{N}', n)$. But then $(\mathfrak{M}', m)$ and $(\mathfrak{N}', n)$ have the same first-order theory by the following argument. Because $(\mathfrak{M}', m) \equiv_R (\mathfrak{N}', n)$ holds, $\exists$loise (the existential player) has a winning strategy in all games where $\forall$belard (the universal player) only plays *immediate* R-successors or points named by a nominal. But since the models are generated, in the first-order back-and-forth game he can only play worlds which are accessible by a finite R-transition from either the root or one of the named worlds. But then she can compute a winning answer for the classic Ehrenfeucht-Fraïssé game from her winning generated back-and-forth strategy. This contradicts the claim that $\mathfrak{M}' \models \varphi(\bar{m})$ and $\mathfrak{N}' \models \neg\varphi(\bar{n})$.

iii. $\Rightarrow$ *i.* We use a "van Benthem style" diagram-chasing argument. We only provide the outline. Let $\varphi(\bar{x})$ be as in the hypothesis and $BC(\varphi(\bar{x}))$ be the bounded consequences of $\varphi(\bar{x})$ (that is, the consequences of $\varphi(\bar{x})$ that belong to the bounded fragment). We will show that $BC(\varphi(\bar{x})) \models \varphi(\bar{x})$, from which the result follows by compactness. (In this notation we interpret the $\bar{x}$ as constants, or equivalently we use the local version of first-order consequence.)

If $BC(\varphi(\bar{x}))$ is inconsistent we are done. Otherwise, let $\mathfrak{M}, \bar{m}$ be a model of $BC(\varphi(\bar{x}))$ and $\mathfrak{N}, \bar{n}$ be a model of $\varphi(\bar{x})$ together with the bounded theory of $\mathfrak{M}, \bar{m}$. (Such a model can easily be shown to exist.) Take ω-saturated extensions $\mathfrak{M}^+, \bar{m}$ and $\mathfrak{N}^+, \bar{n}$. Create a family F of finite functions between M^+ and N^+ as follows: $f : \bar{x} \to \bar{y}$ is in F iff $\mathfrak{M}^+, \bar{x}$ and $\mathfrak{N}^+, \bar{y}$ make the same bounded formulas true. It is easy to show that F is a generated back and forth system linking $\bar{m}$ and $\bar{n}$. Now we can start diagram chasing: $\mathfrak{N} \models \varphi(\bar{n})$ then (by elementary extension) $\mathfrak{N}^+ \models \varphi(\bar{n})$, then (by invariance) $\mathfrak{M}^+ \models \varphi(\bar{m})$, then (passing to an elementary submodel) $\mathfrak{M} \models \varphi(\bar{m})$ as desired. QED

3.3 Hybrid bisimulations

We have just seen that by weakening the notion of an Ehrenfeucht-Fraïssé game we can link the bounded fragment (and hence $\mathcal{H}(\downarrow, @)$) with generated submodels. But in spite of its binding apparatus, $\mathcal{H}(\downarrow, @)$ has a distinctly modal flavor. Isn't it also possible to strengthen the notion of *bisimulation* (the standard notion of equivalence between models used in modal logic) with clauses for $\downarrow$ and

@, and so characterize $\mathcal{H}(\downarrow, @)$ in intrinsically modal terms? That's what we will do in this section. The approach has an advantage over the use of generated back-and-forth systems: preservation results can be easily obtained for reducts as well.

Recall that for ordinary propositional modal logics, bisimulations are non-empty binary relations linking the domains of models, with the restriction that only worlds with identical atomic information and matching accessibility relations should be connected (see Definition 3.7 [4]; here bisimulations are called p-relations). Now, if we want to extend this notion to $\mathcal{H}(\downarrow, @)$, we need to take care of assignments to world variables as well. To this end, hybrid bisimulations will not simply link worlds, rather they will link pairs $(\bar{m}, m)$, where m is a world and $\bar{m}$ is an assignment. We start by defining k-bisimulations, which are the correct notion of bisimulation for formulas φ such that $\mathsf{WVAR}(\varphi) \subseteq \{x_1, \ldots, x_k\}$.

k-bisimulation. Let $\mathfrak{M}$ and $\mathfrak{N}$ be two hybrid models. Let $\stackrel{k}{\sim}$ be a binary relation between ${}^kM \times M$ and ${}^kN \times N$. So $\stackrel{k}{\sim}$ relates tuples $((m_1, \ldots, m_k), m)$ with tuples $((n_1, \ldots, n_k), n)$. We write these tuples as $(\bar{m}, m)$. Notice that $\bar{m}$ can be seen as an assignment over $(x_1, \ldots, x_k)$. A non-empty relation $\stackrel{k}{\sim}$ is called a *k-bisimulation* if it satisfies the following properties

(prop) If $(\bar{m}, m) \stackrel{k}{\sim} (\bar{n}, n)$, then $m \in V^{\mathfrak{M}}(a)$ iff $n \in V^{\mathfrak{N}}(a)$, for $a \in \mathsf{PROP} \cup \mathsf{NOM}$.

(var) If $(\bar{m}, m) \stackrel{k}{\sim} (\bar{n}, n)$, then for all $j \leq k$, $m_j = m$ iff $n_j = n$.

(forth) If $(\bar{m}, m) \stackrel{k}{\sim} (\bar{n}, n)$ and $R^{\mathfrak{M}}mm'$, then there exists an $n' \in N$ such that $R^{\mathfrak{N}}nn'$ and $(\bar{m}, m') \stackrel{k}{\sim} (\bar{n}, n')$.

(back) A similar condition from $\mathfrak{N}$ to $\mathfrak{M}$.

(@) If $(\bar{m}, m) \stackrel{k}{\sim} (\bar{n}, n)$, then for every nominal $i \in \mathsf{NOM}$, if $m' \in V^{\mathfrak{M}}(i)$ and $n' \in V^{\mathfrak{N}}(i)$ then $(\bar{m}, m') \stackrel{k}{\sim} (\bar{n}, n')$, and for every $j \leq k$, $(\bar{m}, m_j) \stackrel{k}{\sim} (\bar{n}, n_j)$

($\downarrow$) If $(\bar{m}, m) \stackrel{k}{\sim} (\bar{n}, n)$, then for every $j \leq k$, $(\bar{m}^j_m, m) \stackrel{k}{\sim} (\bar{n}^j_n, n)$.

Note that since $\downarrow$ and @ are self-dual, we can collapse the back and forth clauses for these modalities into one. We write $\mathfrak{M} \stackrel{k}{\sim} \mathfrak{N}$ if there exists a k-bisimulation between the two models.

To extend the notion to the full language we need to add only one further condition.

ω-bisimulation. Let $\mathfrak{M}$ and $\mathfrak{N}$ be two hybrid models. An ω-bisimulation between $\mathfrak{M}$ and $\mathfrak{N}$ is a non-empty family of k-bisimulations satisfying the following *storage rule*:

(sto) If $(\bar{m}, m) \stackrel{k}{\sim} (\bar{n}, n)$, then $((\bar{m}, m), m) \stackrel{k+1}{\sim} ((\bar{n}, n), n)$.

Let $\bar{m}$ ($\bar{n}$) be a tuple in ${}^kM$ (${}^kN$). Then $(\mathfrak{M}, \bar{m}) \overset{\omega}{\sim} (\mathfrak{N}, \bar{n})$ means that there exists an ω-bisimulation between $\mathfrak{M}$ and $\mathfrak{N}$ such that $(\bar{m}, \bar{m}(0)) \overset{k}{\sim} (\bar{n}, \bar{n}(0))$.

Notice that the modular definition of k and ω-bisimulation will lead to results for reducts of the language as well. For instance if we delete $\downarrow$ from the language, we just delete the **($\downarrow$)** clause from the definition of bisimulation and we obtain the appropriate notion for the language $\mathcal{H}(@)$. Of course, if we through away the variables we don't need the assignment tuples anymore, and the bisimulation becomes just a relation between worlds, as usual. Then for the language without $\downarrow, @$ and variables, the standard definition of bisimulation applies (the condition **(prop)** takes care of the nominals). If we add @ to this language, we just have to add the following clause

(@′) For all nominals i, if $V^{\mathfrak{M}}(i) = \{m\}$ and $V^{\mathfrak{N}}(i) = \{n\}$, then $m \sim n$.

Preservation results for all these alternatives can be given (the required proofs follow much the same lines as the proofs below).

The first important fact about hybrid bisimulations is that they preserve truth:

Proposition 3.7 *Let $\mathfrak{M}$ and $\mathfrak{N}$ be two hybrid models, $m \in M$, $n \in N$. Then,*

i. If $\mathfrak{M} \overset{k}{\sim} \mathfrak{N}$, then for all formulas φ over the signature $\langle \mathsf{PROP}, \mathsf{NOM}, \{x_1, \ldots, x_k\}\rangle$, $(\bar{m}, m) \overset{k}{\sim} (\bar{n}, n)$ implies $\mathfrak{M}, \bar{m}, m \Vdash \varphi \Leftrightarrow \mathfrak{N}, \bar{n}, n \Vdash \varphi$.

ii. If $(\mathfrak{M}, m) \overset{\omega}{\sim} (\mathfrak{N}, n)$, then for all sentences *φ over the signature $\langle \mathsf{PROP}, \mathsf{NOM}, \mathsf{WVAR}\langle$, $\mathfrak{M}, m \Vdash \varphi \Leftrightarrow \mathfrak{N}, n \Vdash \varphi$. (Recall that for sentences the choice of assignment is irrelevant.)*

Proof.

i. By a straightforward inductive argument.

ii. Let $(\mathfrak{M}, m) \overset{\omega}{\sim} (\mathfrak{N}, n)$ and let φ be a hybrid sentence. Then it contains some variables, say $\{x_1, \ldots, x_k\}$. We have $(\langle m\rangle, m) \overset{1}{\sim} (\langle n\rangle, n)$, so $k-1$ applications of the storage rule gives $(\bar{m}, m) \overset{k}{\sim} (\bar{n}, n)$, where $\bar{m}$ is a k-tuple consisting of m's and similarly for $\bar{n}$. By *i.*, $\mathfrak{M}, \bar{m}, m \Vdash \varphi \Leftrightarrow \mathfrak{N}, \bar{n}, n \Vdash \varphi$, whence since φ is a sentence $\mathfrak{M}, m \Vdash \varphi \Leftrightarrow \mathfrak{N}, n \Vdash \varphi$. QED

The notion of k-bisimulation has a distinct modal flavor. But a very first-order notion is hidden behind: partial isomorphism.

Proposition 3.8 *Let $k \geq 2$, and let $\mathfrak{M} \overset{k}{\sim} \mathfrak{N}$. If $(\bar{m}, m) \overset{k}{\sim} (\bar{n}, n)$, then the function f defined as $f(m) = n$ and $f(m(i)) = n(i)$ is a partial isomorphism between $\{m(1), \ldots, m(k), m\}$ and $\{n(1), \ldots, n(k), n\}$.*

Proof. The map f is a bijection by **(var)** and **(@)**. By **(prop)** and **(@)**, f preserves nominals and propositional variables. To see that it preserves the accessibility relation suppose $R^{\mathfrak{M}}xy$. There are three cases.

(Case 1: $x = m$, $y = m_i$.) Then by **(forth)** there exists an n' such that $R^{\mathfrak{N}}nn'$ and $(\bar{m}, m_i) \overset{k}{\sim} (\bar{n}, n')$. But $\bar{m}(i) = m_i$, so by **(var)**, $n' = \bar{n}(i)$, whence $R^{\mathfrak{N}}nf(m(i))$.

(Case 2: $x = m_i$, $y = m$.) Let $j \neq i$. Such a j exists because we assumed that $k \geq 2$. By **($\downarrow$)**, $(\bar{m}^j_m, m) \overset{k}{\sim} (\bar{n}^j_n, n)$. Then by **(@)**, $(\bar{m}^j_m, m_i) \overset{k}{\sim} (\bar{n}^j_n, n_i)$. Now continue as in case 1.

(Case 3: $x = m_i$, $y = m_j$.) By **(@)**, $(\bar{m}, m_i) \overset{k}{\sim} (\bar{n}, n_i)$. Now continue as in case 1.
Thus $R^{\mathfrak{M}}xy$ implies $R^{\mathfrak{N}}f(x)f(y)$. For the other direction use **(back)** in the same way. QED

Thus there is a clear link with our earlier work on generated back-and-forth systems. After Proposition 3.8 the following theorem shouldn't come as a surprise:

Theorem 3.9 *Let $(\mathfrak{M}, \bar{m})$ and $(\mathfrak{N}, \bar{n})$ be two models. Then the following are equivalent:*
i. $(\mathfrak{M}, \bar{m}) \overset{\omega}{\sim} (\mathfrak{N}, \bar{n})$.
ii. $(\mathfrak{M}, \bar{m}) \equiv_R (\mathfrak{N}, \bar{n})$.

It is time to draw together the threads developed in the previous section. First we note their consequences for expressivity over models. Then we note the consequences for frames and what this tells us about hybrid completeness.

3.4 Expressivity over models

We have the following five-fold characterization of $\mathcal{H}(\downarrow, @)$:

Theorem 3.10 *Let $\varphi(\bar{x})$ be a first-order formula in the hybrid signature (with equality). Then the following are equivalent.*
i. $\varphi(\bar{x})$ is equivalent to the standard translation of a $\mathcal{H}(\downarrow, @)$ formula.
ii. $\varphi(\bar{x})$ is invariant for generated submodels.
iii. $\varphi(\bar{x})$ is invariant for generated back-and-forth systems.
iv. $\varphi(\bar{x})$ is invariant for ω-bisimulation.
v. $\varphi(\bar{x})$ is equivalent to a formula in the bounded fragment of first-order logic.

Proof. By Corollary 3.4, Theorem 3.6 and Proposition 3.7. QED

But these have obvious consequences for the ordinary modal correspondence language. In particular, if we consider *nominal-free* hybrid sentences, then we obtain a five-fold characterization of the fragment of first-order logic in the classical modal signature which is invariant for generated submodels:

Corollary 3.11 *Let $\varphi(x)$ be a first-order formula in the modal signature with equality. Then the following are equivalent.*
i. $\varphi(x)$ is equivalent to the standard translation of a nominal-free $\mathcal{H}(\downarrow, @)$ sentence.
ii. $\varphi(x)$ is invariant for generated submodels (now in the standard modal sense).

iii. $\varphi(x)$ *is invariant for R-generated back-and-forth systems, where an R-generated back-and-forth system is a back-and-forth system satisfying only the* **$\Diamond$-extension** *rule.*

iv. $\varphi(x)$ *is invariant for ω-bisimulation.*

v. $\varphi(x)$ *is equivalent to a formula in the bounded fragment of first-order logic without constants.*

3.5 Frames and completeness

Recall that a frame $\mathfrak{F}$ is a pair $\langle W, R\rangle$ (that is, a model without a valuation). Since the late 1950s, one of the central topics in modal logic has been linking modal formulas to properties of frames and investigating when they give rise to complete axiomatizations for the frame classes they define. The work in the previous section easily yields a characterization of the frame-defining abilities of pure nominal-free sentences. Moreover, the axiomatic investigations of [6] show that there is a perfect match between definability and completeness for pure nominal-free sentences. By combining these results we obtain matching definability and completeness results for a wide range of first-order definable frame classes.

In modal correspondence theory, the first-order language (with equality) over the signature consisting simply of a binary symbol R is called the (first-order) *frame language.* We shall call a formula φ in the frame language containing exactly one free variable a *frame condition.* The class of frames defined by a frame condition $\varphi(x)$ is the class in which the universal closure $\forall x.\varphi(x)$ is true; we call this class $\textsc{frames}(\forall x.\varphi(x))$.

Before proceeding further, two simple observations are in order. First, note that if we apply the standard translation ST to a pure nominal-free sentence α , then $ST(\alpha)$ is a frame condition with free-variable x. Furthermore, note that for any frame $\mathfrak{F} = \langle W, R\rangle$ we have that $\mathfrak{F} \Vdash \alpha$ iff $\mathfrak{F} \Vdash \forall x.ST(\alpha)$; this is an immediate consequence of the definition of frame validity.

Theorem 3.12 *Let $\mathsf{K}[\mathcal{H}(\downarrow, @)]$ be the axiomatization for $\mathcal{H}(\downarrow, @)$ given in [6], and for any hybrid sentence α let $\mathsf{K}[\mathcal{H}(\downarrow, @)] + \alpha$ be the system obtained by adding α as an additional axiom. Then, if $\varphi(x)$ is a frame condition and $\varphi(x)$ is invariant under generated submodels (in the usual modal sense) we have that:*

i. If $\varphi(x)$ is in the bounded fragment, then the pure nominal free sentence $\downarrow x.HT(\varphi(x))$ defines $\textsc{frames}(\forall x.\varphi(x))$. Moreover, the logic $\mathsf{K}[\mathcal{H}(\downarrow, @)]$ $+ \downarrow x.HT(\varphi(x))$ is strongly complete with respect to $\textsc{frames}(\forall x.\varphi(x))$.

ii. If $\varphi(x)$ is not in the bounded fragment, there is a nominal free sentence α such that α defines $\textsc{frames}(\forall x.\varphi(x))$, and $ST(\alpha)$ is equivalent to $\varphi(x)$. Moreover, $\mathsf{K}[\mathcal{H}(\downarrow, @)] + \alpha$ is strongly complete with respect to $\textsc{frames}(\forall x.\varphi(x))$.

Conversely, if α is a pure nominal-free sentence, then α defines the class of frames $\textsc{frames}(\forall x.ST(\alpha(x)))$, and $\mathsf{K}[\mathcal{H}(\downarrow, @)] + \alpha$ is complete with respect to $\textsc{frames}(\forall x.ST(\alpha(x)))$.

PROOF. The converse condition was proved in [6], so we examine the other direction.

For item *i.*, we first remark that as $\varphi(x)$ belongs to the *frame* language, it contains no unary predicate symbols, hence $HT(\varphi(x))$ is a *pure* formula; that $\downarrow x.HT(\varphi(x))$ is a pure nominal-free sentence is thus clear. Now, by Corollary 3.3, for any model $\mathfrak{M} = (\mathfrak{F}, V)$ and any $m \in M$,

$$(\mathfrak{F}, V) \models \varphi[m] \text{ iff } (\mathfrak{F}, V), m \Vdash \downarrow x.HT(\varphi).$$

But this means that

$$(\mathfrak{F}, V) \models \forall x.\varphi \text{ iff } (\mathfrak{F}, V) \Vdash \downarrow x.HT(\varphi).$$

As $\varphi(x)$ contains no unary predicate symbols (and $\downarrow x.HT(\varphi)$ no propositional variables) V is irrelevant, and hence

$$\mathfrak{F} \models \forall x.\varphi(x) \text{ iff } \mathfrak{F} \Vdash \downarrow x.HT(\varphi).$$

This implies that $\downarrow x.HT(\varphi(x))$ defines $\text{FRAMES}(\forall x.\varphi(x))$. Completeness follows using the arguments of [6].

For item *ii.*, we know that $\varphi(x)$ being invariant under generated submodels is equivalent to a formula in the bounded fragment; but is it equivalent to a *frame condition* $\varphi'(x)$? In fact, this can be established by modifying the diagram chasing argument used in the proof of Theorem 3.6. The key point to observe is that instead of showing that $BC(\varphi(x)) \models \varphi(x)$, we can show by the same method that $FC(\varphi(x)) \models \varphi(x)$, where FC are all the frame conditions implied by $\varphi(x)$. Thus there is an equivalent frame condition $\varphi'(x)$, and we can take α to be $\downarrow x.HT(\varphi'(x))$. The remainder of the proof is as for item *i.* QED

4 'It's a poor sort of memory that only works backwards'

As the Queen in *Alice through the Looking-Glass* pointed out, wouldn't it be more satisfying if we could just remember a bit of what is to come? To remember (even if vaguely) of when we will be able to understand how to define the *correct modal fragment* for a given, specific need; and have general results that will let us measure in advance which is the minimal complexity and which are the properties of such a language?

Some beautiful results have already been given to us, e.g. the guarded fragments of Andréka, van Benthem and Németi [1] ("if a logic can be mapped *here* then it is decidable") or the work on complexity of modal logics of Edith Spaan [9] ("if a logic is able to express *this* then it has at least *this* complexity). Also the conditions for failure of interpolation provided in [3] are in a similar line.

Which are the well behaved fragments and, most interestingly, why? These are indeed important questions which the modern approach to modal logics is starting to unravel.

And hybrid logics? Are they *just one of the many*? Play with them for a while and you will perhaps answer the question by yourself. Nominals seems to add to modal logic a new dimension (like filling a hole that only now we notice it was there). Modal logic is locality itself; this is the main reason of its good behavior. And once a local point of view is adopted, once we evaluate formulas in a particular point in the model, then the concept of "terms as formulas" comes just so naturally.

Hybrid logics will probably play a role in the next years. They offer high expressive power, an elegant proof theory and plenty of connections with other fields like temporal reasoning and knowledge representation. For the moment they taught us a bit more about the structure of the landscape of fragments we are swimming in.

References

[1] H. Andréka, J. van Benthem, and I. Németi. Back and forth between modal logic and classical logic. *Bulletin of the Interest Group in Pure and Applied Logics*, 3(3):685–720, 1995.

[2] C. Areces, P. Blackburn, and M. Marx. Hybrid logics. Characterization, interpolation and complexity. Submitted to the *Journal of Symbolic Logic*, February 1999.

[3] C. Areces and M. Marx. Failure of interpolation in combined modal logics, 1997. To appear in the *Notre Dame Journal of Formal Logic*.

[4] J. van Benthem. *Modal Logic and Classical Logic*. Bibliopolis, Naples, 1983.

[5] J. van Benthem. Correspondence theory. In Gabbay and Guenthner [7], pages 167–248. Extensions of classical logic.

[6] P. Blackburn and M. Tzakova. Hybrid languages and temporal logics. *Logic Journal of the IGPL*, 7(1):27–54, 1999.

[7] D. Gabbay and F. Guenthner, editors. volume 165 of *Synthese Library*. D. Reidel Publishing Co., Dordrecht, 1984. Extensions of classical logic.

[8] W. Hodges. *Model theory*. Cambridge University Press, Cambridge, 1993.

[9] E. Spaan. *Complexity of modal logics*. PhD thesis, University of Amsterdam, Institute for Logic, Language and Computation, 1993.