

## NEIGHBOURHOOD STRUCTURES: BISIMILARITY AND BASIC MODEL THEORY\*

HELLE HVID HANSEN<sup>a</sup>, CLEMENS KUPKE<sup>b</sup>, AND ERIC PACUIT<sup>c</sup>

<sup>a</sup> Eindhoven University of Technology, FM group, P.O. Box 513, 5600 MB Eindhoven, Netherlands.  
*e-mail address:* h.h.hansen@tue.nl

<sup>b</sup> Imperial College London, Department of Computing, 180 Queen’s Gate, London SW7 2AZ, UK.  
*e-mail address:* ckupke@doc.ic.ac.uk

<sup>c</sup> Stanford University, Department of Philosophy, Stanford, CA 94305-2155, USA.  
*e-mail address:* pacuit@stanford.edu

**ABSTRACT.** Neighbourhood structures are the standard semantic tool used to reason about non-normal modal logics. The logic of all neighbourhood models is called classical modal logic. In coalgebraic terms, a neighbourhood frame is a coalgebra for the contravariant powerset functor composed with itself, denoted by  $2^2$ . We use this coalgebraic modelling to derive notions of equivalence between neighbourhood structures.  $2^2$ -bisimilarity and behavioural equivalence are well known coalgebraic concepts, and they are distinct, since  $2^2$  does not preserve weak pullbacks. We introduce a third, intermediate notion whose witnessing relations we call precocongruences (based on pushouts). We give back-and-forth style characterisations for  $2^2$ -bisimulations and precocongruences, we show that on a single coalgebra, precocongruences capture behavioural equivalence, and that between neighbourhood structures, precocongruences are a better approximation of behavioural equivalence than  $2^2$ -bisimulations. We also introduce a notion of modal saturation for neighbourhood models, and investigate its relationship with definability and image-finiteness. We prove a Hennessy-Milner theorem for modally saturated and for image-finite neighbourhood models. Our main results are an analogue of Van Benthem’s characterisation theorem and a model-theoretic proof of Craig interpolation for classical modal logic.

### 1. INTRODUCTION

Neighbourhood semantics [13] forms a generalisation of Kripke semantics, and it has become the standard tool for reasoning about non-normal modal logics in which (Kripke valid) principles such as  $\Box p \wedge \Box q \rightarrow \Box(p \wedge q)$  and  $\Box p \rightarrow \Box(p \vee q)$  are considered not to

*1998 ACM Subject Classification:* F.1.1, F.3.2, F.4.1, I.2.4.

*Key words and phrases:* Neighbourhood semantics, non-normal modal logic, bisimulation, behavioural equivalence, invariance.

\* This is an extended and revised version of [22].

<sup>a</sup> Supported by NWO grant 612.000.316.

<sup>b</sup> Supported by NWO under FOCUS/BRICKS grant 642.000.502.

<sup>c</sup> Supported by NSF grant OISE 0502312.

hold. In a neighbourhood model, with each state one associates a collection of subsets of the universe (called its neighbourhoods), and a modal formula  $\Box\varphi$  is true at a state  $s$  if the truth set of  $\varphi$  is a neighbourhood of  $s$ . The modal logic of all neighbourhood models is called classical modal logic.

Neighbourhood semantics was invented in 1970 by Scott and Montague (independently in [41] and [31]); and Segerberg [42] presents some basic results about neighbourhood models and the classical modal logics that correspond to them. These and other salient results were incorporated by Chellas in his textbook [13]. During the past 15-20 years, non-normal modal logics have emerged in the areas of computer science and social choice theory, where system (or agent) properties are formalised in terms of various notions of ability in strategic games (e.g. [4, 38]). These logics have in common that they are monotonic, meaning they contain the above-mentioned formula  $\Box p \rightarrow \Box(p \vee q)$ . The corresponding property of neighbourhood models is that neighbourhood collections are closed under supersets. Non-monotonic modal logics occur in deontic logic (see e.g. [17]) where monotonicity can lead to paradoxical obligations, and in the modelling of knowledge and related epistemic notions (cf. [43, 33]). Furthermore, the topological semantics of modal logic can be seen as neighbourhood semantics (see [11] and references).

Neighbourhood frames are easily seen to be coalgebras for the contravariant powerset functor composed with itself, denoted  $2^2$ . From a coalgebra point of view, neighbourhood structures are interesting since they constitute a general framework for studying coalgebraic modal logics in the style of Pattinson [35], where modalities are defined in terms of predicate liftings. It can easily be shown that any (unary) modality defined in this way, can be viewed as a neighbourhood modality. Furthermore, in much work on coalgebra (cf. [39]) it is often assumed that the functor preserves weak pullbacks, however, it is not always clear whether this requirement is really needed. In [19], weaker functor requirements for congruences are studied, and  $2^2$  provides an example of a functor which does not preserve weak pullbacks in general, but only the special ones consisting of kernel pairs.

From the modal logic point of view, coalgebra is interesting since it offers an abstract theory which can be instantiated to neighbourhood models, and help us generalise the well-known Kripke notions such as bisimilarity and image-finiteness to neighbourhood models. For monotonic neighbourhood structures, these questions have already been addressed (cf. [36, 20, 21]), but as mentioned in [36], if one starts from elementary intuitions, it is not immediately clear how to generalise monotonic bisimulation to arbitrary neighbourhood structures. The theory of coalgebra provides us not with one, but with several notions of state equivalence in  $F$ -coalgebras for an arbitrary functor  $F$ .  $F$ -bisimilarity and behavioural equivalence are well known concepts, and it is generally known that the two notions coincide if and only if the functor  $F$  preserves weak pullbacks [39]. This is, for example, the case over Kripke frames which are coalgebras for the covariant powerset functor  $\mathcal{P}$ , and it explains some of the fundamental properties of Kripke bisimulation: (i) Kripke bisimulations are characterised by back-and-forth conditions, which makes it possible to efficiently compute Kripke bisimilarity over finite models as a greatest fixed point. (ii) The Hennessy-Milner theorem for normal modal logic states that over the class of finite Kripke models, two states are Kripke bisimilar if and only if they satisfy the same modal formulas. (iii) Van Benthem's characterisation theorem [7, 8] tells us that Kripke bisimilarity characterises the modal fragment of first-order logic. These properties of Kripke bisimulations form the starting points of our investigation into equivalence notions in neighbourhood structures and classical modal logic.

As neighbourhood structures are coalgebras for a functor that does not preserve weak pullbacks, it is to be expected that only behavioural equivalence will give rise to a Hennessy-Milner theorem for classical modal logic. However, it turns out to be very difficult to give a back-and-forth style characterisation of behavioural equivalence. This motivates our introduction of a third equivalence notion whose witnessing relations we call pre-congruences, since they can be seen as a two-coalgebra analogue of the pre-congruences from [1].

The main contributions of this paper are: (1) the introduction of pre-congruences and basic results which relate them to bisimulations and behavioural equivalence. In particular, we show that on a single coalgebra, the largest pre-congruence is behavioural equivalence (Theorem 3.12), and that over neighbourhood models, pre-congruences are a better approximation of behavioural equivalence than  $2^2$ -bisimilarity; (2) the definition of a notion of modal saturation for neighbourhood models, which leads to a behavioural-equivalence-somewhere-else result (Theorem 4.27) by showing that ultrafilter extensions are a Hennessy-Milner class; (3) a Van Benthem style characterisation of the classical modal fragment of first-order logic (Theorem 5.5); and (4) a model-theoretic proof of Craig interpolation for classical modal logic (Theorem 5.11).

In section 2 we define basic notions and notation. In section 3, we define pre-congruences and investigate their relationship with bisimulations and behavioural equivalence. We also instantiate all three notions to the concrete case of neighbourhood frames, provide back-and-forth style characterisations for  $2^2$ -bisimulations and pre-congruences, and prove the results mentioned in (1). In section 4, we introduce our notion of modal saturation for neighbourhood models, and use it to prove a Hennessy-Milner theorem for the class of finite neighbourhood models. We then use general coalgebraic constructions to define image-finite neighbourhood models and ultrafilter extensions of neighbourhood models, and show that these are also Hennessy-Milner classes. Finally, in section 5 we prove our main results as described in (3) and (4) above. In particular, we demonstrate that  $2^2$ -bisimulations are a useful tool for proving Craig interpolation of classical modal logic.

Since neighbourhood structures are of general interest outside the world of coalgebra, we have tried to keep this paper accessible to readers who are not familiar with coalgebraic modal logic. This means that some of our results could be obtained by instantiating more general results in coalgebra. When this is the case, we give a brief explanation in the form of a remark of how the general coalgebraic framework instantiates to neighbourhood structures. However, these remarks are not necessary for understanding the main results of the paper. On the other hand, we also hope that these remarks will inspire readers to study the more general results.

## 2. PRELIMINARIES AND NOTATION

In this section, we settle on notation, define the necessary set-theoretic and coalgebraic notions, and introduce neighbourhood semantics for modal logic. For further reading on coalgebra we refer to [39, 44]. We assume the reader is familiar with the Kripke semantics and the basic model theory of normal modal logic. Some knowledge of more advanced topics such as modal saturation and ultrafilter extensions will be useful. All the necessary background information can be found in [10]. Extensive discussions on neighbourhood semantics can be found in [42, 13, 20].

**2.1. Functions and relations.** Let  $X$  and  $Y$  be sets. We denote by  $\mathcal{P}(X)$  the powerset of  $X$ , and by  $X + Y$  the disjoint union of  $X$  and  $Y$ . If  $Y \subseteq X$ , then we write  $\iota_Y$  for the inclusion map  $\iota_Y: Y \hookrightarrow X$ ;  $Y^c$  for the complement  $X \setminus Y$  of  $Y$  in  $X$ ;  $Y \subseteq_\omega X$  if  $Y$  is a finite subset of  $X$ ; and  $\uparrow Y = \{Y' \subseteq X \mid Y \subseteq Y'\}$  for the *upwards closure of  $\{Y\}$  in  $\mathcal{P}(X)$* .

For a function  $f: X \rightarrow Y$  and subsets  $U \subseteq X$  and  $V \subseteq Y$  we define the *direct  $f$ -image of  $U$*  and the  *$f$ -preimage of  $V$*  by putting  $f[U] := \{f(x) \mid x \in U\}$  and  $f^{-1}[V] := \{x \in X \mid f(x) \in V\}$ , respectively. Furthermore we call  $\text{dom}(f) := X$  the domain of  $f$  and we call  $\text{rng}(f) := f[X]$  the *range of  $f$* . More generally, we also define the notions image, preimage, domain and range for a relation  $R \subseteq X \times Y$ . For  $U \subseteq X$  and  $V \subseteq Y$ , we denote the  *$R$ -image of  $U$*  by  $R[U] = \{y \in Y \mid \exists x \in U : xRy\}$ , and the  *$R$ -preimage of  $V$*  by  $R^{-1}[V] = \{x \in X \mid \exists y \in V : xRy\}$ . The *domain of  $R$*  is  $\text{dom}(R) = R^{-1}[Y]$ , and the *range of  $R$*  is  $\text{rng}(R) = R[X]$ . We will often work with a relation in terms of its projection maps. Let  $R \subseteq X_1 \times X_2$  be a relation. The maps  $\pi_1: R \rightarrow X_1$  and  $\pi_2: R \rightarrow X_2$  denote the *projections* defined for all  $\langle x_1, x_2 \rangle \in R$  by  $\pi_i(\langle x_1, x_2 \rangle) = x_i$ , for  $i = 1, 2$ .  $R$  is called a *bitotal relation* if  $\pi_1$  and  $\pi_2$  are surjective. Note that for  $U_i \subseteq X_i$ ,  $i = 1, 2$ , we have  $R[U_1] = \pi_2[\pi_1^{-1}[U_1]]$  and  $R^{-1}[U_2] = \pi_1[\pi_2^{-1}[U_2]]$ .

If  $R \subseteq X \times X$ , then we denote by  $R^e$  the smallest equivalence relation on  $X$  which contains  $R$ , and if  $R$  is an equivalence relation on  $X$  then  $X/R$  is the set of  $R$ -equivalence classes. A relation  $R \subseteq X_1 \times X_2$ , can be viewed as a relation  $R_{X_1+X_2}$  on  $X_1 + X_2$  by composing the projections with the canonical inclusion maps  $\iota_1: X_1 \rightarrow X_1 + X_2$  and  $\iota_2: X_2 \rightarrow X_1 + X_2$ . More precisely,  $R_{X_1+X_2} = \{\langle \iota_1(x_1), \iota_2(x_2) \rangle \mid \langle x_1, x_2 \rangle \in R\}$ .

Throughout this paper the notion of coherence will be used extensively.

**Definition 2.1.** Let  $X_1$  and  $X_2$  be sets,  $R \subseteq X_1 \times X_2$  a relation,  $U_1 \subseteq X_1$  and  $U_2 \subseteq X_2$ . The pair  $\langle U_1, U_2 \rangle$  is  *$R$ -coherent* if:  $R[U_1] \subseteq U_2$  and  $R^{-1}[U_2] \subseteq U_1$ . For a set  $X$ , a relation  $R \subseteq X \times X$  and  $U \subseteq X$ , we say that  $U$  is  *$R$ -coherent*, if  $\langle U, U \rangle$  is  $R$ -coherent.

If  $R \subseteq X_1 \times X_2$ , then trivially,  $\langle \emptyset, \emptyset \rangle$  and  $\langle X_1, X_2 \rangle$  are  $R$ -coherent. Note that if  $R$  is an equivalence relation, then an  $R$ -coherent subset  $U$  is often called  *$R$ -closed*. We list a number of useful properties of  $R$ -coherence in the following two lemmas. Their easy, but instructive, proofs are left to the reader.

**Lemma 2.2.** Let  $R \subseteq X_1 \times X_2$  be a relation with projections  $\pi_i: R \rightarrow X_i$ ,  $i = 1, 2$ . For all  $U_1 \subseteq X_1$  and  $U_2 \subseteq X_2$ , the following are equivalent:

- (1)  $\langle U_1, U_2 \rangle$  is  $R$ -coherent.
- (2) for all  $\langle x_1, x_2 \rangle \in R$ :  $x_1 \in U_1 \Leftrightarrow x_2 \in U_2$ .
- (3)  $\pi_1^{-1}[U_1] = \pi_2^{-1}[U_2]$ .
- (4)  $U_1 + U_2$  is  $R_{X_1+X_2}$ -coherent.

**Lemma 2.3.** Let  $R \subseteq X \times X$  be a relation and  $U \subseteq X$ . The following are equivalent:

- (1)  $U$  is  $R$ -coherent.
- (2)  $U$  is  $R^e$ -coherent, i.e.  $R^e$ -closed.
- (3)  $U$  is a union of  $R^e$ -equivalence classes.
- (4)  $U^c$  is  $R^e$ -coherent.

**2.2. Classical modal logic and neighbourhood semantics.** Let  $\text{At} = \{p_j \mid j \in \omega\}$  be a countable set of atomic sentences. The *basic modal language over  $\text{At}$* , denoted  $\mathcal{L}(\text{At})$ , is

defined by the grammar:

$$\varphi ::= \perp \mid p_j \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi,$$

where  $j \in \omega$ . We define  $\top$ ,  $\rightarrow$  and  $\leftrightarrow$  in the usual way. We will assume  $\mathbf{At}$  to be fixed, and to ease notation, we write  $\mathcal{L}$  instead of  $\mathcal{L}(\mathbf{At})$ .

**Definition 2.4.** A *neighbourhood frame* is a pair  $\langle S, \nu \rangle$  where  $S$  is a set of states and  $\nu: S \rightarrow \mathcal{P}(\mathcal{P}(S))$  is a neighbourhood function which assigns to each state  $s \in S$  its collection of neighbourhoods  $\nu(s)$ . A *neighbourhood model* based on a neighbourhood frame  $\langle S, \nu \rangle$  is a triple  $\langle S, \nu, V \rangle$  where  $V: \mathbf{At} \rightarrow \mathcal{P}(S)$  is a valuation function.

Given a neighbourhood model  $\mathcal{M}$ , a state  $s$  in  $\mathcal{M}$  and an  $\mathcal{L}$ -formula  $\varphi$ , we write  $\mathcal{M}, s \models \varphi$  to denote that  $\varphi$  is true at  $s$  in  $\mathcal{M}$ , and  $\mathcal{M}, s \not\models \varphi$ , if  $\varphi$  is not true at  $s$  in  $\mathcal{M}$ . Truth of the atomic propositions is defined via the valuation:  $\mathcal{M}, s \models p_j$  iff  $s \in V(p_j)$ , and inductively over the boolean connectives as usual. Truth of modal formulas is given by,

$$\mathcal{M}, s \models \Box\phi \quad \text{iff} \quad \llbracket \phi \rrbracket^{\mathcal{M}} \in \nu(s), \quad (2.1)$$

where  $\llbracket \phi \rrbracket^{\mathcal{M}} = \{t \in S \mid \mathcal{M}, t \models \phi\}$  denotes the *truth set of  $\phi$  in  $\mathcal{M}$* . Let also  $\mathcal{N}$  be a neighbourhood model. Two states,  $s$  in  $\mathcal{M}$  and  $t$  in  $\mathcal{N}$ , are *modally equivalent* (notation:  $\mathcal{M}, s \equiv \mathcal{N}, t$  or simply  $s \equiv t$ ), if they satisfy the same modal  $\mathcal{L}$ -formulas, i.e.,  $s \equiv t$  if and only if for all  $\varphi \in \mathcal{L}$ :  $\mathcal{M}, s \models \varphi$  iff  $\mathcal{N}, t \models \varphi$ . A subset  $X \subseteq S$  is *modally coherent*, if for all  $s, t \in S$  such that  $s \equiv t$ :  $s \in X$  iff  $t \in X$  i.e.,  $X$  is  $\equiv$ -coherent.

Let  $\Phi \cup \{\varphi\} \subseteq \mathcal{L}$ . We write  $\Phi \models \varphi$  if  $\varphi$  is a *local semantic consequence* of  $\Phi$  over the class of all neighbourhood models, i.e., for any neighbourhood model  $\mathcal{M}$  and state  $s$  in  $\mathcal{M}$ , if  $\mathcal{M}, s \models \Phi$  then  $\mathcal{M}, s \models \varphi$ . In particular, if  $\Phi \not\models \perp$  then  $\Phi$  is called *consistent*, which means that  $\Phi$  is satisfiable in some neighbourhood model, and  $\models \varphi$  means that  $\varphi$  is valid in all neighbourhood models. We define *classical modal logic*  $\mathbf{E}$  to be the theory of neighbourhood models, that is, for all  $\mathcal{L}$ -formulas  $\varphi$ :  $\varphi \in \mathbf{E}$  iff  $\models \varphi$ . We will not be concerned with proof theory or axiomatics. For these matters, the reader is referred to [13].

The structure preserving maps between neighbourhood structures will be referred to as bounded morphisms. These have previously been studied in the context of algebraic duality [14], and monotonic neighbourhood structures (which we define in Remark 2.7 below).

**Definition 2.5.** If  $\mathcal{M}_1 = \langle S_1, \nu_1, V_1 \rangle$  and  $\mathcal{M}_2 = \langle S_2, \nu_2, V_2 \rangle$  are neighbourhood models, and  $f: S_1 \rightarrow S_2$  is a function, then  $f$  is a (*frame*) *bounded morphism from  $\langle S_1, \nu_1 \rangle$  to  $\langle S_2, \nu_2 \rangle$*  (notation:  $f: \langle S_1, \nu_1 \rangle \rightarrow \langle S_2, \nu_2 \rangle$ ) if for all  $s \in S_1$  and all  $X \subseteq S_2$ :

$$f^{-1}[X] \in \nu_1(s) \quad \text{iff} \quad X \in \nu_2(f(s)). \quad (2.2)$$

If also  $s \in V_1(p_j)$  iff  $f(s) \in V_2(p_j)$ , for all  $p_j \in \mathbf{At}$ , and all  $s \in S_1$ , then  $f$  is a *bounded morphism from  $\mathcal{M}_1$  to  $\mathcal{M}_2$*  (notation:  $f: \mathcal{M}_1 \rightarrow \mathcal{M}_2$ ).

Bounded morphisms preserve truth of modal formulas.

**Lemma 2.6.** *Let  $\mathcal{M}_1 = \langle S_1, \nu_1, V_1 \rangle$  and  $\mathcal{M}_2 = \langle S_2, \nu_2, V_2 \rangle$  be two neighbourhood models and  $f: \mathcal{M}_1 \rightarrow \mathcal{M}_2$  a bounded morphism. For each modal formula  $\varphi \in \mathcal{L}$  and state  $s \in S_1$ ,  $\mathcal{M}_1, s \models \varphi$  iff  $\mathcal{M}_2, f(s) \models \varphi$ .*

*Proof.* By a straightforward induction on the formula structure. Details left to the reader.  $\square$

Neighbourhood frames and bounded (frame) morphisms form a category which we denote by  $\text{NbhdFr}$ . Similarly, neighbourhood models and bounded morphisms form a category  $\text{Nbhd}$ . This can easily be verified directly, but it also follows from the straightforward coalgebraic modelling of neighbourhood structures which we describe now.

**2.3. Coalgebraic modelling.** We will work in the category  $\text{Set}$  of sets and functions. Let  $F: \text{Set} \rightarrow \text{Set}$  be a functor. An  $F$ -coalgebra is a pair  $\langle X, \xi \rangle$  where  $X$  is a set, and  $\xi: X \rightarrow F(X)$  is a function, sometimes called the *coalgebra map*. Given two  $F$ -coalgebras,  $\langle X_1, \xi_1 \rangle$  and  $\langle X_2, \xi_2 \rangle$ , a function  $f: X_1 \rightarrow X_2$  is an  $F$ -coalgebra morphism if  $F(f) \circ \xi_1 = \xi_2 \circ f$ , that is, the following diagram commutes:

$$\begin{array}{ccc} X_1 & \xrightarrow{f} & X_2 \\ \xi_1 \downarrow & & \downarrow \xi_2 \\ F(X_1) & \xrightarrow{F(f)} & F(X_2) \end{array}$$

The category of  $F$ -coalgebras and  $F$ -coalgebra morphisms is denoted by  $\text{Coalg}(F)$ . All notions pertaining to  $F$ -coalgebras are parametric in the functor  $F$ , but if  $F$  is clear from the context or immaterial, we will often leave it out and simply speak of coalgebras, coalgebra morphisms, and so on. Several examples of systems which can be modelled as coalgebras can be found in [39, 40].

The contravariant powerset functor  $2: \text{Set} \rightarrow \text{Set}$  maps a set  $X$  to  $\mathcal{P}(X)$ , and a function  $f: X \rightarrow Y$  to the inverse image function  $f^{-1}[-]: \mathcal{P}(Y) \rightarrow \mathcal{P}(X)$ . The functor  $2^2$  is defined as the composition of  $2$  with itself. That is, for any set  $X$  and any function  $f: X \rightarrow Y$ ,

$$\begin{aligned} 2^2(X) &= \mathcal{P}(\mathcal{P}(X)), \\ 2^2(f)(U) &= \{D \subseteq Y \mid f^{-1}[D] \in U\} \text{ for all } U \in 2^2(X). \end{aligned}$$

It should be clear that  $\text{NbhdFr}$  and  $\text{Coalg}(2^2)$  have the same objects. Similarly, given a neighbourhood model  $\langle S, \nu, V \rangle$ , we can view the valuation  $V: \text{At} \rightarrow \mathcal{P}(S)$  in its transposed form  $\hat{V}: S \rightarrow \mathcal{P}(\text{At})$  where  $p_j \in \hat{V}(s)$  iff  $s \in V(p_j)$ . It is now easy to see that  $\langle S, \nu, V \rangle$  uniquely corresponds to a coalgebra  $\langle \nu, \hat{V} \rangle: S \rightarrow 2^2(S) \times \mathcal{P}(\text{At})$  for the functor  $2^2(-) \times \mathcal{P}(\text{At})$ . Moreover, it is straightforward to show that a function  $f: S_1 \rightarrow S_2$  is a bounded morphism between the neighbourhood frames  $\mathcal{S}_1 = \langle S_1, \nu_1 \rangle$  and  $\mathcal{S}_2 = \langle S_2, \nu_2 \rangle$  iff  $f$  is a coalgebra morphism from  $\mathcal{S}_1$  to  $\mathcal{S}_2$ . Similarly,  $2^2(-) \times \mathcal{P}(\text{At})$ -coalgebra morphisms are simply the same as bounded morphisms between neighbourhood models. Hence  $\text{NbhdFr} = \text{Coalg}(2^2)$  and  $\text{Nbhd} = \text{Coalg}(2^2(-) \times \mathcal{P}(\text{At}))$ . From now on, we will switch freely between the coalgebraic setting and the neighbourhood setting.

In the course of this paper, we will relate some of our results and definitions to existing ones for monotonic modal logic and normal modal logic. We briefly remind the reader of their definitions and their relationship with neighbourhood structures and coalgebras.

**Remark 2.7.** A neighbourhood frame/model is *monotonic*, if for all  $s \in S$ , the collection of neighbourhoods  $\nu(s)$  is *upwards closed*, i.e., if  $U \subseteq V$  and  $U \in \nu(s)$  then  $V \in \nu(s)$ . Monotonic modal logic is the theory of monotonic neighbourhood models (cf. [13, 20]). It was shown in [21] that monotonic neighbourhood frames are coalgebras for the subfunctor  $\text{Mon}$  of  $2^2$  which is defined by  $\text{Mon}(X) = \{U \in \mathcal{P}(\mathcal{P}(X)) \mid U \text{ is upwards closed}\}$  on a set  $X$ .

**Remark 2.8.** It is well known that Kripke frames and their bounded morphisms can be seen as the category of coalgebras and coalgebra morphisms for the covariant powerset functor  $\mathcal{P}: \mathbf{Set} \rightarrow \mathbf{Set}$  which maps a set  $X$  to the powerset  $\mathcal{P}(X)$ , and a function  $f: X \rightarrow Y$  to the direct image function  $f[-]: \mathcal{P}(X) \rightarrow \mathcal{P}(Y)$ .

Kripke frames/models are in 1-1 correspondence with so-called augmented neighbourhood frames/models (cf. [13]). A neighbourhood frame  $\langle S, \nu \rangle$  is *augmented*, if it is monotonic and for all  $s \in S$ ,  $\bigcap \nu(s) \in \nu(s)$ . In other words, in an augmented neighbourhood frame, each neighbourhood collection is the upwards closure of a unique, smallest neighbourhood. Given a Kripke model  $\mathcal{K} = \langle S, R, V \rangle$ , we obtain an augmented neighbourhood model  $\mathcal{K}^{\text{aug}} = \langle S, \nu, V \rangle$ , by taking  $\nu(s) = \uparrow R[s]$  for all  $s \in S$ . Conversely, given an augmented neighbourhood model  $\mathcal{M} = \langle S, \nu, V \rangle$ , we define the Kripke model  $\mathcal{M}^{\text{krp}} = \langle S, R, V \rangle$  by taking  $R[s] = \bigcap \nu(s)$  for all  $s \in S$ . It should be easy to see that these transformations are inverses of each other. It is also straightforward to show that for any two Kripke models  $\mathcal{K}_1$  and  $\mathcal{K}_2$ , a function is a Kripke bounded morphism from  $\mathcal{K}_1$  to  $\mathcal{K}_2$  iff  $f$  is a (neighbourhood) bounded morphism from  $\mathcal{K}_1^{\text{aug}}$  to  $\mathcal{K}_2^{\text{aug}}$ . Hence the category of Kripke frames is isomorphic to the category of augmented neighbourhood frames. Moreover, a Kripke model  $\mathcal{K}$  and its corresponding augmented model  $\mathcal{K}^{\text{aug}}$  are pointwise equivalent, i.e., for all states  $s$  in  $\mathcal{K}$  and any  $\mathcal{L}$ -formula  $\varphi$ :  $\mathcal{K}, s \models \varphi$  iff  $\mathcal{K}^{\text{aug}}, s \models \varphi$ . This can be proved by an easy induction on  $\varphi$  (cf. [13]). Normal modal logic is the logic of all Kripke models, or equivalently, of all augmented neighbourhood models.

**2.4. Basic constructions.** Finally, we will need a number of technical constructions. Disjoint unions of neighbourhood structures lift disjoint unions of sets to neighbourhood structures such that the inclusion maps are bounded morphisms. Disjoint unions are instances of the category theoretical notion of *coproducts*, and hence they satisfy a *universal property* (which we will use in several proofs). We give the concrete definition of disjoint unions neighbourhood models and their universal property, The definition for neighbourhood frames is obtained by leaving out the part about the valuations.

**Definition 2.9.** Let  $\mathcal{M}_1 = \langle S_1, \nu_1, V_1 \rangle$  and  $\mathcal{M}_2 = \langle S_2, \nu_2, V_2 \rangle$  be two neighbourhood models. The *disjoint union of  $\mathcal{M}_1$  and  $\mathcal{M}_2$*  is the neighbourhood model  $\mathcal{M}_1 + \mathcal{M}_2 = \langle S_1 + S_2, \nu, V \rangle$  where for all  $p_j \in \text{At}$ ,  $V(p_j) = V_1(p_j) + V_2(p_j)$ ; and for  $i = 1, 2$ , for all  $X \subseteq S_1 + S_2$ , and  $s \in S_i$ :  $X \in \nu(s)$  iff  $X \cap S_i \in \nu_i(s)$ .  $\mathcal{M}_1 + \mathcal{M}_2$  has the following universal property: If  $\mathcal{N}$  is a neighbourhood model and  $f_j: \mathcal{M}_j \rightarrow \mathcal{N}$ ,  $j = 1, 2$ , are bounded morphisms, then there is a unique bounded morphism  $f: \mathcal{M}_1 + \mathcal{M}_2 \rightarrow \mathcal{N}$  such that for  $j = 1, 2$ ,  $f_j = f \circ \iota_j$ , where  $\iota_j: \mathcal{M}_j \rightarrow \mathcal{M}_1 + \mathcal{M}_2$  is the inclusion map.

In the sequel we will also use coequalisers, pushouts and pullbacks. The general definition of these notions can be found in any standard book on category theory (for example [2]). We are interested in particular instances of these notions in  $\mathbf{Set}$ , and we therefore only give the concrete definitions using the well known constructions. We also give the universal property of coequalisers and pushouts, which we will also use.

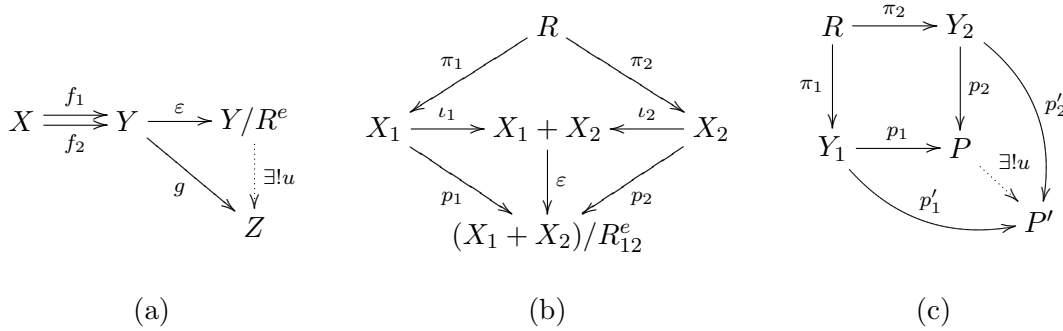
**Definition 2.10.** (coequaliser) Let  $f_1, f_2: X \rightarrow Y$  be a pair of functions. The *coequaliser of  $f_1$  and  $f_2$  in  $\mathbf{Set}$*  is the natural quotient map  $\varepsilon: Y \rightarrow Y/R^e$  where  $R = \{\langle f_1(x), f_2(x) \mid X \rangle\}$ . For any function  $g: Y \rightarrow Z$  such that  $g \circ f_1 = g \circ f_2$  there is a unique function  $u: Y/R^e \rightarrow Z$  such that  $u \circ \varepsilon = g$ . The coequaliser of a relation  $R \subseteq X \times X$  is the coequaliser of its projections  $\pi_1, \pi_2: R \rightarrow X$ .

(pushout) Let  $R \subseteq X_1 \times X_2$  be a relation with projections  $\pi_1: R \rightarrow X_1$  and  $\pi_2: R \rightarrow X_2$ . The *pushout of  $R$  in  $\mathbf{Set}$*  is the triple  $\langle P, p_1, p_2 \rangle$ , where  $P := (X_1 + X_2)/R_{12}^e$ ,  $R_{12} = R_{X_1 + X_2} = \{ \langle \iota_1(x_1), \iota_2(x_2) \rangle \mid \langle x_1, x_2 \rangle \in R \}$ ,  $\varepsilon: X_1 + X_2 \rightarrow (X_1 + X_2)/R_{12}^e$  is the coequaliser of  $\iota_1 \circ \pi_1$  and  $\iota_2 \circ \pi_2$ , and  $p_i = \varepsilon \circ \iota_i$ ,  $i \in \{1, 2\}$ . The construction is illustrated in Figure 1(b). Moreover, if  $P', p'_1: Y_1 \rightarrow P'$  and  $p'_2: Y_2 \rightarrow P'$  are such that  $p'_1 \circ \pi_1 = p'_2 \circ \pi_2$ , then there exists a unique function  $u: P \rightarrow P'$  such that  $p'_1 = u \circ p_1$  and  $p'_2 = u \circ p_2$ , as illustrated in Figure 1(c).

(pullback) Let  $f_1: X_1 \rightarrow Y$  and  $f_2: X_2 \rightarrow Y$  be functions. The *pullback of  $f_1$  and  $f_2$  in  $\mathbf{Set}$*  is the triple  $\langle \text{pb}(f_1, f_2), \pi_1, \pi_2 \rangle$ , where  $\text{pb}(f_1, f_2) := \{ \langle x_1, x_2 \rangle \in X_1 \times X_2 \mid f_1(x_1) = f_2(x_2) \}$ ; and  $\pi_1: \text{pb}(f_1, f_2) \rightarrow X_1$  and  $\pi_2: \text{pb}(f_1, f_2) \rightarrow X_2$  are the projections.

Coproducts and coequalisers are a special form of colimit. It is known that for any functor  $F: \mathbf{Set} \rightarrow \mathbf{Set}$ , all colimits exist in  $\mathbf{Coalg}(F)$  and they are constructed essentially as in  $\mathbf{Set}$ , see [39, Section 4.4]. We have already seen how this works for coproducts. For coequalisers, it means that the coequaliser of two  $F$ -coalgebra morphisms  $f_1, f_2: \langle X, \xi \rangle \rightarrow \langle Y, \gamma \rangle$  in  $\mathbf{Coalg}(F)$  is the same map  $e: Y \rightarrow Y/R^e$  which is the coequaliser of  $f_1$  and  $f_2$  in  $\mathbf{Set}$ , and there is a coalgebra structure  $\lambda: Y/R^e \rightarrow F(Y/R^e)$  such that  $e$  is an  $F$ -coalgebra morphism from  $\langle Y, \gamma \rangle$  to  $\langle Y/R^e, \lambda \rangle$ .

Figure 1: Coequalisers and Pushouts.



### 3. EQUIVALENCE NOTIONS

In this section we will study various notions of “observational equivalence” for neighbourhood frames in detail. In the first part we list the three coalgebraic equivalence notions that we are going to consider. In the second part we work out in detail what these three equivalence notions mean on neighbourhood frames.

**3.1. Three coalgebraic notions of equivalence.** The main observation for defining equivalences between coalgebras is that coalgebra morphisms preserve the behaviour of coalgebra states. This basic idea motivates the well-known coalgebraic definitions of bisimilarity and behavioural equivalence. In the following  $F$  denotes an arbitrary  $\mathbf{Set}$  functor.



**Definition 3.1.** Let  $\langle X_1, \xi_1 \rangle$  and  $\langle X_2, \xi_2 \rangle$  be F-coalgebras.

(1) A relation  $R \subseteq X_1 \times X_2$  is an (F-)bisimulation between  $\langle X_1, \xi_1 \rangle$  and  $\langle X_2, \xi_2 \rangle$ , if there exists a function  $\rho: R \rightarrow F(R)$  such that the projections  $\pi_i: R \rightarrow X_i$  are F-coalgebra morphisms from  $\langle R, \rho \rangle$  to  $\langle X_i, \xi_i \rangle$ ,  $i \in \{1, 2\}$ . Two states  $x_1$  and  $x_2$  are (F-)bisimilar (notation:  $x_1 \Leftrightarrow x_2$ ), if they are linked by some F-bisimulation. The relation  $\Leftrightarrow$  is called F-bisimilarity.

(2) Two states  $x_1 \in X_1$  and  $x_2 \in X_2$  are *behaviourally equivalent* (notation:  $x_1 \Leftrightarrow_b x_2$ ), if there exists an F-coalgebra  $\langle Y, \gamma \rangle$  and F-coalgebra morphisms  $f_i: \langle X_i, \xi_i \rangle \rightarrow \langle Y, \gamma \rangle$  for  $i = 1, 2$  such that  $f_1(x_1) = f_2(x_2)$ . The triple  $\langle \langle Y, \gamma \rangle, f_1, f_2 \rangle$  is called a *cocongruence* between  $\langle X_1, \xi_1 \rangle$  and  $\langle X_2, \xi_2 \rangle$ . If  $\langle \langle Y, \gamma \rangle, f_1, f_2 \rangle$  is a cocongruence, then we also refer to  $R = \text{pb}(f_1, f_2)$  as a cocongruence between  $\langle X_1, \xi_1 \rangle$  and  $\langle X_2, \xi_2 \rangle$ . The relation  $\Leftrightarrow_b$  is called *behavioural equivalence*.

**Remark 3.2.** Cocongruences were introduced by Kurz in [26]. In *loc.cit.*, Kurz refers to (the kernel of) an epimorphism as a behavioural equivalence. We have chosen to follow the terminology of [1, 19] and use the word congruence for kernels. We reserve behavioural equivalence to denote the equivalence notion associated with congruences and cocongruences.

For any functor F, F-bisimilarity implies behavioural equivalence (this fact will also follow from Proposition 3.10). However, the converse only holds if F preserves weak pullbacks. Precongruences were introduced in [1] as an alternative to bisimulations for functors that do not preserve weak pullbacks.

**Definition 3.3.** Let  $\langle X, \xi \rangle$  be an F-coalgebra and  $R \subseteq X \times X$  a relation.  $R$  is a *congruence* on  $\langle X, \xi \rangle$  if the coequaliser  $\varepsilon: X \rightarrow X/R$  of  $R$  is an F-coalgebra morphism, i.e., there exists a unique coalgebra structure  $\lambda: X/R \rightarrow F(X/R)$  such that  $\varepsilon$  is a coalgebra morphism from  $\langle X, \xi \rangle$  to  $\langle X/R, \lambda \rangle$ . We call  $\langle X, \xi \rangle / R := \langle X/R, \lambda \rangle$  the *quotient of  $\langle X, \xi \rangle$  with  $R$* .  $R$  is a *precongruence* on  $\langle X, \xi \rangle$  if  $R^e$  is a congruence.

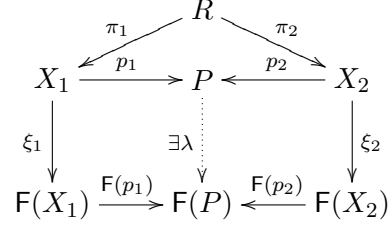
Since any F-coalgebra morphism  $f: \langle X, \xi \rangle \rightarrow \langle Y, \gamma \rangle$  factors through  $X/\ker(f)$ , it follows that  $R$  is a congruence on  $\langle X, \xi \rangle$  iff  $R = \ker(f) = \text{pb}(f, f)$  for some F-coalgebra morphism  $f: \langle X, \xi \rangle \rightarrow \langle X', \xi' \rangle$ .

**Lemma 3.4.** *Let  $\langle X, \xi \rangle$  be an F-coalgebra. Behavioural equivalence, the largest congruence and the largest precongruence on  $\langle X, \xi \rangle$  all coincide.*

*Proof.* The lemma follows from results in [1] and [19, Lemma 5.10], but we also provide a quick argument here. Clearly, a congruence is also a precongruence and a precongruence is contained in a congruence. Hence the largest congruence is the largest precongruence. We refer to [1] for more details. Similarly, a congruence is clearly a cocongruence, and any cocongruence is contained in a congruence, since the category of F-coalgebras has coequalisers: if  $R = \text{pb}(f_1, f_2)$  for F-coalgebra morphisms  $f_1, f_2: X \rightarrow Y$ , then  $R \subseteq \ker(e \circ f_1)$ , where  $e$  is the coequaliser of  $f_1$  and  $f_2$ . See also [19, Lemma 5.10]. Hence the largest congruence is behavioural equivalence  $\square$

Precocongruences can be seen as a generalisation of precongruences to relations between coalgebras obtained by replacing coequalisers by pushouts.

**Definition 3.5.** Let  $\langle X_1, \xi_1 \rangle$  and  $\langle X_2, \xi_2 \rangle$  be  $F$ -coalgebras, and let  $R \subseteq X_1 \times X_2$  be a relation with pushout  $\langle P, p_1, p_2 \rangle$ . The relation  $R$  is called a *precongruence* between  $\langle X_1, \xi_1 \rangle$  and  $\langle X_2, \xi_2 \rangle$ , if there exists a coalgebra map  $\lambda: P \rightarrow F(P)$  such that the pushout maps  $p_1: X_1 \rightarrow P$  and  $p_2: X_2 \rightarrow P$  are  $F$ -coalgebra morphisms, i.e., the diagram on right commutes. In other words,  $R$  is a precongruence if and only if its pushout  $\langle P, p_1, p_2 \rangle$  is a congruence. If two states  $x_1$  and  $x_2$  are related by some precongruence, we write  $x_1 \leftrightarrow_p x_2$ .



The following lemma tells us that we can think of precongruences as the relations that are precongruences on the coproduct (disjoint union), and it provides a useful criterion for proving that a relation is a precongruence.

**Lemma 3.6.** Let  $\langle X_1, \xi_1 \rangle$  and  $\langle X_2, \xi_2 \rangle$  be  $F$ -coalgebras, and let  $R \subseteq X_1 \times X_2$  be a relation with pushout  $\langle P, p_1, p_2 \rangle$ . The following are equivalent:

- (1)  $R$  is a precongruence between  $\langle X_1, \xi_1 \rangle$  and  $\langle X_2, \xi_2 \rangle$ .
- (2)  $F(p_1) \circ \xi_1 \circ \pi_1 = F(p_2) \circ \xi_2 \circ \pi_2$ , i.e.,  $R \subseteq \text{pb}(F(p_1) \circ \xi_1, F(p_2) \circ \xi_2)$ .
- (3)  $R_{X_1+X_2}$  is a precongruence on  $\langle X_1, \xi_1 \rangle + \langle X_2, \xi_2 \rangle$ .

*Proof.* (1  $\Leftrightarrow$  2): Item 2 holds iff the outer part of the diagram in Def. 3.5 commutes, so the implication (1  $\Rightarrow$  2) is immediate. Conversely, if item 2 holds, then by the universal property of the pushout  $\langle P, p_1, p_2 \rangle$  there is a (unique) function  $\lambda: P \rightarrow F(P)$  such that  $\lambda \circ p_1 = F(p_1) \circ \xi_1$  and  $\lambda \circ p_2 = F(p_2) \circ \xi_2$ . Hence  $R$  is a precongruence,

(1  $\Rightarrow$  3): If the pushout maps are morphisms, there exists by the universal property of the disjoint union  $\langle X_1, \xi_1 \rangle + \langle X_2, \xi_2 \rangle$  in  $\text{Coalg}(F)$ , a unique  $F$ -coalgebra morphism  $u: X_1 + X_2 \rightarrow P$  such that  $p_i = u \circ \iota_i$ ,  $i \in \{1, 2\}$ . By the definition of the pushout (cf. Figure 1(b)), it must be the case that  $u$  is equal to the natural quotient map  $\varepsilon: X_1 + X_2 \rightarrow P$ , and hence  $R_{X_1+X_2}$  is a precongruence.

(3  $\Rightarrow$  1): If  $R_{X_1+X_2}$  is a precongruence on the disjoint union, then the quotient map  $\varepsilon: X_1 + X_2 \rightarrow (X_1 + X_2)/R_{X_1+X_2}^e$  is an  $F$ -coalgebra morphism. Since  $p_i = \varepsilon \circ \iota_i$ ,  $i \in \{1, 2\}$ , and the canonical inclusions  $\iota_i: X_i \rightarrow X_1 + X_2$ ,  $i \in \{1, 2\}$ , are also  $F$ -coalgebra morphisms, it follows that the pushout maps are  $F$ -coalgebra morphisms.  $\square$

An interesting property of precongruences, is that, like bisimulations, they can be characterised by a form of relation lifting.

**Definition 3.7.** Let  $R \subseteq X_1 \times X_2$  be a relation and let  $\langle P, p_1, p_2 \rangle$  be the pushout of  $\langle R, \pi_1, \pi_2 \rangle$ . We define the  $F$ -lifting  $Lif(F)(R) \subseteq F(X_1) \times F(X_2)$  of  $R$  by

$$Lif(F)(R) := \text{pb}(F(p_1), F(p_2)).$$

Note that  $Lif(F)$  is independent of the concrete representation of the pushout. This follows easily from the fact that pushouts are unique up-to isomorphism. The definition of  $Lif(F)$  goes back to an idea by Kurz ([25]) for defining a relation lifting of functors that do not preserve weak pullbacks.

**Lemma 3.8.** Let  $\langle X_1, \xi_1 \rangle$  and  $\langle X_2, \xi_2 \rangle$  be  $F$ -coalgebras, and let  $R \subseteq X_1 \times X_2$  be a relation.  $R$  is precongruence iff for all  $\langle x_1, x_2 \rangle \in R$ :  $\langle \xi_1(x_1), \xi_2(x_2) \rangle \in Lif(F)(R)$ .

*Proof.* Immediate from Lemma 3.6 and the definition of  $Lif(F)$ .  $\square$

The characterisation of pre-cocongruences in Lemma 3.8 makes it easy to show that between any two coalgebras, there exists a largest, and necessarily unique, pre-cocongruence. First, note that for any relations  $R' \subseteq R \subseteq X_1 \times X_2$  with pushouts  $\langle P', p'_1, p'_2 \rangle$  and  $\langle P, p_1, p_2 \rangle$ , respectively, there exists by the universal property of  $P'$  a unique map  $u: P' \rightarrow P$  such that  $p_i = u \circ p'_i$ ,  $i \in \{1, 2\}$ . Consequently,  $F(p_i) = F(u) \circ F(p'_i)$ ,  $i \in \{1, 2\}$ , and for all  $t_1 \in F(X_1)$ ,  $t_2 \in F(X_2)$ :  $F(p'_1)(t_1) = F(p'_2)(t_2)$  implies that  $F(p_1)(t_1) = F(p_2)(t_2)$ . Hence,

$$R' \subseteq R \quad \Rightarrow \quad \text{Lif}(F)(R') \subseteq \text{Lif}(F)(R). \quad (3.1)$$

**Lemma 3.9.** *Let  $\langle X_1, \xi_1 \rangle$  and  $\langle X_2, \xi_2 \rangle$  be F-coalgebras. The union of all pre-cocongruences between  $\langle X_1, \xi_1 \rangle$  and  $\langle X_2, \xi_2 \rangle$  is again a pre-cocongruence.*

*Proof.* Let  $R$  be the union of all pre-cocongruences between  $\langle X_1, \xi_1 \rangle$  and  $\langle X_2, \xi_2 \rangle$ , and  $\langle P, p_1, p_2 \rangle$  the pushout of  $R$ . If  $\langle x_1, x_2 \rangle \in R$ , then there is a pre-cocongruence  $R' \subseteq R$  such that  $\langle x_1, x_2 \rangle \in R'$ . Letting  $\langle P', p'_1, p'_2 \rangle$  be the pushout of  $R'$ , it follows that  $\langle \xi_1(x_1), \xi_2(x_2) \rangle \in \text{Lif}(F)(R')$ , and hence by (3.1) that  $\langle \xi_1(x_1), \xi_2(x_2) \rangle \in \text{Lif}(F)(R)$ . We conclude by Lemma 3.8 that  $R$  is a pre-cocongruence.  $\square$

In the following proposition we give a first comparison between pre-cocongruences, bisimulations and cocongruences.

**Proposition 3.10.** *Let  $\langle X_1, \xi_1 \rangle$  and  $\langle X_2, \xi_2 \rangle$  be F-coalgebras, and let  $R$  be a relation between  $X_1$  and  $X_2$ .*

- (1) *If  $R$  is a bisimulation, then  $R$  is a pre-cocongruence.*
- (2) *If  $R$  is a pre-cocongruence, then  $R$  is contained in a cocongruence.*

*Consequently, for all  $x_1 \in X_1$  and  $x_2 \in X_2$ :*

$$x_1 \stackrel{\text{b}}{\sim} x_2 \text{ implies } x_1 \stackrel{\text{p}}{\sim} x_2 \text{ implies } x_1 \stackrel{\text{c}}{\sim} x_2.$$

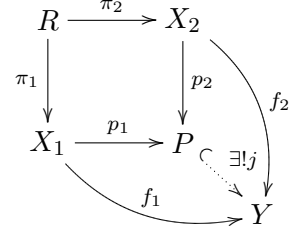
*Proof.* Let  $R \subseteq X_1 \times X_2$  be a relation with projections  $\pi_1: R \rightarrow X_1$  and  $\pi_2: R \rightarrow X_2$ , and pushout  $\langle P, p_1, p_2 \rangle$ . Item 1: Assume  $R$  is a bisimulation. By composing the projections with the canonical inclusion morphisms into the coproduct, we have a pair of parallel F-coalgebra morphisms  $\iota_1 \circ \pi_1, \iota_2 \circ \pi_2: R \rightarrow X_1 + X_2$ . The quotient map  $\varepsilon: X_1 + X_2 \rightarrow (X_1 + X_2)/R_{X_1 + X_2}^e$  is now the coequaliser of  $\iota_1 \circ \pi_1$  and  $\iota_2 \circ \pi_2$  in  $\text{Coalg}(F)$ , hence an F-coalgebra morphism. Since  $p_i = \varepsilon \circ \iota_i$ ,  $i = 1, 2$ ,  $p_1$  and  $p_2$  are F-coalgebra morphisms. Item 2: If  $R$  is a pre-cocongruence, then the pushout maps  $p_1$  and  $p_2$  are F-coalgebra morphisms. The claim now follows from the fact that  $R \subseteq \text{pb}(p_1, p_2)$ .  $\square$

Proposition 3.10 alone does not yet tell us whether pre-cocongruences are a better approximation of behavioural equivalence than F-bisimulations, but in the next subsection, we will see that, in general, the implications of Proposition 3.10 are strict. The following lemma provides us with a criterion which ensures that a cocongruence is a pre-cocongruence.

**Lemma 3.11.** *If  $\langle X_1, \xi_1 \rangle$  and  $\langle X_2, \xi_2 \rangle$  are F-coalgebras and  $R \subseteq X_1 \times X_2$  is a bitotal cocongruence between  $\langle X_1, \xi_1 \rangle$  and  $\langle X_2, \xi_2 \rangle$ , then  $R$  is a pre-cocongruence.*

*Proof.* Let  $R$  be a cocongruence with projection maps  $\pi_1: R \rightarrow X_1$  and  $\pi_2: R \rightarrow X_2$  and pushout  $\langle P, p_1, p_2 \rangle$ . Then there exist an F-coalgebra  $\langle Y, \gamma \rangle$  and F-coalgebra morphisms  $f_i: X_i \rightarrow Y$  for  $i \in \{1, 2\}$  such that  $R = \text{pb}(f_1, f_2)$ . We are going to define a function  $\lambda: P \rightarrow F(P)$  such that  $p_i$  is an F-coalgebra morphism from  $\langle X_i, \xi_i \rangle$  to  $\langle P, \lambda \rangle$  for  $i \in \{1, 2\}$ .

By the universal property of the pushout there has to be a function  $j: P \rightarrow Y$  such that  $j \circ p_i = f_i$  for  $i \in \{1, 2\}$ , as shown in the diagram to the right. We claim that this function is injective. First, it follows from the definition of the pushout that both  $p_1$  and  $p_2$  are surjective, because  $R$  is bitotal. Let now  $z_1, z_2 \in P$  and suppose that  $j(z_1) = j(z_2)$ . The surjectivity of the  $p_i$ 's implies that there are  $s_1 \in X_1$  and  $s_2 \in X_2$  such that  $p_1(s_1) = z_1$  and  $p_2(s_2) = z_2$ . Hence  $j(p_1(s_1)) = j(p_2(s_2))$  which in turn yields  $f_1(s_1) = f_2(s_2)$ . This implies that  $\langle s_1, s_2 \rangle \in R$  and consequently,  $p_1(s_1) = p_2(s_2)$ , i.e.,  $z_1 = z_2$ . This demonstrates that  $j$  is injective and thus there is some surjective map  $e: Y \rightarrow P$  with  $e \circ j = \text{id}_P$ . Now define  $\lambda := F(e) \circ \lambda \circ j$ . It is straightforward to check that for  $i \in \{1, 2\}$ , the function  $p_i: \langle X_i, \xi_i \rangle \rightarrow \langle P, \lambda \rangle$  is an F-coalgebra morphism.  $\square$



We will now show that on a single F-coalgebra, an equivalence relation is a precocongruence iff it is a congruence. It then follows immediately that the largest congruence is a precocongruence.

**Theorem 3.12.** *Let  $\langle X, \xi \rangle$  an F-coalgebra.*

- (1) *If  $R \subseteq X \times X$  is an equivalence relation then:  $R$  is a precocongruence on  $\langle X, \xi \rangle$  iff  $R$  is a congruence on  $\langle X, \xi \rangle$ .*
- (2) *For all  $x_1, x_2 \in X$ :  $x_1 \leftrightarrow_b x_2$  iff  $x_1 \leftrightarrow_p x_2$ .*

*Proof.* To prove item 1, first, observe that if  $R \subseteq X \times X$  is an equivalence relation, then  $\langle x, x \rangle \in R$  for all  $x \in X$ , hence  $p_1(x) = p_2(x)$  for all  $x \in X$ , i.e.,  $p_1 = p_2$ . It follows that the pushout of  $R$  is of the form  $\langle P, p, p \rangle$  and  $R = \ker(p)$ . Hence if  $R$  is also a precocongruence, then  $p$  is a coalgebra morphism and  $R = \ker(p)$  is a congruence. Conversely, if  $R$  is a congruence, then  $R$  is clearly a bitotal cocongruence on  $\langle X, \xi \rangle$  and so by Lemma 3.11, a precocongruence. Item 2 of the lemma follows from item 1 and Lemma 3.4.  $\square$

We have introduced precocongruences as a generalisation of precongruences to relations between different coalgebras. However, we point out that this generalisation is conceptual rather than set-theoretic, since on a single coalgebra, a precongruence is not necessarily a precocongruence (as we will see in Example 3.18 below). In fact, one might say that precocongruences specialise precongruences in the one-coalgebra case, since the converse does hold.

**Lemma 3.13.** *Let  $\langle X, \xi \rangle$  be an F-coalgebra and  $R \subseteq X \times X$ . If  $R$  is a precocongruence on  $\langle X, \xi \rangle$ , then  $R$  is also a precongruence on  $\langle X, \xi \rangle$ .*

*Proof.* Let  $\langle P, p_1, p_2 \rangle$  be the pushout of  $R$ , and let  $\varepsilon_R: X \rightarrow X/R^e$  be the natural quotient map (i.e., the coequaliser of  $R$ ). By the universal property of the pushout in **Set**, there is a unique map  $u: P \rightarrow X/R^e$  such that  $u \circ p_1 = \varepsilon_R = u \circ p_2$ . It follows that  $F(u) \circ F(p_1) = F(\varepsilon_R) = F(u) \circ F(p_2)$ , and hence for all  $x, y \in X$ :  $F(p_1)(\xi(x)) = F(p_2)(\xi(y))$  implies that  $F(\varepsilon_R)(\xi(x)) = F(\varepsilon_R)(\xi(y))$ . Consequently, using Lemma 3.6(2) and the fact that  $R$  is a precongruence iff  $R \subseteq \ker(F(\varepsilon_R) \circ \xi)$  (this can easily be shown using the universal property of coequalisers, see also [1]), we conclude that if  $R$  is a precocongruence, then  $R$  is also a precongruence.  $\square$

**3.2. Equivalences between neighbourhood frames.** In this subsection, we will investigate behavioural equivalence, bisimilarity and the equivalence notion arising from pre-congruences over  $2^2$ -coalgebras, i.e., neighbourhood frames. First, we obtain set-theoretic, back-and-forth style predicates for  $2^2$ -bisimulations and  $2^2$ -precongruences. Next, we provide examples which show that the implications from Proposition 3.10 are strict. However, we also show that on a single neighbourhood frame all three equivalence notions coincide. Finally, we compare the three equivalence notions with bisimulations over monotonic neighbourhood frames and Kripke frames.

**Remark 3.14.** For simplicity of presentation, we have chosen to only treat equivalence notions on neighbourhood frames, but the results of this section can easily be extended to neighbourhood *models*, i.e.,  $2^2(-) \times \mathcal{P}(\text{At})$ -coalgebras. For example, working out the details of the definition of  $2^2(-) \times \mathcal{P}(\text{At})$ -bisimulation results in the expected characterisation: A relation  $R$  is  $2^2(-) \times \mathcal{P}(\text{At})$ -bisimulation and if and only if  $R$  is a  $2^2$ -bisimulation and for all  $\langle s, t \rangle \in R$ ,  $s$  and  $t$  satisfy the same atomic propositions. Similar statements hold for cocongruences and precongruences.

Let us start out by considering  $2^2$ -bisimulations. Recall from Def. 3.1 that a relation  $R \subseteq S_1 \times S_2$  is a  $2^2$ -bisimulation between two  $2^2$ -coalgebras  $\mathcal{S}_1 = \langle S_1, \nu_1 \rangle$  and  $\mathcal{S}_2 = \langle S_2, \nu_2 \rangle$  if the projection maps  $\pi_1$  and  $\pi_2$  are bounded morphisms ( $2^2$ -coalgebra morphisms) from some  $2^2$ -coalgebra  $\langle R, \mu \rangle$  to  $\mathcal{S}_1$  and  $\mathcal{S}_2$  respectively. By Definition 2.5 of a bounded morphism this means that for  $\langle s_1, s_2 \rangle \in R$  and  $i = 1, 2$ :

$$U \in \nu_i(s_i) \quad \text{iff} \quad \pi_i^{-1}[U] \in \mu(\langle s_1, s_2 \rangle) \quad \text{for } U \subseteq S_i.$$

This leads to two “minimal requirements” on the neighbourhood functions  $\nu_1$  and  $\nu_2$  for pairs  $\langle s_1, s_2 \rangle$  contained in a  $2^2$ -bisimulation. For all  $U_i, U'_i \subseteq S_i$ ,  $i = 1, 2$ :

- (1)  $\pi_i^{-1}[U_i] = \pi_i^{-1}[U'_i]$  implies  $U_i \in \nu_i(s_i)$  iff  $U'_i \in \nu_i(s_i)$ ,
- (2)  $\pi_1^{-1}[U_1] = \pi_2^{-1}[U_2]$  implies  $U_1 \in \nu_1(s_1)$  iff  $U_2 \in \nu_2(s_2)$ .

Using the notion of  $R$ -coherence we can reformulate the previous requirements and prove that they in fact characterise  $2^2$ -bisimulations.

**Proposition 3.15.** *Let  $\mathcal{S}_1 = \langle S_1, \nu_1 \rangle$  and  $\mathcal{S}_2 = \langle S_2, \nu_2 \rangle$  be neighbourhood frames. A relation  $R \subseteq S_1 \times S_2$  is a  $2^2$ -bisimulation between  $\mathcal{S}_1$  and  $\mathcal{S}_2$  iff for all  $\langle s_1, s_2 \rangle \in R$ , for all  $U_1, U'_1 \subseteq S_1$  and for all  $U_2, U'_2 \subseteq S_2$  the following two conditions are satisfied:*

- (1) (a) if  $\text{dom}(R) \cap U_1 = \text{dom}(R) \cap U'_1$  then  $U_1 \in \nu_1(s_1)$  iff  $U'_1 \in \nu_1(s_1)$ , and  
 (b) if  $\text{rng}(R) \cap U_2 = \text{rng}(R) \cap U'_2$  then  $U_2 \in \nu_2(s_2)$  iff  $U'_2 \in \nu_2(s_2)$ .
- (2) if the pair  $\langle U_1, U_2 \rangle$  is  $R$ -coherent, then:  $U_1 \in \nu_1(s_1)$  iff  $U_2 \in \nu_2(s_2)$ .

*Proof.* It is a matter of routine checking that every  $2^2$ -bisimulation  $R$  fulfills conditions 1 and 2. Let now  $R \subseteq S_1 \times S_2$  be a relation that fulfills the conditions 1 and 2 for all  $\langle s_1, s_2 \rangle \in R$ . We define the neighbourhood function  $\mu: R \rightarrow 2^2(R)$  by  $\mu(\langle s_1, s_2 \rangle) := \{\pi_1^{-1}[U] \mid U \in \nu_1(s_1)\} \cup \{\pi_2^{-1}[V] \mid V \in \nu_2(s_2)\}$ . In order to show that  $R$  is a  $2^2$ -bisimulation it suffices to prove that for  $i = 1, 2$  the projection functions  $\pi_i: \langle R, \mu \rangle \rightarrow \mathcal{S}_i$  are bounded morphisms. We only provide the details for the proof that  $\pi_1$  is a bounded morphism. We have to demonstrate that for all  $\langle s_1, s_2 \rangle \in R$  and all  $U \subseteq S_1$  we have

$$U \in \nu_1(s_1) \quad \text{iff} \quad \pi_1^{-1}[U] \in \mu(\langle s_1, s_2 \rangle). \quad (3.2)$$

Let  $\langle s_1, s_2 \rangle \in R$  and  $U \subseteq S_1$ . By definition of  $\mu(\langle s_1, s_2 \rangle)$  the direction from left to right in (3.2) is immediate. In order to prove the other implication in (3.2) suppose that  $\pi_1^{-1}[U] \in$

$\mu(\langle s_1, s_2 \rangle)$  for some  $U \subseteq S_1$ . According to the definition of  $\mu(\langle s_1, s_2 \rangle)$  the following cases can occur:

**Case:**  $\pi_1^{-1}[U] = \pi_1^{-1}[U']$  for some  $U' \in \nu_1(s_1)$ . Then  $\text{dom}(R) \cap U = \text{dom}(R) \cap U'$  and hence  $U$  must be also in  $\nu_1(s_1)$  by condition 1 of the proposition.

**Case:**  $\pi_1^{-1}[U] = \pi_2^{-1}[V]$  for some  $V \in \nu_2(s_1)$ , i.e., the pair  $\langle U, V \rangle$  is  $R$ -coherent. Condition 2 therefore yields  $U \in \nu_1(s_2)$  as required.  $\square$

Another way of formulating condition 1a in Proposition 3.15, is to say that if  $U_1 \in \nu_1(s_1)$  and  $U'_1 \notin \nu_1(s_1)$ , then there is a  $u \in (U_1 \setminus U'_1) \cup (U'_1 \setminus U_1)$  such that  $u \in \text{dom}(R)$ . Similarly for condition 1b. Informally, one can say that condition 1 requires that the relation  $R$  must witness the difference between subsets when one is a neighbourhood and the other is not. We will now show that precocongruences are characterised by condition 2 only, hence condition 1 is unnecessary (unwanted even) for the purpose of approximating behavioural equivalence.

Let  $\langle S_1, \nu_1 \rangle$  and  $\langle S_2, \nu_2 \rangle$  be two  $2^2$ -coalgebras and  $R \subseteq S_1 \times S_2$  a relation with pushout  $\langle P, p_1, p_2 \rangle$ . We have:

$$\begin{aligned} & R \text{ is a precocongruence} \\ \text{iff } & \forall \langle s_1, s_2 \rangle \in R : 2^2(p_1)(\nu_1(s_1)) = 2^2(p_2)(\nu_2(s_2)) \\ \text{iff } & \forall \langle s_1, s_2 \rangle \in R . \forall V \subseteq P : p_1^{-1}[V] \in \nu_1(s_1) \Leftrightarrow p_2^{-1}[V] \in \nu_2(s_2) \end{aligned} \quad (3.3)$$

We now show that, in fact, (3.3) is equivalent with condition 2 of Proposition 3.15.

**Proposition 3.16.** *Let  $\mathcal{S}_1 = \langle S_1, \nu_1 \rangle$  and  $\mathcal{S}_2 = \langle S_2, \nu_2 \rangle$  be neighbourhood frames, and  $R \subseteq S_1 \times S_2$  a relation. We have:  $R$  is a precocongruence between  $\mathcal{S}_1$  and  $\mathcal{S}_2$  if and only if for all  $\langle s_1, s_2 \rangle \in R$  and for all  $U_1 \subseteq S_1$  and  $U_2 \subseteq S_2$  such that  $\langle U_1, U_2 \rangle$  is  $R$ -coherent:  $U_1 \in \nu_1(s_1)$  iff  $U_2 \in \nu_2(s_2)$ .*

*Proof.* Let  $\mathcal{S}_1, \mathcal{S}_2$  and  $R$  be as stated. Furthermore, let  $\pi_i: R \rightarrow S_i$ ,  $i \in \{1, 2\}$ , be the projections of  $R$ ,  $R_{12} = R_{S_1 + S_2}$ , and  $\langle P, p_1, p_2 \rangle$  the pushout of  $R$ . We will prove that for all  $U_1 \subseteq S_1$  and  $U_2 \subseteq S_2$ :

$$\langle U_1, U_2 \rangle \text{ is } R\text{-coherent} \quad \text{iff} \quad U_1 = p_1^{-1}[Y] \text{ and } U_2 = p_2^{-1}[Y] \text{ for some } Y \subseteq P. \quad (3.4)$$

The proposition then follows from (3.3) and (3.4). To prove the direction from left to right in (3.4), assume  $U_1 \subseteq S_1$ ,  $U_2 \subseteq S_2$  and  $\langle U_1, U_2 \rangle$  is  $R$ -coherent. From Lemmas 2.2 and 2.3, we get that  $U_1 + U_2$  is  $R_{12}^e$ -coherent. Let  $\varepsilon: S_1 + S_2 \rightarrow P$  be the quotient map associated with  $R_{12}^e$ . We claim that we can take  $Y = \varepsilon[U_1 + U_2]$ , the set of  $R_{12}^e$ -equivalence classes intersecting  $U_1 + U_2$ . To see that  $p_1^{-1}[\varepsilon[U_1 + U_2]] = U_1$  and  $p_2^{-1}[\varepsilon[U_1 + U_2]] = U_2$ , we have for all  $i \in \{1, 2\}$  and  $s_i \in S_i$ :

$$\begin{aligned} s_i \in p_i^{-1}[\varepsilon[U_1 + U_2]] & \iff p_i(s_i) \in \varepsilon[U_1 + U_2] \\ & \iff \exists s' \in U_1 + U_2 : \langle s_i, s' \rangle \in R_{12}^e \\ (U_1 + U_2 \text{ } R_{12}^e\text{-coh.}) & \iff s_i \in U_1 + U_2 \\ & \iff s_i \in U_i. \end{aligned}$$

To prove the direction from right to left in (3.4), let  $Y \subseteq P$  be arbitrary. We have for all  $\langle s_1, s_2 \rangle \in R$ :

$$\langle s_1, s_2 \rangle \in \pi_1^{-1}[p_1^{-1}[Y]] \quad \text{iff} \quad p_1(s_1) \in Y \quad \text{iff} \quad p_2(s_2) \in Y \quad \text{iff} \quad \langle s_1, s_2 \rangle \in \pi_2^{-1}[p_2^{-1}[Y]].$$

where the middle equivalence follows from the fact that  $\langle s_1, s_2 \rangle \in R$  implies  $p_1(s_1) = p_2(s_2)$ . We have now shown that  $\pi_1^{-1}[p_1^{-1}[Y]] = \pi_2^{-1}[p_2^{-1}[Y]]$ , hence by Lemma 2.2, the pair  $\langle p_1^{-1}[Y], p_2^{-1}[Y] \rangle$  is  $R$ -coherent.  $\square$

Since we know that on a single coalgebra, congruences are precocongruences (Theorem 3.12), we get the following characterisation.

**Corollary 3.17.** *Let  $\langle S, \nu \rangle$  be a neighbourhood frame and  $R \subseteq S \times S$  an equivalence relation. We have:  $R$  is a congruence on  $\langle S, \nu \rangle$  iff*

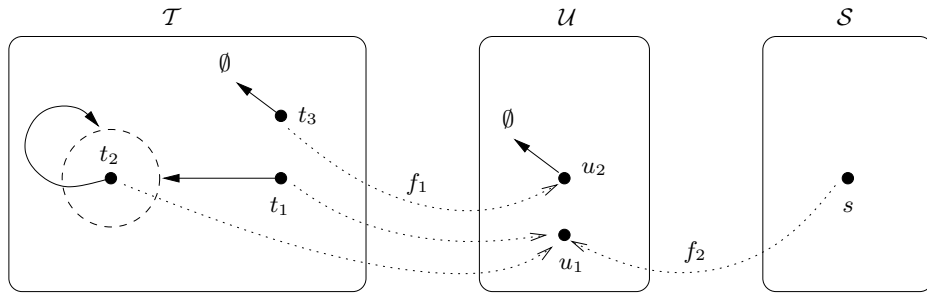
$$\text{for all } \langle s_1, s_2 \rangle \in R \text{ and all } R\text{-coherent } U \subseteq S: U \in \nu(s_1) \text{ iff } U \in \nu(s_2). \quad (3.5)$$

*Proof.* Let  $R \subseteq S \times S$  be an equivalence relation. We first prove a small claim: *Claim:* A pair  $\langle U_1, U_2 \rangle$  is  $R$ -coherent iff  $U_1 = U_2 = U$  for some  $R$ -coherent subset  $U \subseteq S$ . *Proof of Claim:* Recall that a pair  $\langle U_1, U_2 \rangle$  is  $R$ -coherent iff  $R[U_1] \subseteq U_2$  and  $R^{-1}[U_2] \subseteq U_1$ . Since  $R$  is an equivalence relation,  $R$  is reflexive, and it follows that if  $\langle U_1, U_2 \rangle$  is  $R$ -coherent, then  $U_1 \subseteq R[U_1] \subseteq U_2$  and  $U_2 \subseteq R^{-1}[U_2] \subseteq U_1$ , hence  $U_1 = U_2$ . Conversely, if  $U$  is some  $R$ -coherent subset of  $S$ , then by definition,  $\langle U, U \rangle$  is  $R$ -coherent.

We now have:  $R$  is a congruence iff (Thm. 3.12)  $R$  is a precocongruence iff (Prop. 3.16) for all  $\langle s_1, s_2 \rangle \in R$  and for all  $U_1, U_2 \subseteq S$  such that  $\langle U_1, U_2 \rangle$  is  $R$ -coherent:  $U_1 \in \nu(s_1)$  iff  $U_2 \in \nu(s_2)$ . Using the above claim, this last statement is equivalent with (3.5).  $\square$

We will now demonstrate with two examples that  $2^2$ -bisimilarity, precocongruences and behavioural equivalence differ on neighbourhood frames. It is tempting to think of the elements of neighbourhoods as successor states, but these examples show that this leads to wrong intuitions. For example, contrary to the intuition we have from Kripke bisimulations, behavioural equivalence in neighbourhood frames does not require that nonempty neighbourhoods are somehow matched by nonempty neighbourhoods. Moreover, states that are not contained in any neighbourhood of some state  $s$ , can influence the existence of a bisimulation or cocongruence at  $s$ .

**Example 3.18.** Consider the two neighbourhood frames,  $\mathcal{T} = \langle T, \nu_{\mathcal{T}} \rangle$  and  $\mathcal{S} = \langle S, \nu_{\mathcal{S}} \rangle$  where  $T = \{t_1, t_2, t_3\}$ ,  $\nu_{\mathcal{T}}(t_1) = \nu_{\mathcal{T}}(t_2) = \{\{t_2\}\}$ ,  $\nu_{\mathcal{T}}(t_3) := \{\emptyset\}$ , and  $S = \{s\}$ ,  $\nu_{\mathcal{S}}(s) = \emptyset$ . The two states  $t_1$  and  $s$  are behaviourally equivalent. To see this, let  $\mathcal{U} = \langle U, \nu_{\mathcal{U}} \rangle$  be the neighbourhood frame where  $U = \{u_1, u_2\}$ ,  $\nu_{\mathcal{U}}(u_1) = \emptyset$  and  $\nu_{\mathcal{U}}(u_2) = \{\emptyset\}$ . Let  $f_1: \mathcal{T} \rightarrow \mathcal{U}$  and  $f_2: \mathcal{S} \rightarrow \mathcal{U}$  be the functions with graphs  $\text{Gr}(f_1) = \{\langle t_1, u_1 \rangle, \langle t_2, u_1 \rangle, \langle t_3, u_2 \rangle\}$  and  $\text{Gr}(f_2) = \{\langle s, u_1 \rangle\}$ , respectively, as illustrated in the following picture:



It can easily be verified that  $f_1$  and  $f_2$  are bounded morphisms. For example, the bounded morphism condition (2.2) holds for  $f_1$  at  $t_1$  and  $t_2$ , since their only neighbourhood  $\{t_2\}$  is not the inverse  $f_1$ -image of any subset of  $U$ . Since  $f_1(t_1) = f_2(s)$ ,  $t_1$  and  $s$  are behaviourally equivalent. In fact,  $R := \text{pb}(f_1, f_2) = \{\langle t_1, s \rangle, \langle t_2, s \rangle\}$  is a precocongruence.

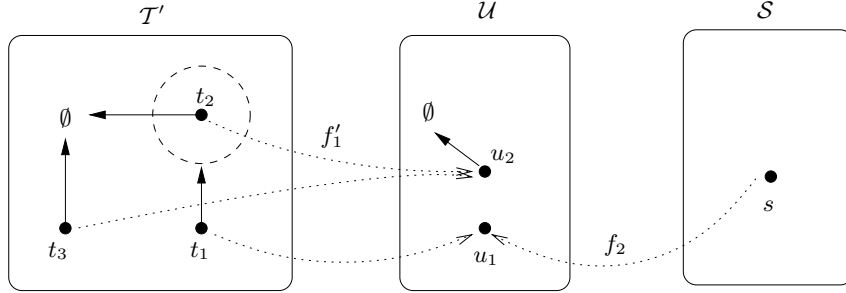
This can be verified using the characterisation given in Proposition 3.16. Note that there is no subset  $U \subseteq S$  such that  $\langle \{t_2\}, U \rangle$  is  $R$ -coherent.

However,  $t_1$  and  $s$  are not  $2^2$ -bisimilar. For suppose  $R$  is a  $2^2$ -bisimulation between  $\mathcal{T}$  and  $\mathcal{S}$ , then  $\langle t_3, s \rangle \notin R$ , since  $\langle \emptyset, \emptyset \rangle$  is  $R$ -coherent,  $\emptyset \in \nu_{\mathcal{T}}(t_3)$  and  $\emptyset \notin \nu_{\mathcal{S}}(s)$ . Hence  $t_3 \notin \text{dom}(R)$ , and it follows that  $\text{dom}(R) \cap \{t_2\} = \text{dom}(R) \cap \{t_2, t_3\}$ . Now, since  $\{t_2\} \in \nu_{\mathcal{T}}(t_1)$  and  $\{t_2, t_3\} \notin \nu_{\mathcal{T}}(t_1)$ , we can conclude from condition 1a of Proposition 3.15 that  $t_1$  cannot be  $R$ -related to any state in  $\mathcal{S}$ , in particular not to  $s$ . Since  $R$  was an arbitrary  $2^2$ -bisimulation,  $t_1$  and  $s$  are not  $2^2$ -bisimilar.

Consider, now the relation  $R' = \langle \{t_1, t_2\} \rangle$  on the neighbourhood frame  $\mathcal{T}$ . The reader can check that  $R'$  is a precongurence, but not a precocongurence, on  $\mathcal{T}$ .

The above example shows that between neighbourhood frames, precocongurances are a better approximation of behavioural equivalence than  $2^2$ -bisimilarity. However, the next example shows that also precocongurances cannot capture behavioural equivalence, in general.

**Example 3.19.** We consider now a small variation on the picture given in Example 3.18. The neighbourhood frames  $\mathcal{S}$ ,  $\mathcal{U}$  and the function  $f_2$  are the same as before, but on  $\mathcal{T}$  we now take as neighbourhood function  $\nu'_{\mathcal{T}}(t_1) = \{\{t_2\}\}$ ,  $\nu'_{\mathcal{T}}(t_2) = \nu'_{\mathcal{T}}(t_3) = \{\emptyset\}$ , and let  $\mathcal{T}' = \langle \mathcal{T}, \nu'_{\mathcal{T}} \rangle$ . Instead of the function  $f_1$ , we take the function  $f'_1: \mathcal{T} \rightarrow \mathcal{U}$  with graph  $\text{Gr}(f'_1) = \langle \{t_1, u_1\}, \{t_2, u_2\}, \{t_3, u_2\} \rangle$ . Again, it is straightforward to check that  $f'_1$  is a bounded morphism, and hence  $t_1$  and  $s$  are behaviourally equivalent.



However, there is no precocongurence containing the pair  $\langle t_1, s \rangle$ . Suppose  $R' \subseteq \mathcal{T}' \times \mathcal{S}$  is an arbitrary precocongurence between  $\mathcal{T}'$  and  $\mathcal{S}$ . Since  $\langle \emptyset, \emptyset \rangle$  is  $R'$ -coherent,  $\emptyset \in \nu'_{\mathcal{T}'}(t_2)$  and  $\emptyset \notin \nu_{\mathcal{S}}(s)$ , it follows from Proposition 3.16 that  $\langle t_2, s \rangle \notin R'$ . This implies that  $\langle \{t_2\}, \emptyset \rangle$  is  $R'$ -coherent, but  $\{t_2\} \in \nu'_{\mathcal{T}'}(t_1)$  and  $\emptyset \notin \nu_{\mathcal{S}}(s)$ , so  $\langle t_1, s \rangle \notin R'$ .

To sum it up: Example 3.18 showed that precocongurances are a clear improvement when compared to  $2^2$ -bisimulations. Example 3.19, however, demonstrates that precocongurances are still incomplete as a proof principle for behavioural equivalence over neighbourhood frames.

From Theorem 3.12 of the previous subsection, we know that on a single neighbourhood frame, precocongurances do capture behavioural equivalence. Using the results of this subsection it follows easily that, in fact, also  $2^2$ -bisimilarity captures behavioural equivalence on a single structure.

**Proposition 3.20.** *If  $\mathcal{S} = \langle S, \nu \rangle$  is a neighbourhood frame, and  $R \subseteq S \times S$  is an equivalence relation, then:*

*$R$  is a  $2^2$ -bisimulation iff  $R$  is a precocongurence iff  $R$  is a congruence.*



Consequently, for all  $s_1, s_2 \in S$ :  $s_1 \leftrightarrow s_2$  iff  $s_1 \leftrightarrow_p s_2$  iff  $s_1 \leftrightarrow_b s_2$ .

*Proof.* If  $R \subseteq S \times S$  is an equivalence relation, then in particular  $\text{dom}(R) = \text{rng}(R) = S$ , and hence condition 1 of Proposition 3.15 is trivially satisfied. It follows from the characterisations in Propositions 3.15 and 3.16 that  $R$  is a  $2^2$ -bisimulation iff  $R$  is a precongurence. The second equivalence is an instance of the more general result in Theorem 3.12. The final claim is an immediate consequence of the main claim and Lemma 3.4.  $\square$

**Remark 3.21.** Alternatively, Proposition 3.20 follows from the result in [19] that congruences are  $\mathbf{F}$ -bisimulations in case the functor  $\mathbf{F}$  weakly preserves kernel pairs - a property that the functor  $2^2$  has as the following argument shows: Let  $f: S \rightarrow T$  be a function and consider its kernel  $\ker(f) := \{\langle s, s' \rangle \in S \times S \mid f(s) = f(s')\}$  with projections  $\pi_i: \ker(f) \rightarrow S$  for  $i = 1, 2$ . We have to show that for every pair of sets  $N_1, N_2 \in \ker(2^2(f))$  there exists a set  $N \in 2^2(\ker(f))$  such that  $2^2(\pi_i)(N) = N_i$  for  $i = 1, 2$ . Let  $N_1, N_2$  be elements of  $2^2(S)$  such that  $2^2(f)(N_1) = 2^2(f)(N_2)$ . We put  $N := \{\pi_1^{-1}(U_1) \mid U_1 \in N_1\} \cup \{\pi_2^{-1}(U_2) \mid U_2 \in N_2\}$ . It is now easy to check that  $2^2(\pi_i)(N) = N_i$  for  $i = 1, 2$  as required.

#### 4. HENNESSY-MILNER CLASSES

The Hennessy-Milner theorem for normal modal logic states that over the class of finite Kripke models, two states are Kripke bisimilar if and only if they satisfy the same modal formulas. It is well known (see e.g. [10]), that this Hennessy-Milner theorem can be generalised to hold over any class of modally saturated Kripke models, in particular, over the class of image-finite Kripke models.

In this section, we define modal saturation and image-finiteness for neighbourhood models and show that each of these properties leads to a Hennessy-Milner style theorem. In the last subsection we describe ultrafilter extensions of neighbourhood models, and show that they are modally saturated.

First, we make precise what we mean by a Hennessy-Milner class of neighbourhood models. Since we have three equivalence notions for neighbourhood models, we have, in principle, three types of Hennessy-Milner classes. However, Examples 3.18 and 3.19 of section 3 showed that even over the class of finite neighbourhood models, two states can be behaviourally equivalent, and hence modally equivalent, without being linked by a precongurence or a bisimulation. This means that precongurences and bisimulations do not fit well with the expressivity of the modal language. We therefore define Hennessy-Milner classes with respect to behavioural equivalence.

**Definition 4.1.** A class  $\mathbf{K}$  of neighbourhood models is a *Hennessy-Milner class*, if for any  $\mathcal{M}_1$  and  $\mathcal{M}_2$  in  $\mathbf{K}$  containing states  $s_1$  and  $s_2$ , respectively, we have:  $\mathcal{M}_1, s_1 \equiv \mathcal{M}_2, s_2$  iff  $\mathcal{M}_1, s_1 \leftrightarrow_b \mathcal{M}_2, s_2$ .

The following lemma provides an easy, but useful, criterion for proving that a class of models is a Hennessy-Milner class.

**Lemma 4.2.** *Let  $\mathbf{K}$  be a class of neighbourhood models. If for any  $\mathcal{M}_1, \mathcal{M}_2 \in \mathbf{K}$ , the modal equivalence relation  $\equiv$  is a congruence on  $\mathcal{M}_1 + \mathcal{M}_2$ , then  $\mathbf{K}$  is a Hennessy-Milner class.*

*Proof.* Let  $\mathcal{M}_1$  and  $\mathcal{M}_2$  be neighbourhood models in  $\mathbf{K}$ , and let  $\iota_i: \mathcal{M}_i \rightarrow \mathcal{M}_1 + \mathcal{M}_2$  denote the canonical inclusion morphisms. Assume that we have states  $s_1$  and  $s_2$  such that  $\mathcal{M}_1, s_1 \equiv \mathcal{M}_2, s_2$ . Since truth is invariant under bounded morphisms, we have  $\iota_1(s_1) \equiv \iota_2(s_2)$  in  $\mathcal{M}_1 + \mathcal{M}_2$ . By assumption,  $\equiv$  is a congruence on  $\mathcal{M}_1 + \mathcal{M}_2$ , hence  $\varepsilon: \mathcal{M}_1 + \mathcal{M}_2 \rightarrow (\mathcal{M}_1 + \mathcal{M}_2)/\equiv$  is a bounded morphism (as illustrated by the diagram), and  $\langle s_1, s_2 \rangle \in \text{pb}(\varepsilon \circ \iota_1, \varepsilon \circ \iota_2)$ , hence  $s_1 \leftrightarrow_b s_2$ .  $\square$

$$\begin{array}{ccc} \mathcal{M}_1 & \xrightarrow{\iota_1} & \mathcal{M}_1 + \mathcal{M}_2 & \xleftarrow{\iota_2} & \mathcal{M}_2 \\ & & \downarrow \varepsilon & & \\ & & (\mathcal{M}_1 + \mathcal{M}_2)/\equiv & & \end{array}$$

**4.1. Modally saturated models.** In Lemma 4.2 we saw that in order to prove a Hennessy-Milner theorem, we are interested in neighbourhood models on which modal equivalence is a congruence. Let  $\mathcal{M} = \langle S, \nu, V \rangle$  be a neighbourhood model. By applying the characterisations of congruences on neighbourhood frames in Corollary 3.17 and adding the condition for the atomic propositions, we find that  $\equiv$  is a congruence on  $\mathcal{M}$  iff for all  $s, t \in S$  such that  $s \equiv t$ :

$$\begin{aligned} \text{(c1)} \quad & \text{for all } p \in \text{At} : & s \in V(p) & \iff & t \in V(p), \text{ and} \\ \text{(c2)} \quad & \text{for all modally coherent } X \subseteq S : & X \in \nu(s) & \iff & X \in \nu(t). \end{aligned} \tag{4.1}$$

Clearly, condition (c1) holds in all neighbourhood models, since modally equivalent states must make the same atomic propositions true. One way of making condition (c2) hold, is to ensure that all modally coherent neighbourhoods are definable.

**Lemma 4.3.** *Let  $\mathcal{M} = \langle S, \nu, V \rangle$  be a neighbourhood model. If for all  $s \in S$  and all modally coherent  $X \in \nu(s)$ , there exists a modal  $\mathcal{L}$ -formula  $\varphi$  such that  $X = \llbracket \varphi \rrbracket^{\mathcal{M}}$ , then modal equivalence is a congruence on  $\mathcal{M}$ .*

*Proof.* Let  $X$  be a modally coherent neighbourhood of some state, and assume  $X = \llbracket \varphi \rrbracket^{\mathcal{M}}$ . We have for any  $s, t \in S$  such that  $s \equiv t$ :  $X \in \nu(s)$  iff  $\mathcal{M}, s \models \Box \varphi$  iff  $\mathcal{M}, t \models \Box \varphi$  iff  $X \in \nu(t)$ .  $\square$

For finite models, a standard argument shows that any modally coherent neighbourhood  $X$  is definable by a formula of the form  $\delta = \bigvee_{i \leq n} \bigwedge_{j \leq k} \delta_{i,j}$  where  $n, k < \omega$ . For infinite models, the same argument would yield a formula with an infinite disjunction and conjunction, which is not a well-formed formula of our finitary language. Modal saturation is a compactness property which allows us to replace infinite conjunctions and disjunctions with finite ones<sup>1</sup>. Thus we can essentially use the same argument as in finite models to show that modally coherent neighbourhoods are definable (and we do so in Lemma 4.5 below). We will use the following notation. Let  $\Psi$  be a set of modal  $\mathcal{L}$ -formulas and  $\mathcal{M} = \langle S, \nu, V \rangle$  a neighbourhood model. We define  $\neg \Psi = \{\neg \psi \mid \psi \in \Psi\}$ ,  $\llbracket \bigwedge \Psi \rrbracket^{\mathcal{M}} = \bigcap_{\psi \in \Psi} \llbracket \psi \rrbracket^{\mathcal{M}}$ , and  $\llbracket \bigvee \Psi \rrbracket^{\mathcal{M}} = \bigcup_{\psi \in \Psi} \llbracket \psi \rrbracket^{\mathcal{M}}$ . A set  $\Psi$  of  $\mathcal{L}$ -formulas is *satisfiable in a subset*  $X \subseteq S$  of  $\mathcal{M}$ , if  $\llbracket \bigwedge \Psi \rrbracket^{\mathcal{M}} \cap X \neq \emptyset$ . A set  $\Psi$  of  $\mathcal{L}$ -formulas is *finitely satisfiable in*  $X \subseteq S$ , if any finite subset  $\Psi_0 \subseteq \Psi$  is satisfiable in  $X$ .

**Definition 4.4.** Let  $\mathcal{M} = \langle S, \nu, V \rangle$  be a neighbourhood model. A subset  $X \subseteq S$  is called *modally compact* if for all sets  $\Psi$  of modal  $\mathcal{L}$ -formulas,  $\Psi$  is satisfiable in  $X$  whenever  $\Psi$  is finitely satisfiable in  $X$ . The neighbourhood model  $\mathcal{M}$  is *modally saturated*, if for all  $s \in S$

<sup>1</sup>This perspective on modal saturation was pointed out to us by H.P. Gumm (personal correspondence).

and all modally coherent neighbourhoods  $X \in \nu(s)$ , both  $X$  and the complement  $X^c$  are modally compact.

To see why modal compactness is really a compactness property, note that for a subset  $X$  in a neighbourhood model  $\mathcal{M}$ ,  $X \subseteq \llbracket \bigvee \Psi \rrbracket^{\mathcal{M}}$  iff  $\{\neg\psi \mid \psi \in \Psi\}$  is not satisfiable in  $X$ . Hence  $X$  is modally compact, if and only if, for all  $\Psi \subseteq \mathcal{L}$  such that  $X \subseteq \llbracket \bigvee \Psi \rrbracket^{\mathcal{M}}$  there is a  $\Psi_0 \subseteq_{\omega} \Psi$  such that  $X \subseteq \llbracket \bigvee \Psi_0 \rrbracket^{\mathcal{M}}$ . Clearly, any finite set is modally compact. Note also that, in Definition 4.4, due to the fact that  $\llbracket \bigwedge \Psi \rrbracket^{\mathcal{M}} \subseteq X$  if and only if  $X^c \subseteq \llbracket \bigvee \neg\Psi \rrbracket^{\mathcal{M}}$ , we have that  $X^c$  is modally compact, if and only if, for all  $\Psi \subseteq \mathcal{L}$  such that  $\llbracket \bigwedge \Psi \rrbracket^{\mathcal{M}} \subseteq X$ , there is a  $\Psi_0 \subseteq_{\omega} \Psi$  such that  $\llbracket \bigwedge \Psi_0 \rrbracket^{\mathcal{M}} \subseteq X$ .

**Lemma 4.5.** *Let  $\mathcal{M} = \langle S, \nu, V \rangle$  be a modally saturated neighbourhood model. For all  $X \subseteq S$ :  $X$  is modally coherent iff  $X$  is definable by a modal  $\mathcal{L}$ -formula.*

*Proof.* If  $X = \llbracket \varphi \rrbracket^{\mathcal{M}}$  for some  $\varphi \in \mathcal{L}$ , then clearly  $X$  is modally coherent. For the converse implication, assume  $X$  is modally coherent, i.e.,  $X$  is a union of modal equivalence classes  $X = \bigcup_{c \in C} [x_c]_{\equiv}$ . For  $c \in C$  and  $y \neq x_c$  there is a modal  $\mathcal{L}$ -formula  $\delta_{c,y}$  such that  $x_c \models \delta_{c,y}$  and  $y \models \neg\delta_{c,y}$ , so by taking  $\Delta_c = \{\delta_{c,y} \mid y \neq x_c\}$ , we have  $[x_c]_{\equiv} = \llbracket \bigwedge \Delta_c \rrbracket^{\mathcal{M}} \subseteq X$  for each  $c \in C$ . By modal compactness of  $X^c$ , for each  $c \in C$  there is a finite subset  $\Delta_c^0 \subseteq_{\omega} \Delta_c$  such that  $[x_c]_{\equiv} \subseteq \llbracket \bigwedge \Delta_c^0 \rrbracket^{\mathcal{M}} \subseteq X$ . Defining  $\delta_c = \bigwedge \Delta_c^0$  for each  $c \in C$ , we therefore have  $X = \bigcup_{c \in C} \llbracket \delta_c \rrbracket^{\mathcal{M}}$ . Now by modal compactness of  $X$ , we get a finite subset  $\Delta_0 \subseteq_{\omega} \{\delta_c \mid c \in C\}$  such that  $X = \llbracket \bigvee \Delta_0 \rrbracket^{\mathcal{M}}$ . That is,  $X$  is definable by the formula  $\delta = \bigvee \Delta_0$ .  $\square$

**Proposition 4.6.** *If  $\mathcal{M}$  is a modally saturated neighbourhood model, then modal equivalence is a congruence on  $\mathcal{M}$ . It follows that modally equivalent states in  $\mathcal{M}$  are behaviourally equivalent.*

*Proof.* Immediate consequence of Lemmas 4.3 and 4.5.  $\square$

**Corollary 4.7.** *The class of finite neighbourhood models is a Hennessy-Milner class.*

*Proof.* Since the disjoint union of two finite neighbourhood models is again finite, it suffices by Lemma 4.2 and Proposition 4.6 to show that finite neighbourhood models are modally saturated. But this is immediate, since any set of states in a finite neighbourhood model  $\mathcal{M}$ , is necessarily finite, and hence modally compact, so  $\mathcal{M}$  is modally saturated.  $\square$

The question remains whether the class of all modally saturated neighbourhood models is a Hennessy-Milner class. We conjecture that if  $\mathcal{M}$  and  $\mathcal{N}$  are modally saturated then modal equivalence is a congruence on  $\mathcal{M} + \mathcal{N}$ . If this is the case, then the Hennessy-Milner theorem follows from Lemma 4.2.

**Remark 4.8.** In [36] the following definition of modal saturation for monotonic neighbourhood models was introduced, and it was shown that over the class of modally saturated monotonic neighbourhood models modal equivalence implies monotonic bisimilarity. A monotonic neighbourhood model  $\langle S, \nu, V \rangle$  is *monotonic modally saturated*, if for all  $s \in S$  and all sets  $\Psi$  of modal  $\mathcal{L}$ -formulas the following hold:

- (m1-mon) For all  $X \in \nu(s)$ , if  $\Psi$  is finitely satisfiable in  $X$ , then  $\Psi$  is satisfiable in  $X$ .
- (m2-mon) If for all  $\Psi_0 \subseteq_{\omega} \Psi$ , there is an  $X \in \nu(s)$  such that  $X \subseteq (\bigwedge \Psi_0)$ , then there is an  $X \in \nu(s)$  such that  $X \subseteq (\bigwedge \Psi)$ .

In a monotonic neighbourhood model  $\mathcal{M}$ , (m1-mon) clearly implies that all modally coherent neighbourhoods are modally compact. The converse also holds, since for any neighbourhood  $X$  of some state  $s$ , the closure  $X'$  of  $X$  with respect to modal equivalence, i.e.,  $X' = \bigcup_{x \in X} [x]_{\equiv}$ , is also a neighbourhood of  $s$  by monotonicity, and for any  $\Psi \subseteq \mathcal{L}$ ,  $\Psi$  is satisfiable in  $X$  if and only if  $\Psi$  is satisfiable in  $X'$ . However, it is not clear whether monotonic modal saturation and (neighbourhood) modal saturation coincide in all monotonic models. We suspect that neither implies the other due to the following. The condition (m2-mon) says that all neighbourhood collections are closed under arbitrary intersections of definable neighbourhoods, a property which we expect can be shown to fail in some modally saturated neighbourhood model. On the other hand, it is not clear why the complements of modally coherent neighbourhoods should be modally compact in a monotonic modally saturated model. Unfortunately, at the moment we have no examples that confirm these intuitions.

**Remark 4.9.** A Kripke model  $\mathcal{K} = \langle S, R, V \rangle$  is *Kripke modally saturated*, if for all  $s \in S$  and all sets  $\Psi$  of modal  $\mathcal{L}$ -formulas:

(m1-krip) If  $\Psi$  is finitely satisfiable in  $R[s]$ , then  $\Psi$  is satisfiable in  $R[s]$ ,

and over the class of modally saturated Kripke models, modal equivalence implies Kripke bisimilarity (see e.g. [10]). From the above definitions, it is clear that for any augmented neighbourhood model  $\mathcal{M}$ , if  $\mathcal{M}$  is monotonic modally saturated or (neighbourhood) modally saturated, then  $\mathcal{M}^{\text{krip}}$  is Kripke modally saturated. However, if  $\mathcal{M}^{\text{krip}}$  is Kripke modally saturated, then modally coherent neighbourhoods may fail to be modally compact in  $\mathcal{M}$ . This is shown by Example 4.18 (page 24) in the next subsection. Hence Kripke modal saturation does not imply monotonic modal saturation nor (neighbourhood) modal saturation. Note that (m2-mon) holds over any augmented neighbourhood model.

As we have seen in Remarks 4.8 and 4.9, the notions of neighbourhood, monotonic and Kripke modal saturation do not restrict in a natural way. Moreover, in the next subsection (Example 4.18), we will see that image-finite neighbourhood models are not necessarily modally saturated. These observations could be interpreted as arguments for saying that our definition of modal saturation for neighbourhood models is not the right one. On the other hand, Definition 4.4 arises in a natural manner, it implies Kripke modal saturation over Kripke models, in subsection 4.3 we show that ultrafilter extensions of neighbourhood models are modally saturated, and in subsection 5.2 we will see that when viewing neighbourhood models as first-order models, then  $\omega$ -saturation implies modal saturation (Lemma 5.6). We believe these are good arguments for Definition 4.4 being the right notion after all. However, further investigations are needed to support this claim. It would be useful to have a better understanding of what an abstract notion of modal saturation for F-coalgebras should be.

**4.2. Image-finite neighbourhood models.** In normal modal logic, we know that image-finite Kripke models are modally saturated, and hence form a Hennessy-Milner class with respect to Kripke bisimilarity. In this section, we describe image-finite neighbourhood models and prove that they form a Hennessy-Milner class, despite the fact that, in general, they are not modally saturated.

**Remark 4.10.** We obtain our notion of an image-finite neighbourhood model by instantiating a widely used categorical definition. Similarly, we could obtain the Hennessy-Milner

result of this section by using a far more general theorem from coalgebraic modal logic. Our motivation for giving an “elementary” proof is that we want to equip the working modal logician with some intuition concerning image-finite neighbourhood models. We outline how the result could be obtained as a corollary from coalgebraic work in Remark 4.16 below.

In contrast with the Kripke case, image-finite neighbourhood models are not necessarily modally saturated. Instead, we will show that they satisfy the condition of the following lemma.

**Lemma 4.11.** *Let  $\mathcal{M} = \langle S, \nu, V \rangle$  be a neighbourhood model. If for any states  $s_1, s_2 \in S$  and any modally coherent subset  $X \subseteq S$  there is a formula  $\varphi \in \mathcal{L}$  such that for any  $i \in \{1, 2\}$ ,  $X \in \nu(s_i)$  if and only if  $\llbracket \varphi \rrbracket^{\mathcal{M}} \in \nu(s_i)$ , then modal equivalence is a congruence on  $\mathcal{M}$ .*

*Proof.* Immediate by the characterisation given by conditions (c1) and (c2) on page 18.  $\square$

A Kripke model is image-finite if every state has only finitely many successors (cf. [10]). For neighbourhood models, the notion of image-finiteness is less obvious, but as with bisimilarity, universal coalgebra provides us with an abstract notion of image-finiteness for coalgebras which we instantiate for the  $2^2$ -functor. The general construction behind this definition is that of taking the finitary part of a functor. Recall that we denote the inclusion map of  $Y \subseteq X$  by  $\iota_Y : Y \hookrightarrow X$ . Given any functor  $F : \mathbf{Set} \rightarrow \mathbf{Set}$ , define the functor  $F_\omega$  by letting

$$F_\omega(X) = \bigcup \{F(\iota_Y)[FY] \mid \iota_Y : Y \hookrightarrow X, Y \subseteq_\omega X\}$$

for a set  $X$ , and for a function  $f : X \rightarrow Y$ ,  $F_\omega(f)$  is the restriction of  $F(f)$  to  $F_\omega(X)$ . It is known that  $F_\omega$  is the unique finitary (or  $\omega$ -accessible) subfunctor of  $F$  which agrees with  $F$  on all finite sets (see e.g. [3, 34]), and  $F_\omega$  is called the finitary part of  $F$ . We now give a characterisation of the finitary part of  $2^2$ . For a subset inclusion map  $\iota_B : B \hookrightarrow X$  and  $D \subseteq X$ , note that  $\iota_B^{-1}[D] = D \cap B$ . If  $U \in 2^2_\omega(X)$  and  $B \subseteq X$  is such that for all  $D \subseteq X$ :  $D \in U \iff D \cap B \in U$ , then we call  $B$  a *base set for  $U$* .

**Lemma 4.12.** *Let  $X$  be a set. We have:*

$$2^2_\omega(X) = \{U \in 2^2(X) \mid \exists B \subseteq_\omega X. \forall D \subseteq X : (D \in U \iff D \cap B \in U)\}.$$

*Proof.* The proof is obtained by spelling out the definitions.  $\square$

**Definition 4.13.** We define the class of image-finite neighbourhood frames as the class  $\mathbf{Coalg}(2^2_\omega)$  of  $2^2_\omega$ -coalgebras. The class of image-finite neighbourhood models is the class of neighbourhood models based on an image-finite neighbourhood frame.

So, image-finite neighbourhood frames are the neighbourhood frames in which all neighbourhood collections are determined by a finite base set. It should be clear that a finite neighbourhood frame  $\langle S, \nu \rangle$  is image-finite, since for all  $s \in S$ ,  $S$  is a finite base set for  $\nu(s)$ . In proving that image-finite neighbourhood models form a Hennessy-Milner class, we use the following lemma.

**Lemma 4.14.** *Let  $S$  be a set and  $\theta$  an equivalence relation on  $S$ . Moreover, let  $B \subseteq S$  and denote by  $B_\theta \subseteq B$  a set of representatives of the  $\theta$ -classes intersecting  $B$ . For all  $X, X' \subseteq S$ , if  $X$  and  $X'$  are both  $\theta$ -coherent, then  $X \cap B = X' \cap B$  iff  $X \cap B_\theta = X' \cap B_\theta$ .*

*Proof.* Let  $S, B$  and  $B_\theta \subseteq B$  be as stated, and assume that  $X$  and  $X'$  are  $\theta$ -coherent subsets of  $S$ . It is clear that  $X \cap B = X' \cap B$  implies  $X \cap B_\theta = X' \cap B_\theta$ . For the other implication, assume  $X \cap B_\theta = X' \cap B_\theta$ . We have:  $s \in X \cap B$  implies there is an  $s' \in B_\theta$  such that  $s\theta s'$ .

Since  $X$  is  $\theta$ -coherent,  $s' \in X \cap B_\theta = X' \cap B_\theta$ . Now since  $X'$  is  $\theta$ -coherent,  $s \in X'$ , and thus  $s \in X' \cap B$ . Hence we have shown  $X \cap B \subseteq X' \cap B$ . The other inclusion is shown similarly.  $\square$

**Proposition 4.15.** *The class of image-finite neighbourhood models is a Hennessy-Milner class.*

*Proof.* The class of image-finite neighbourhood models is closed under disjoint unions, since for any functor  $F$ , the category  $\text{Coalg}(F)$  has coproducts (cf. [39]). By Lemma 4.2 it suffices to show that in an image-finite neighbourhood model, modal equivalence is a congruence. So let  $\mathcal{M} = \langle S, \nu, V \rangle$  be image-finite, and let  $s, t \in S$ . We then have finite base sets  $B_s, B_t \subseteq_\omega S$  for  $\nu(s)$  and  $\nu(t)$ , respectively. Let  $B_{st} = B_s \cup B_t$ . By Lemma 4.11 it suffices to find for any modally coherent  $X \subseteq S$ , a formula  $\varphi \in \mathcal{L}$  such that

$$X \cap B_{st} = \llbracket \varphi \rrbracket^{\mathcal{M}} \cap B_{st}, \quad (4.2)$$

since then  $X \cap B_s = \llbracket \varphi \rrbracket^{\mathcal{M}} \cap B_s$  and  $X \cap B_t = \llbracket \varphi \rrbracket^{\mathcal{M}} \cap B_t$ , and hence  $X \in \nu(s)$  iff  $\llbracket \varphi \rrbracket^{\mathcal{M}} \in \nu(s)$ , similarly for  $t$ , and consequently, if  $s \equiv t$ , then  $X \in \nu(s)$  if and only if  $X \in \nu(t)$ .

We now show how to obtain such a  $\varphi$ . Let  $X \subseteq S$  be modally coherent and let  $B'_{st} \subseteq B_{st}$  be a set of representatives of the  $\equiv$ -classes intersecting  $B_{st}$ . Since  $B_{st}$  is finite, so is  $B'_{st}$ . Assume  $B'_{st} = \{s_1, \dots, s_n\}$ . Now there are modal formulas  $\varphi_1, \dots, \varphi_n \in \mathcal{L}$  which characterise  $s_1, \dots, s_n$ , respectively, within  $B'_{st}$ , that is,  $\mathcal{M}, s_i \models \varphi_j$  iff  $i = j$ , for  $1 \leq i, j \leq n$ . Namely, for each  $s_i \in B'_{st}$ , we have for all  $s_j \in B'_{st} \setminus \{s_i\}$ ,  $s_i \not\models s_j$ . Hence there is a formula  $\varphi_{i,j}$  such that  $\mathcal{M}, s_i \models \varphi_{i,j}$  and  $\mathcal{M}, s_j \not\models \varphi_{i,j}$ . Take  $\varphi_i = \bigwedge_{j=1, j \neq i}^n \varphi_{i,j}$ ,  $i = 1, \dots, n$ . We now define  $\varphi = \bigvee \{\varphi_i \mid s_i \in X \cap B'_{st}\}$ . To see that  $\varphi$  satisfies (4.2) it suffices by Lemma 4.14 to show that  $X \cap B'_{st} = \llbracket \varphi \rrbracket^{\mathcal{M}} \cap B'_{st}$ . Clearly, by definition of  $\varphi$ , if  $s_i \in X \cap B'_{st}$  then  $s_i \in \llbracket \varphi \rrbracket^{\mathcal{M}} \cap B'_{st}$ . Conversely, if  $s_j \in \llbracket \varphi \rrbracket^{\mathcal{M}} \cap B'_{st}$  then  $\mathcal{M}, s_j \models \varphi_i$  for some  $i$  such that  $s_i \in X \cap B'_{st}$ . Since  $\varphi_i$  characterises  $s_i$  in  $B'_{st}$ , it follows that  $s_j = s_i \in X \cap B'_{st}$ .  $\square$

**Remark 4.16.** As we already mentioned, Proposition 4.15 is a consequence of a more general result in coalgebraic modal logic, which we briefly explain here. In coalgebraic modal logic, the semantics of modalities is given by predicate liftings. A predicate lifting for a functor  $F: \text{Set} \rightarrow \text{Set}$  is a natural transformation  $\lambda: 2 \rightarrow 2 \circ F$ . Given a set  $\Lambda$  of predicate liftings for  $F$ , the finitary coalgebraic modal language  $\mathcal{L}(\Lambda)$  is the multi-modal language which contains a modality  $[\lambda]$  for each  $\lambda \in \Lambda$ . Given an  $F$ -coalgebra  $\mathcal{X} = \langle X, \xi \rangle$ , the truth of formulas is defined in the standard inductive manner for the basic Boolean connectives. The truth of a modal formula  $[\lambda]\phi$  is defined by:  $\mathcal{X}, x \models [\lambda]\phi$  iff  $\xi(x) \in \lambda_X(\llbracket \phi \rrbracket^{\mathcal{X}})$ . Atomic propositions can also be interpreted using constant predicate liftings. We refer to [35] for details.

Using currying, every predicate lifting  $\lambda: 2 \rightarrow 2 \circ F$  corresponds to a natural transformation  $\hat{\lambda}: F \rightarrow 2^2$ , called the transpose of  $\lambda$ . A set  $\Lambda$  of predicate liftings for  $F$  is called *separating* if the source of transposes  $\{\hat{\lambda} \mid \lambda \in \Lambda\}$  is jointly injective. Schröder shows in [40, Theorem 41, Corollary 45]) that if  $F: \text{Set} \rightarrow \text{Set}$  is a finitary functor, and  $\Lambda$  is a separating set of predicate liftings, then the finitary coalgebraic modal language  $\mathcal{L}(\Lambda)$  is expressive for  $F$ -coalgebras, meaning that over the class of  $F$ -coalgebras,  $\mathcal{L}(\Lambda)$ -equivalence implies behavioural equivalence.

We can instantiate the result for the finitary functor  $2_\omega^2 \times \mathcal{P}(\text{At})$  and classical modal logic. The basic modal language and its interpretation over neighbourhood models is the finitary coalgebraic modal logic given by  $\Lambda = \{\lambda\} \cup \{\rho_i \mid i < \omega\}$ , where  $\lambda: 2 \rightarrow 2 \circ 2_\omega^2$  is defined by  $\lambda_X(A) = \{U \in 2_\omega^2(X) \mid A \in U\}$ , and the  $\rho_i$ ,  $i < \omega$ , are constant predicate liftings

that interpret the atomic propositions. It is known that  $\{\lambda\} \cup \{\rho_i \mid i < \omega\}$  is separating iff  $\{\lambda\}$  is separating. The transposite  $\hat{\lambda}: 2_{\omega}^2 \rightarrow 2^2$  is simply the inclusion map, i.e.,  $\hat{\lambda}_X = \iota_{2_{\omega}^2(X)}$  for all sets  $X$ , so trivially  $\{\hat{\lambda}\}$  is jointly injective, hence  $\{\lambda\}$  is separating. It now follows from Schröder's result that over the class of image-finite neighbourhood models, modal equivalence implies behavioural equivalence.

We now show that the notion of image-finiteness for neighbourhood frames restricts to the subclasses of neighbourhood frames that correspond with Kripke frames and monotonic neighbourhood frames, respectively.

Monotonic neighbourhood frames are coalgebras for the subfunctor  $\text{Mon}$  of  $2^2$  (cf. Remark 2.7) which sends a set  $X$  to the collection of all subsets of  $\mathcal{P}(X)$  which are closed under supersets. Due to monotonicity, given a function  $f: X \rightarrow Y$ , we can describe  $\text{Mon}(f)$  in terms of the direct image of  $f$ , namely, for all  $V \in \text{Mon}(X)$ ,  $\text{Mon}(f)(V) = \bigcup \{\uparrow f[D] \mid D \in V\}$ . Recall that for a subset  $B \subseteq X$ ,  $\uparrow B = \{B' \subseteq X \mid B \subseteq B'\}$ . Image-finite monotonic neighbourhood frames, are then nothing but  $\text{Mon}_{\omega}$ -coalgebras. By simply working out the definitions, we find that for a set  $X$  and  $U \in \text{Mon}(X)$ :

$$U \in \text{Mon}_{\omega}(X) \quad \text{iff} \quad \exists C_1, \dots, C_n \subseteq_{\omega} X : U = \uparrow C_1 \cup \dots \cup \uparrow C_n.$$

The neighbourhood collections in an image-finite monotonic neighbourhood model are thus generated by finite sets of finite neighbourhoods which are minimal with respect to  $\subseteq$  in  $\mathcal{P}(X)$ . Such minimal neighbourhoods will be referred to as core neighbourhoods. More precisely, if  $\mathcal{M} = \langle S, \nu, V \rangle$  is a neighbourhood model,  $s \in S$  and  $C \in \nu(s)$  is such that for all  $D \subsetneq C$ ,  $D \notin \nu(s)$ ,  $C$  is called a *core neighbourhood* of  $s$ . The collection of core neighbourhoods of  $s$  is denoted  $\nu^c(s)$ . This terminology follows [37, 20] where image-finite monotonic neighbourhood models were called *locally core finite*.

Finally, recall that a Kripke model  $\langle S, R, V \rangle$  is image-finite, if for all  $s \in S$ , the set of  $R$ -successors  $R[s]$  is finite.

**Proposition 4.17.** *Let  $\mathcal{M} = \langle S, \nu, V \rangle$  be a neighbourhood model.*

- (1) *If  $\mathcal{M}$  is a monotonic neighbourhood model, then  $\mathcal{M}$  is image-finite as a monotonic neighbourhood model iff  $\mathcal{M}$  is image-finite as a neighbourhood model.*
- (2) *If  $\mathcal{M}$  is augmented, then  $\mathcal{M}^{\text{krip}}$  is image-finite as a Kripke model iff  $\mathcal{M}$  is image-finite as a neighbourhood model.*

*Proof.* To prove item 1, let  $\mathcal{M}$  be monotonic. Since  $\text{Mon}$  is a subfunctor of  $2^2$ , also  $\text{Mon}_{\omega}$  is a subfunctor of  $2_{\omega}^2$ . It follows that any image-finite monotonic model is also image-finite as a neighbourhood model. Concretely, one can show that for all  $s \in S$ , the union of core neighbourhoods  $B = \bigcup \nu^c(s)$  is a finite base set for  $\nu(s)$ . For the other direction, assume  $\mathcal{M}$  is image-finite as a neighbourhood model. Let  $s \in S$ , and assume  $B \subseteq_{\omega} S$  is a finite base set for  $\nu(s)$ . We first show that every neighbourhood is in the upwards closure of some finite core neighbourhood:  $U \in \nu(s)$  implies  $B \cap U \in \nu(s)$ , and since  $B \cap U$  is finite, there must be a finite  $C \in \nu^c(s)$  such that  $C \subseteq B \cap U \subseteq U$ . Suppose now that  $C \in \nu^c(s)$  is an arbitrary core neighbourhood of  $s$ . As  $B$  is a base set for  $\nu(s)$ ,  $C \cap B \in \nu(s)$ , and hence by  $\subseteq$ -minimality of  $C$ ,  $C \subseteq B$ . It now follows from the finiteness of  $B$ , that  $s$  has only finitely many core neighbourhoods  $C_1, \dots, C_n$  of finite cardinality, and  $\nu(s) = \uparrow C_1 \cup \dots \cup \uparrow C_n$ .

For item 2, let  $\mathcal{M}^{\text{krip}} = \langle S, R, V \rangle$ , i.e., for all  $s \in S$ ,  $\nu(s) = \uparrow R[s]$ , and  $\nu^c(s) = \{R[s]\}$ . This immediately shows that if  $\mathcal{M}^{\text{krip}}$  is image-finite then  $\mathcal{M}$  is image-finite as a monotonic model, and hence by item 1, also as a neighbourhood model. Conversely, if  $\mathcal{M}$  is image-finite,

then by item 1  $\mathcal{M}$  is image-finite as a monotonic model, hence for all  $s \in S$ ,  $\bigcup \nu^c(s) = R[s]$  is finite.  $\square$

The following example demonstrates that image-finite neighbourhood models are not necessarily modally saturated, and it also shows that a Kripke modally saturated model, is not necessarily modally saturated as a (monotonic) neighbourhood model.

**Example 4.18.** Consider the Kripke model  $\mathcal{K} = \langle S, R, V \rangle$  where  $S = \mathbb{N}$ , the set of natural numbers, and  $R$  is the usual relation  $>$  on  $\mathbb{N}$ , that is, for  $m, n \in \mathbb{N}$ ,  $\langle m, n \rangle \in R$  iff  $m > n$ , and  $R[m] = \{n \in \mathbb{N} \mid n < m\}$ . Finally, the valuation  $V$  is defined as  $V(p_i) = \emptyset$ , for all atomic propositions  $p_i \in \text{At}$ .  $\mathcal{K}$  is an image-finite Kripke model, hence by Proposition 4.17 the augmented neighbourhood model  $\mathcal{K}^{\text{aug}}$  corresponding to  $\mathcal{K}$  is also image-finite as a (monotonic) neighbourhood model. Since  $\mathcal{K}$  is image-finite,  $\mathcal{K}$  is Kripke modally saturated. However,  $\mathcal{K}^{\text{aug}}$  is not modally saturated as a neighbourhood model nor as a monotonic model. To see this, first note that the set  $\mathbb{N}$  is trivially modally coherent and by monotonicity  $\mathbb{N}$  is also a neighbourhood of every  $n \in \mathbb{N}$ . Now, consider the set of modal  $\mathcal{L}$ -formulas,  $\Psi = \{\diamond^n \Box \perp \mid n \in \mathbb{N}\}$ . Note that by transitivity,  $\mathcal{K}, m \models \diamond^n \Box \perp$  iff  $m \geq n$ . Since  $\mathcal{K}$  and  $\mathcal{K}^{\text{aug}}$  are pointwise equivalent, and every finite subset  $\Psi_0 \subseteq_{\omega} \Psi$  is satisfiable in  $\mathcal{K}$  at the maximal  $n \in \mathbb{N}$  such that  $\diamond^n \Box \perp \in \Psi_0$ , it follows that  $\Psi$  is finitely satisfiable in the neighbourhood  $\mathbb{N}$  in  $\mathcal{K}^{\text{aug}}$ . However,  $\Psi$  is clearly not satisfiable in  $\mathbb{N}$ . We have thus shown that  $\mathbb{N}$  is not modally compact, hence  $\mathcal{K}^{\text{aug}}$  is not (monotonic) modally saturated.

**4.3. Ultrafilter extensions.** In this section, we prove a behavioural-equivalence-somewhere-else result by showing that any two modally equivalent states of neighbourhood models have behaviourally equivalent representatives in the ultrafilter extensions of these neighbourhood models. To this end, we define ultrafilter extensions of neighbourhood models, and we prove analogues of results known for ultrafilter extensions of Kripke models. In particular, we show that ultrafilter extensions are modally saturated. This result will be used in our proof of Craig interpolation in subsection 5.3.

Just as ultrafilter extensions of Kripke models are obtained from algebraic duality (see e.g. [10]), ultrafilter extensions of neighbourhood models are a by-product of a more general duality between coalgebras and certain algebras on the category of Boolean algebras, as described in e.g. [24, 27]. Our definition of ultrafilter extensions of neighbourhood frames is obtained by instantiating the more general definition of ultrafilter extensions of  $\mathbf{F}$ -coalgebras presented in [27] to  $\mathbf{F} = 2^2$ . The basic properties follow from the category theoretical framework. With quite some effort, the behavioural-equivalence-somewhere-else result can be obtained as a special case of a more general theorem in [24]. However, instead of requiring knowledge of the (rather abstract) theory in [24, 27], we have chosen to give a direct, concrete description of ultrafilter extensions of neighbourhood models, and to use standard model-theoretic techniques to prove basic properties. We believe that such a presentation will make the results of this section and the proof of the Craig interpolation theorem better accessible to readers whose background is mainly in modal logic. For the interested reader, we give a brief summary of the construction from [27] in Remark 4.21.

Let us begin by introducing some terminology and notation, and recalling some facts concerning ultrafilters.

**Definition 4.19.** Let  $S$  be a non-empty set. A set  $\mathbf{u} \subseteq \mathcal{P}(S)$  is called an *ultrafilter over  $S$*  if  $S \in \mathbf{u}$ ,  $U_1, U_2 \in \mathbf{u}$  implies  $U_1 \cap U_2 \in \mathbf{u}$ ,  $U_1 \in \mathbf{u}$  and  $U_1 \subseteq U_2 \subseteq S$  implies  $U_2 \in \mathbf{u}$ , and for



all  $U \subseteq S$  we have:  $U \in \mathbf{u}$  iff  $S \setminus U \notin \mathbf{u}$ . The collection of ultrafilters over  $S$  will be denoted by  $\text{Uf}(S)$ . For a set  $S$  and a subset  $U \subseteq S$ , we define

$$\hat{U} := \{\mathbf{u} \in \text{Uf}(S) \mid U \in \mathbf{u}\}.$$

For a set  $S$  and  $s \in S$ , we define

$$\mathbf{u}_s := \{U \subseteq S \mid s \in U\}.$$

It can easily be confirmed that  $\mathbf{u}_s \in \text{Uf}(S)$ . The induced map  $\mathbf{u}: S \rightarrow \text{Uf}(S)$  is called the *principal ultrafilter map* and  $\mathbf{u}_s$  is the *principal ultrafilter* generated by  $s$ .

The duality between Stone spaces and Boolean algebras gives rise to the following two contravariant functors.  $\mathbb{P}: \mathbf{Set}^{\text{op}} \rightarrow \mathbf{BA}$  maps a set  $X$  to its Boolean algebra of subsets. The functor  $\mathbb{U}: \mathbf{BA} \rightarrow \mathbf{Set}^{\text{op}}$  maps a Boolean algebra to the set of its ultrafilters. Both functors can be regarded as subfunctors of the contravariant powerset functor  $2$ , as they both map a morphism  $f$  in their respective categories to the inverse image function  $f^{-1}$ . Composing these functors, we find that for a set  $X$ ,  $\mathbb{U}\mathbb{P}(X) = \text{Uf}(X)$ , and for a function  $f: X \rightarrow Y$ ,  $\mathbb{U}\mathbb{P}(f) = (f^{-1})^{-1}$ . Hence  $\text{Uf}$  can be regarded as a subfunctor of  $2^2$ .

The following definition of ultrafilter extensions of neighbourhood models is obtained by instantiating the corresponding coalgebraic notion for  $\mathbf{F}$ -coalgebras in [27] to the case that  $\mathbf{F} = 2^2$ . We sketch the main ideas of the construction in Remark 4.21 below. In fact, the definition of the neighbourhood relation of the ultrafilter extension goes back to the definition of the canonical neighbourhood model in [42].

**Definition 4.20.** Let  $\mathcal{M} = \langle S, \nu, V \rangle$  be a neighbourhood model. The *ultrafilter extension* of  $\mathcal{M}$  is defined as the triple  $\mathcal{M}^u := \langle \text{Uf}(S), \mu, V^u \rangle$ , where

- $\text{Uf}(S)$  is the set of ultrafilters over the set  $S$ ,
- $\mu: \text{Uf}(S) \rightarrow 2^2(\text{Uf}(S))$  is defined by

$$\mu(\mathbf{u}) := \{\hat{U} \subseteq \text{Uf}(S) \mid U \subseteq S, \boxtimes U \in \mathbf{u}\},$$

where for any  $U \subseteq S$  we put  $\boxtimes U := \{s \in S \mid U \in \nu(s)\}$ ,

- $V^u(p) := \{\mathbf{u} \in \text{Uf}(S) \mid V(p) \in \mathbf{u}\}$ .

**Remark 4.21.** In [27] the neighbourhood functor  $2^2$  is denoted by  $\mathcal{H}$ . Given the coalgebraic modal logic for neighbourhood frames with one predicate lifting for the interpretation of the  $\square$ -operator (see Remark 4.16) one can define a functor  $\mathbb{L}: \mathbf{BA} \rightarrow \mathbf{BA}$  such that the category of  $\mathbb{L}$ -algebras provides the algebraic semantics of the logic. For a Boolean algebra  $\mathcal{A} = \langle A, +, -, 0 \rangle$ ,  $\mathbb{L}(\mathcal{A})$  is the free Boolean algebra generated by  $\{\square a \mid a \in A\}$ . Let  $\mathbf{Alg}(\mathbb{L})$  be the category of  $\mathbb{L}$ -algebras over  $\mathbf{BA}$ . The functors  $\mathbb{P}: \mathbf{Set}^{\text{op}} \rightarrow \mathbf{BA}$  and  $\mathbb{U}: \mathbf{BA} \rightarrow \mathbf{Set}^{\text{op}}$  are extended to functors  $\bar{\mathbb{P}}: \mathbf{Coalg}(2^2)^{\text{op}} \rightarrow \mathbf{Alg}(\mathbb{L})$  and  $\bar{\mathbb{U}}: \mathbf{Alg}(\mathbb{L}) \rightarrow \mathbf{Coalg}(2^2)^{\text{op}}$ . The ultrafilter extension of a  $2^2$ -coalgebra  $\langle S, \nu \rangle$  is then obtained as  $\bar{\mathbb{U}}\bar{\mathbb{P}}(\langle S, \nu \rangle)$ . The lifting of  $\mathbb{P}$  and  $\mathbb{U}$  relies on the existence of two natural transformations:  $\delta: \mathbb{L}\mathbb{P} \rightarrow \mathbb{P}2^2$  and  $h: \mathbb{U}\mathbb{L} \rightarrow 2^2\mathbb{U}$  whose components at a set  $X$  are defined as follows (cf. Def. 2.6.5 and Ex. 3.6 of [27]):

$$\begin{aligned} \delta_X(\square U) &= \{N \in 2^2(X) \mid U \in N\} \\ h_X(\mathbf{u}) &= \{\hat{U} \subseteq \mathbb{U}\mathbb{P}(X) \mid \square U \in U\} \end{aligned}$$

The liftings  $\bar{\mathbb{P}}$  and  $\bar{\mathbb{U}}$  are now given as follows on objects:  $\bar{\mathbb{P}}$  maps a  $2^2$ -coalgebra  $\langle X, \nu \rangle$  to  $\bar{\mathbb{P}}(\langle X, \nu \rangle) = \langle \mathbb{L}\mathbb{P}(X), \mathbb{P}(\nu) \circ \delta_X \rangle$  as illustrated here:

$$\mathbb{L}\mathbb{P}(X) \xrightarrow{\delta_X} \mathbb{P}2^2(X) \xrightarrow{\mathbb{P}(\nu)} \mathbb{P}(X)$$

$\bar{\mathbb{U}}$  maps a  $\langle \mathcal{A}, \alpha \rangle$  in  $\mathbf{Alg}(\mathbb{L})$  to  $\bar{\mathbb{U}}(\langle \mathcal{A}, \alpha \rangle) = \langle \mathbb{U}(\mathcal{A}), h_{\mathcal{A}} \circ \mathbb{U}(\alpha) \rangle$ :

$$\mathbb{U}(\mathcal{A}) \xrightarrow{\mathbb{U}(\alpha)} \mathbb{UL}(\mathcal{A}) \xrightarrow{h_{\mathcal{A}}} 2^2(\mathbb{U}(\mathcal{A}))$$

By working out the details, the reader can now confirm that the composition  $\bar{\mathbb{U}}\bar{\mathbb{P}}$  yields the ultrafilter extension of neighbourhood frames provided in Definition 4.20.

The construction of the ultrafilter extension in Definition 4.20 can be seen as an extension of the  $\mathbf{Set}$ -functor  $\mathbf{Uf}: \mathbf{Set} \rightarrow \mathbf{Set}$  to a functor  $(\_)^u: \mathbf{Nbhd} \rightarrow \mathbf{Nbhd}$  such that for any neighbourhood model  $\mathcal{M}$ , the principal ultrafilter map  $\mathbf{u}$  is truth-preserving injective map from  $\mathcal{M}$  into  $\mathcal{M}^u$ . In order to see that the construction  $(\_)^u$  of the ultrafilter extension is functorial we show that bounded morphisms between neighbourhood models induce bounded morphisms between the corresponding ultrafilter extensions.

**Lemma 4.22.** *Let  $\mathcal{M}_1 = \langle S_1, \nu_1, V_1 \rangle$  and  $\mathcal{M}_2 = \langle S_2, \nu_2, V_2 \rangle$  be neighbourhood models and let  $f: S_1 \rightarrow S_2$  be a bounded morphism from  $\mathcal{M}_1$  to  $\mathcal{M}_2$ . The function  $f^u := \mathbf{Uf}(f)$  is a bounded morphism from  $\mathcal{M}_1^u = \langle \mathbf{Uf}(S_1), \mu_1, V_1^u \rangle$  to  $\mathcal{M}_2^u = \langle \mathbf{Uf}(S_2), \mu_2, V_2^u \rangle$ .*

*Proof.* It can easily be confirmed that for any subset  $U \subseteq S_2$ :  $(f^u)^{-1}[\hat{U}] = \widehat{f^{-1}[U]}$  and  $f^{-1}[\boxtimes U] = \boxtimes(f^{-1}[U])$ . To prove that  $f^u$  is a bounded morphism, let  $\mathbf{u} \in \mathbf{Uf}(S_1)$  and  $U \subseteq S_2$ . We now have:

$$\begin{aligned} \hat{U} \in \mu_2(f^u(\mathbf{u})) &\text{ iff } \boxtimes U \in f^u(\mathbf{u}) = 2^2(f)(\mathbf{u}) \\ &\text{ iff } f^{-1}[\boxtimes U] = \boxtimes(f^{-1}[U]) \in \mathbf{u} \\ &\text{ iff } \widehat{f^{-1}[U]} = (f^u)^{-1}[\hat{U}] \in \mu_1(\mathbf{u}). \end{aligned}$$

Moreover,  $f^u$  respects valuations:  $V_1(p) \in \mathbf{u}$  iff  $f^{-1}[V_2(p)] \in \mathbf{u}$  iff  $V_2(p) \in f^u(\mathbf{u})$ .  $\square$

The next proposition connects truth of a modal formula in the ultrafilter extension to the truth set of the formula in the original model.

**Proposition 4.23.** *Let  $\mathcal{M} = \langle S, \nu, V \rangle$  be a neighbourhood model with ultrafilter extension  $\mathcal{M}^u$ . For all  $\mathbf{u} \in \mathbf{Uf}(S)$  and for all formulas  $\varphi \in \mathcal{L}$  we have*

$$\mathcal{M}^u, \mathbf{u} \models \varphi \text{ iff } \llbracket \varphi \rrbracket^{\mathcal{M}} \in \mathbf{u}.$$

*Proof.* The standard proof is obtained by induction on the formula  $\varphi$ . Details are left to the reader.  $\square$

Using Proposition 4.23, we now easily show that the principal ultrafilter map  $\mathbf{u}$  preserves the truth of modal formulas. However, it is important to note that, in general,  $\mathbf{u}$  is not a bounded morphism from a model  $\mathcal{M} = \langle S, \nu, V \rangle$  to its ultrafilter extension  $\mathcal{M}^u$ .

**Lemma 4.24.** *Let  $\mathcal{M} = \langle S, \nu, V \rangle$  be a neighbourhood model with ultrafilter extension  $\mathcal{M}^u = \langle \mathbf{Uf}(S), \mu, V^u \rangle$  and let  $\mathbf{u}: S \rightarrow \mathbf{Uf}(S)$  be the injective map from  $S$  to  $\mathbf{Uf}(S)$ . For every modal formula  $\varphi$  we have  $\mathcal{M}, s \models \varphi$  iff  $\mathcal{M}^u, \mathbf{u}_s \models \varphi$ .*

*Proof.* Let  $s \in S$  and let  $\varphi$  be modal formula. Then  $\mathcal{M}, s \models \varphi$  iff  $s \in \llbracket \varphi \rrbracket^{\mathcal{M}}$  iff  $\llbracket \varphi \rrbracket^{\mathcal{M}} \in \mathbf{u}_s$  iff  $\mathcal{M}^u, \mathbf{u}_s \models \varphi$  where the last equivalence is a consequence of Prop. 4.23.  $\square$

Another consequence of Proposition 4.23 is the fact that ultrafilter extensions are modally saturated.

**Proposition 4.25.** *For any neighbourhood model  $\mathcal{M}$ , the ultrafilter extension  $\mathcal{M}^u$  is modally saturated.*

*Proof.* Let  $\mathcal{M} = \langle S, \nu, V \rangle$  and  $\mathcal{M}^u = \langle \text{Uf}(S), \mu, V^u \rangle$ . We show that any  $\hat{U} \subseteq \text{Uf}(S)$  is compact. This suffices since all neighbourhoods in  $\mathcal{M}^u$  are of the form  $\hat{U} \subseteq \text{Uf}(S)$  and for any  $\hat{U}$ ,  $\text{Uf}(S) \setminus \hat{U} = \widehat{U}^c$ . Let  $\Psi$  be a set of formulas with the property that  $\Psi$  is finitely satisfiable in  $\hat{U}$ . For any finite set of formulas  $\{\psi_1, \dots, \psi_n\} \subseteq \Psi$  there exists therefore an ultrafilter  $\mathbf{u} \in \hat{U}$  such that  $\mathcal{M}^u, \mathbf{u} \models \psi_1 \wedge \dots \wedge \psi_n$ . This implies by Prop. 4.23 that

$$\{[\psi_1]^{\mathcal{M}}, \dots, [\psi_n]^{\mathcal{M}}\} \cup \{U\} \subseteq \mathbf{u}$$

Since  $\mathbf{u}$  is closed under finite intersections this implies  $[\psi_1]^{\mathcal{M}} \cap \dots \cap [\psi_n]^{\mathcal{M}} \cap U \in \mathbf{u}$  and hence  $[\psi_1]^{\mathcal{M}} \cap \dots \cap [\psi_n]^{\mathcal{M}} \cap U \neq \emptyset$ . As the set  $\{\psi_1, \dots, \psi_n\}$  was arbitrary we conclude that the set  $X := \{U\} \cup \{[\psi]^{\mathcal{M}} \mid \psi \in \Psi\}$  has the finite intersection property. Hence by the ultrafilter theorem, there exists some ultrafilter  $\mathbf{u}' \in \text{Uf}(S)$  such that  $X \subseteq \mathbf{u}'$ . By construction we get  $\mathbf{u}' \in \hat{U}$  and again by Prop. 4.23, that  $\Psi$  is satisfiable at  $\mathbf{u}' \in \hat{U}$ .  $\square$

We are now able to prove that the class of ultrafilter extensions of neighbourhood models is a Hennessy-Milner class.

**Proposition 4.26.** *The class  $\mathbf{U} := \{\mathcal{M}^u \mid \mathcal{M} \in \text{Nbhd}\}$  of ultrafilter extensions of neighbourhood models is a Hennessy-Milner class.*

*Proof.* Let  $\mathcal{M}_1$  and  $\mathcal{M}_2$  be arbitrary neighbourhood models. By Lemma 4.2 it suffices to show that modal equivalence is a congruence on the disjoint union  $\mathcal{M}_1^u + \mathcal{M}_2^u$  of their ultrafilter extensions. By Proposition 4.25,  $(\mathcal{M}_1 + \mathcal{M}_2)^u$  is modally saturated, hence the quotient map  $\varepsilon: (\mathcal{M}_1 + \mathcal{M}_2)^u \rightarrow (\mathcal{M}_1 + \mathcal{M}_2)^u / \equiv$  is a bounded morphism. Furthermore, denote by  $\iota_i: \mathcal{M}_i \rightarrow \mathcal{M}_1 + \mathcal{M}_2$ ,  $i \in \{1, 2\}$ , the canonical inclusion morphisms. By Lemma 4.22,  $\iota_i^u: \mathcal{M}_i^u \rightarrow (\mathcal{M}_1 + \mathcal{M}_2)^u$ ,  $i \in \{1, 2\}$ , are bounded morphisms, hence there exists, by the universal property of the disjoint union  $\mathcal{M}_1^u + \mathcal{M}_2^u$ , a bounded morphism  $g$  such that the following diagram commutes:

$$\begin{array}{ccccc} \mathcal{M}_1 & \dashrightarrow & \mathcal{M}_1^u & \hookrightarrow & \mathcal{M}_1^u + \mathcal{M}_2^u & \longleftarrow & \mathcal{M}_2^u & \dashleftarrow & \mathcal{M}_2 \\ & & \searrow & & \downarrow g & & \swarrow & & \\ & & \iota_1^u & & (\mathcal{M}_1 + \mathcal{M}_2)^u & & \iota_2^u & & \\ & & & & \downarrow \varepsilon & & & & \\ & & & & (\mathcal{M}_1 + \mathcal{M}_2)^u / \equiv & & & & \end{array}$$

Hence  $\varepsilon \circ g: \mathcal{M}_1^u + \mathcal{M}_2^u \rightarrow (\mathcal{M}_1 + \mathcal{M}_2)^u / \equiv$  is a bounded morphism, and two ultrafilters in  $\mathcal{M}_1^u + \mathcal{M}_2^u$  are modally equivalent if and only if they are identified by  $\varepsilon \circ g$ . It follows that on  $\mathcal{M}_1^u + \mathcal{M}_2^u$ , the modal equivalence relation is the kernel of  $\varepsilon \circ g$ , and hence a congruence.  $\square$

As a corollary we obtain the behavioural-equivalence-somewhere-else result.

**Theorem 4.27.** *Let  $\mathcal{M}_1 = \langle S_1, \nu_1, V_1 \rangle$  and  $\mathcal{M}_2 = \langle S_2, \nu_2, V_2 \rangle$  be neighbourhood models with the respective ultrafilter extensions  $\mathcal{M}_1^u$  and  $\mathcal{M}_2^u$ . For all states  $s_1 \in S_1$  and  $s_2 \in S_2$  we have*

$$\mathcal{M}_1, s_1 \equiv \mathcal{M}_2, s_2 \quad \Rightarrow \quad \mathcal{M}_1^u, \mathbf{u}_{s_1} \leftrightarrow_b \mathcal{M}_2^u, \mathbf{u}_{s_2}.$$

*Proof.* Let  $s_1$  and  $s_2$  be modally equivalent states in  $\mathcal{M}_1$  and  $\mathcal{M}_2$ , respectively. By Lemma 4.24 the states  $u_{s_1}$  and  $u_{s_2}$  of the ultrafilter extensions  $\mathcal{M}_1^u$  and  $\mathcal{M}_2^u$  are modally equivalent as well. The claim is now a direct consequence of Prop 4.26.  $\square$

## 5. MODEL-THEORETIC RESULTS

**5.1. The classical modal fragment of first-order logic.** We will now prove that the three equivalence notions described in section 3 all characterise the modal fragment of first-order logic over the class of neighbourhood models (Theorem 5.5). This result is an analogue of Van Benthem’s characterisation theorem for normal modal logic (cf. [8]): *On the class of Kripke models, modal logic is the Kripke bisimulation-invariant fragment of first-order logic.* It is well known that, when interpreted over Kripke models, the basic modal language  $\mathcal{L}$  can be seen as a fragment of a first-order language which has a binary predicate  $R_\square$ , and a unary predicate  $P$  for each atomic proposition  $p$  in the modal language. Formulas of this first-order language can be interpreted in Kripke models in the obvious way. Van Benthem’s theorem tells us that a first-order formula  $\alpha(x)$  is invariant under Kripke bisimulation if and only if  $\alpha(x)$  is equivalent to a modal formula.

The first step towards a Van Benthem-style characterisation theorem for classical modal logic is to show how  $\mathcal{L}$  can be viewed as a fragment of first-order logic. We will translate modal formulas into a two-sorted first-order language  $\mathcal{L}_1$ , which has previously been employed in proving a Van Benthem style characterisation theorems for topological modal logic [11] and monotonic modal logic [36], and for reasoning about topological models more generally [15]. In Remark 5.8) we will give a more detailed comparison between our characterisation theorem and the characterisation theorem for monotonic modal logic given in [36]. The two sorts of the language  $\mathcal{L}_1$  are denoted  $s$  and  $n$ . Terms of sort  $s$  are intended to represent states, whereas terms of sort  $n$  are intended to represent neighbourhoods. We assume there are countable sets of variables of each sort. To simplify notation, we will not state the type of variables explicitly. Instead we use the following conventions:  $x, y, x', y', x_1, y_2, \dots$  denote variables of sort  $s$  (*state variables*) and  $u, v, u', v', u_1, v_1, \dots$  denote variables of sort  $n$  (*neighbourhood variables*). Furthermore, the language  $\mathcal{L}_1$  contains a unary predicate  $P_i$  (of sort  $s$ ) for each  $i \in \omega$ , a binary relation symbol  $N$  relating elements of sort  $s$  to elements of sort  $n$ , and a binary relation symbol  $E$  relating elements of sort  $n$  to elements of sort  $s$ . The intended interpretation of  $xNu$  is “ $u$  is a neighbourhood of  $x$ ”, and the intended interpretation of  $uEx$  is “ $x$  is an element of  $u$ ”. The language  $\mathcal{L}_1$  is generated by the following grammar:

$$\varphi, \psi ::= x = y \mid u = v \mid P_i x \mid xNu \mid uEx \mid \neg \varphi \mid \varphi \wedge \psi \mid \exists x \varphi \mid \exists u \varphi$$

where  $i \in \omega$ ;  $x$  and  $y$  are state variables of sort  $s$ ; and  $u$  and  $v$  are neighbourhood variables of sort  $n$ . The usual abbreviations (eg.  $\forall$  for  $\neg \exists \neg$ ) apply.

Formulas of  $\mathcal{L}_1$  are interpreted in two-sorted first-order structures of the type  $\mathfrak{M} = \langle D^s, D^n, \{P_i \mid i \in \omega\}, N, E \rangle$  where  $D^s$  and  $D^n$  are the carrier sets of sort  $s$  and sort  $n$ , respectively, and each  $P_i \subseteq D^s$ ,  $N \subseteq D^s \times D^n$  and  $E \subseteq D^n \times D^s$ . The usual definitions of free and bound variables apply. Truth of sentences (formulas with no free variables)  $\varphi \in \mathcal{L}_1$  in a structure  $\mathfrak{M}$  (denoted  $\mathfrak{M} \models \varphi$ ) is defined as expected. If  $x$  is a free state variable in  $\varphi$  (denoted  $\varphi(x)$ ), then we write  $\mathfrak{M} \models \varphi[s]$  to mean that  $\varphi$  is true in  $\mathfrak{M}$  when  $s \in D^s$  is assigned to  $x$ . Note that  $\mathfrak{M} \models \exists x \varphi$  iff there is an element  $s \in D^s$  such that  $\mathfrak{M} \models \varphi[s]$ . If

$\Psi$  is a set of  $\mathcal{L}_1$ -formulas, and  $\mathfrak{M}$  is an  $\mathcal{L}_1$ -model, then  $\mathfrak{M} \models \Psi$  means that for all  $\psi \in \Psi$ ,  $\mathfrak{M} \models \psi$ . Given a class  $\mathbf{K}$  of  $\mathcal{L}_1$ -models, we denote the *semantic consequence relation over  $\mathbf{K}$*  by  $\models_{\mathbf{K}}$ . In particular, for  $\Psi(x) \cup \{\varphi(x)\} \subseteq \mathcal{L}_1$ ,  $\Psi(x) \models_{\mathbf{K}} \varphi(x)$  if for all  $\mathfrak{M} \in \mathbf{K}$  and all  $s$  of sort  $s$  in  $\mathfrak{M}$ ,  $\mathfrak{M} \models \Psi[s]$  implies  $\mathfrak{M} \models \varphi[s]$ . Moreover, a set of formulas  $\Phi(x)$  is  $\mathbf{K}$ -consistent ( $\Phi(x) \not\models_{\mathbf{K}} \perp$ ) if there exists an  $\mathfrak{M} \in \mathbf{K}$  and an  $s$  of sort  $s$  in  $\mathfrak{M}$  such that  $\mathfrak{M} \models \Phi[s]$ .

We can now translate modal  $\mathcal{L}$ -formulas and neighbourhood models to the first-order setting in a natural way:

**Definition 5.1.** Let  $\mathcal{M} = \langle S, \nu, V \rangle$  be a neighbourhood model. The *first-order translation* of  $\mathcal{M}$  is the structure  $\mathcal{M}^\circ = \langle D^s, D^n, \{P_i \mid i \in \omega\}, R_\nu, R_\exists \rangle$  where

- $D^s = S$ ,  $D^n = \nu[S] = \bigcup_{s \in S} \nu(s)$
- $P_i = V(p_i)$  for each  $i \in \omega$ ,
- $R_\nu = \{\langle s, U \rangle \mid s \in D^s, U \in \nu(s)\}$ ,
- $R_\exists = \{\langle U, s \rangle \mid s \in D^s, s \in U\}$ .

**Definition 5.2.** The *standard translation* of the basic modal language is a family of functions  $st_x : \mathcal{L} \rightarrow \mathcal{L}_1$  defined as follows:  $st_x(\perp) = \neg(x = x)$ ,  $st_x(p_i) = P_i x$ ,  $st_x(\neg\varphi) = \neg st_x(\varphi)$ ,  $st_x(\varphi \wedge \psi) = st_x(\varphi) \wedge st_x(\psi)$ , and

$$st_x(\Box\varphi) = \exists u(xNu \wedge (\forall y(uEy \leftrightarrow st_y(\varphi)))).$$

This translation preserves truth; the easy proof is left to the reader.

**Lemma 5.3.** *Let  $\mathcal{M}$  be a neighbourhood model and  $\varphi \in \mathcal{L}$ . For each  $s \in S$ ,  $\mathcal{M}, s \models \varphi$  iff  $\mathcal{M}^\circ \models st_x(\varphi)[s]$ .*

In the Kripke case, every first-order model for the language with  $R_\Box$  can be seen as Kripke model. However, it is not the case that every  $\mathcal{L}_1$ -structure is the translation of a neighbourhood model. Luckily, we can axiomatize the subclass of neighbourhood models up to isomorphism. Let NAX be the following axioms

- (A1):  $\forall u \exists x(xNu)$
- (A2):  $\forall u, v((\forall x(uEx \leftrightarrow vEx)) \rightarrow u = v)$

It is not hard to see that if  $\mathcal{M}$  is a neighbourhood model, then  $\mathcal{M}^\circ \models \text{NAX}$ . The next result states that, in fact, NAX completely characterises the class  $\mathbf{N} := \{\mathfrak{M} \mid \mathfrak{M} \cong \mathcal{M}^\circ \text{ for some neighbourhood model } \mathcal{M}\}$ , where  $\cong$  denotes isomorphism of  $\mathcal{L}_1$ -models.

**Proposition 5.4.** *Suppose  $\mathfrak{M}$  is an  $\mathcal{L}_1$ -model and  $\mathfrak{M} \models \text{NAX}$ . Then there is a neighbourhood model  $\mathfrak{M}_\circ$  such that  $\mathfrak{M} \cong (\mathfrak{M}_\circ)^\circ$ .*

*Proof.* Let  $\mathfrak{M} = \langle D^s, D^n, \{P_i \mid i \in \omega\}, N, E \rangle$  be an  $\mathcal{L}_1$ -model such that  $\mathfrak{M} \models \text{NAX}$ . We will construct from  $\mathfrak{M}$  a neighbourhood model  $\mathfrak{M}_\circ = \langle S, \nu, V \rangle$  such that  $\mathfrak{M} \cong (\mathfrak{M}_\circ)^\circ$ . In case  $D^s = \emptyset$  we also have  $D^n = \emptyset$  by axiom A1 and hence we define  $\mathfrak{M}_\circ$  to be the empty neighbourhood model. In the case  $D^s \neq \emptyset$  we first define a map  $\eta : D^n \rightarrow \mathcal{P}(D^s)$  by  $\eta(u) = \{s \in D^s \mid uEs\}$ . We take  $S = D^s$ . Now define for each  $s \in S$  and each  $X \subseteq S$ :  $X \in \nu(s)$  iff there is a  $u \in D^n$  such that  $sNu$  and  $X = \eta(u)$ , and define for all  $i \in \omega$ ,  $V(p_i) = \{s \in S \mid \mathfrak{M} \models P_i[s]\}$ . Then  $\mathfrak{M}_\circ$  is clearly a well-defined neighbourhood model, and it is not hard to see that the maps  $\text{id} : D^s \rightarrow D^s$  and  $\eta : D^n \rightarrow \bigcup_{s \in D^s} \nu(s)$  yield an isomorphism from  $\mathfrak{M}$  to  $(\mathfrak{M}_\circ)^\circ = \langle S, \nu[S], \{P'_i \mid i \in \omega\}, R_\nu, R_\exists \rangle$  (cf. Definition 5.1). The details are left to the reader.  $\square$

Thus, in a precise way, we can think of models in  $\mathbf{N}$  as neighbourhood models. In particular, if  $\mathfrak{M}$  and  $\mathfrak{N}$  are in  $\mathbf{N}$  we will write  $\mathfrak{M} + \mathfrak{N}$  by which we (strictly speaking) mean the  $\mathcal{L}_1$ -model  $(\mathfrak{M}_\circ + \mathfrak{N}_\circ)^\circ$  (which is also in  $\mathbf{N}$ ). Furthermore, Proposition 5.4 implies that we can work relative to  $\mathbf{N}$  while still preserving nice first-order properties such as compactness and the existence of countably saturated models. These properties are essential in the proof of Theorem 5.5.

**5.2. Characterisation theorem.** We are now able to formulate our characterisation theorem. Let  $\sim$  be a relation on model-state pairs. Over the class  $\mathbf{N}$ , an  $\mathcal{L}_1$ -formula  $\alpha(x)$  is *invariant under  $\sim$* , if for all models  $\mathfrak{M}_1$  and  $\mathfrak{M}_2$  in  $\mathbf{N}$  and all sort  $\mathbf{s}$ -domain elements  $s_1$  and  $s_2$  of  $\mathfrak{M}_1$  and  $\mathfrak{M}_2$ , respectively, we have  $\mathfrak{M}_1, s_1 \sim \mathfrak{M}_2, s_2$  implies  $\mathfrak{M}_1 \models \alpha[s_1]$  iff  $\mathfrak{M}_2 \models \alpha[s_2]$ . Over the class  $\mathbf{N}$ , an  $\mathcal{L}_1$ -formula  $\alpha(x)$  is *equivalent to the translation of a modal formula* if there is a modal formula  $\varphi \in \mathcal{L}$  such that for all models  $\mathfrak{M}$  in  $\mathbf{N}$ , and all  $\mathbf{s}$ -domain elements  $s$  in  $\mathfrak{M}$ ,  $\mathfrak{M} \models \alpha[s]$  iff  $\mathfrak{M} \models st_x(\varphi)[s]$ .

**Theorem 5.5.** *Let  $\alpha(x)$  be an  $\mathcal{L}_1$ -formula. Over the class  $\mathbf{N}$  the following are equivalent:*

- (1)  $\alpha(x)$  is equivalent to the translation of a modal formula,
- (2)  $\alpha(x)$  is invariant under behavioural equivalence,
- (3)  $\alpha(x)$  is invariant under precocongruences,
- (4)  $\alpha(x)$  is invariant under  $2^2$ -bisimilarity.

Our proof of Theorem 5.5 uses essentially the same ingredients as the proof of Van Benthem's theorem (see e.g. [10]) where the main steps are:

- (1) Given a Kripke model  $\mathcal{M}$  we can obtain a modally saturated, elementary extension  $\mathcal{M}^*$  of  $\mathcal{M}$ .
- (2) Between modally saturated Kripke models, modal equivalence is a Kripke bisimulation.

Together, 1 and 2 imply that modally equivalent states  $\mathcal{M}, s$  and  $\mathcal{N}, t$  are Kripke bisimilar in their modally saturated, elementary extensions  $\mathcal{M}^*, s^*$  and  $\mathcal{N}^*, t^*$ . Our analogue of 2 is that in a modally saturated neighbourhood model, modal equivalence is a congruence, which we have shown in Proposition 4.6. If we can show an analogue of 1, it follows that if  $\mathcal{M}, s$  and  $\mathcal{N}, t$  are modally equivalent, then they have behaviourally equivalent representatives in a modally saturated, elementary extension of  $\mathcal{M} + \mathcal{N}$ .

As in the Kripke case, we can obtain an  $\omega$ -saturated, elementary extension of any  $\mathcal{L}_1$ -model in the form of an ultrapower using standard first-order logic techniques (see e.g. [12]). It then only remains to show that an  $\omega$ -saturated neighbourhood model (viewed as a  $\mathcal{L}_1$ -model) is modally saturated. Before we state and prove this lemma, we recall (cf. [12]) the definition of  $\omega$ -saturation. Let  $\mathfrak{M}$  be a first-order  $\mathcal{L}_1$ -model with domain  $M$ . For a subset  $C \subseteq M$ , the  $C$ -*expansion*  $\mathcal{L}_1[C]$  of  $\mathcal{L}_1$  is the two-sorted first-order language obtained from  $\mathcal{L}_1$  by adding a constant  $\underline{c}$  for each  $c \in C$ . Now  $\mathcal{L}_1[C]$ -formulas are interpreted in  $\mathfrak{M}$  by requiring that a new constant  $\underline{c}$  is interpreted as the element  $c$ . The  $\mathcal{L}_1$ -model  $\mathfrak{M}$  is  $\omega$ -*saturated*, if for every finite  $C \subseteq_\omega M$ , and every collection  $\Gamma(x)$  of  $\mathcal{L}_1[C]$ -formulas with one free variable  $x$  the following holds: If  $\Gamma(x)$  is finitely satisfiable in  $\mathfrak{M}$  (equivalently, if  $\Gamma(x)$  is consistent with the  $\mathcal{L}_1[C]$  theory of  $\mathfrak{M}$ ), then  $\Gamma(x)$  is satisfiable in  $\mathfrak{M}$ . It is a classic result of model theory that every model has an  $\omega$ -saturated elementary extension (cf. [12]).

**Lemma 5.6.** *Let  $\mathfrak{M}$  be a model in  $\mathbf{N}$ , and let  $\mathfrak{M}_\circ$  be its corresponding neighbourhood model. If  $\mathfrak{M}$  is  $\omega$ -saturated, then  $\mathfrak{M}_\circ$  is modally saturated.*

*Proof.* Let  $\mathfrak{M}$  be an  $\mathcal{L}_1$ -model in  $\mathbf{N}$ ,  $\mathfrak{M}_\circ = \langle S, \nu, V \rangle$  its corresponding neighbourhood model (cf. Proposition 5.4), and assume that  $\mathfrak{M}$  is  $\omega$ -saturated. Let  $\Psi$  be a set of modal  $\mathcal{L}$ -formulas, and let  $U \subseteq S$  be a neighbourhood of some state  $s$ . Then  $U$  corresponds to a domain element  $u \in D^n$  of  $\mathfrak{M}$  via the isomorphism  $\mathfrak{M} \cong (\mathfrak{M}_\circ)^\circ$ . If  $\Psi$  is finitely satisfiable in  $U$  in  $\mathfrak{M}_\circ$ , then the set of  $\mathcal{L}_1[\{u\}]$ -formulas  $\{\underline{u}\mathbf{N}x\} \cup \{st_x(\psi) \mid \psi \in \Psi\}$  is finitely satisfiable in  $\mathfrak{M}$ , and hence satisfiable, which implies that  $\Psi$  is satisfiable in  $U$ . Similarly, if  $\Psi$  is finitely satisfiable in  $U^c$ , then the set of  $\mathcal{L}_1[\{u\}]$ -formulas  $\{\neg\underline{u}\mathbf{N}x\} \cup \{st_x(\psi) \mid \psi \in \Psi\}$  is finitely satisfiable in  $\mathfrak{M}$ , and hence satisfiable, which implies that  $\Psi$  is satisfiable in  $U^c$ .  $\square$

We are now ready to prove Theorem 5.5.

*Proof of Theorem 5.5.* It is clear that  $2 \Rightarrow 3 \Rightarrow 4$  (cf. Proposition 3.10). To see that  $4 \Rightarrow 2$ , we only need to recall (cf. [39]) that graphs of bounded morphisms are  $2^2$ -bisimulations. Furthermore, as truth of modal formulas is preserved by behavioural equivalence,  $1 \Rightarrow 2$  is clear. We complete the proof by showing that  $2 \Rightarrow 1$ .

Let  $\text{MOC}_{\mathbf{N}}(\alpha) = \{st_x(\varphi) \mid \varphi \in \mathcal{L}, \alpha(x) \models_{\mathbf{N}} st_x(\varphi)\}$  be the set of modal consequences of  $\alpha(x)$  over the class  $\mathbf{N}$ . It suffices to show that  $\text{MOC}_{\mathbf{N}}(\alpha) \models_{\mathbf{N}} \alpha(x)$ , since then by compactness there is a finite subset  $\Gamma(x) \subseteq \text{MOC}_{\mathbf{N}}(\alpha)$  such that  $\Gamma(x) \models_{\mathbf{N}} \alpha(x)$  and  $\alpha(x) \models_{\mathbf{N}} \bigwedge \Gamma(x)$ . It follows that over  $\mathbf{N}$ ,  $\alpha(x)$  is equivalent to  $\bigwedge \Gamma(x)$ , which is the translation of a modal formula. So suppose  $\mathfrak{M}$  is a model in  $\mathbf{N}$  and  $\text{MOC}_{\mathbf{N}}(\alpha)$  is satisfied at some element  $s$  in  $\mathfrak{M}$ . We must show that  $\mathfrak{M} \models \alpha[s]$ . Consider the set  $T(x) = \{st_x(\varphi) \mid \mathfrak{M}_\circ, s \models \varphi\} \cup \{\alpha(x)\}$ .  $T(x)$  is  $\mathbf{N}$ -consistent, since suppose to the contrary that  $T(x)$  is  $\mathbf{N}$ -inconsistent, then by compactness, there is a finite collection of modal formulas  $\varphi_1, \dots, \varphi_n$  such that  $\mathfrak{M}_\circ, s \models \varphi_i$  for all  $i = 1, \dots, n$  and  $\alpha(x) \models_{\mathbf{N}} \neg \bigwedge_{i=1}^n st_x(\varphi_i)$ , which implies that  $\neg \bigwedge_{i=1}^n st_x(\varphi_i) \in \text{MOC}_{\mathbf{N}}(\alpha)$ . But this contradicts the assumption that  $\mathfrak{M} \models \text{MOC}_{\mathbf{N}}(\alpha)[s]$  and  $\mathfrak{M} \models st_x(\varphi_i)[s]$  for all  $i = 1, \dots, n$ . Hence  $T(x)$  is satisfied at an element  $t$  in some  $\mathfrak{N} \in \mathbf{N}$ , and by construction,  $s$  and  $t$  are modally equivalent: For all modal formulas  $\varphi \in \mathcal{L}$ ,  $\mathfrak{M} \models st_x(\varphi)[s]$  implies  $st_x(\varphi) \in T(x)$ , and hence  $\mathfrak{N} \models st_x(\varphi)[t]$ . Conversely,  $\mathfrak{M} \not\models st_x(\varphi)[s]$  iff  $\mathfrak{M} \models \neg st_x(\varphi)[s]$  which implies  $st_x(\neg\varphi) = \neg st_x(\varphi) \in T(x)$ , and hence  $\mathfrak{N} \not\models st_x(\varphi)[t]$ .

Take now an  $\omega$ -saturated, elementary extension  $\mathfrak{U}$  of  $\mathfrak{M} + \mathfrak{N}$ . Note that  $\mathfrak{U} \in \mathbf{N}$ , since validity of NAX is preserved under elementary extensions. Moreover, the images  $s_U$  and  $t_U$  in  $\mathfrak{U}$  of  $s$  and  $t$ , respectively, are also modally equivalent, since modal truth is transferred by elementary maps. Now since  $\mathfrak{U}$  is  $\omega$ -saturated and thus by Lemma 5.6,  $\mathfrak{U}_\circ$  is modally saturated, it follows from Proposition 4.6 that  $s_U$  and  $t_U$  are behaviourally equivalent. The construction is illustrated in the following diagram;  $\preceq$  indicates that the map is elementary.

$$\begin{array}{ccc} \text{MOC}_{\mathbf{N}}(\alpha)[s] \models \mathfrak{M} & \xrightarrow{i} & \mathfrak{M} + \mathfrak{N} \xleftarrow{j} \mathfrak{N} \models \alpha[t] \\ & & \downarrow \preceq \\ & & \mathfrak{U} \end{array}$$

Finally, we can transfer the truth of  $\alpha(x)$  from  $\mathfrak{N}, t$  to  $\mathfrak{M}, s$  by using the invariance of modal formulas under bounded morphisms and standard translations  $(bm+st)$ ; elementary maps

(*elem*); and the assumption that  $\alpha(x)$  is invariant under behavioural equivalence ( $\alpha(x)$ -*beh-inv*).

$$\begin{aligned}
\mathfrak{N} \models \alpha[t] &\iff (\mathfrak{M}_\circ + \mathfrak{N}_\circ)^\circ \models \alpha[j(t)] && (bm+st) \\
&\iff \mathfrak{U} \models \alpha[t_U] && (elem) \\
&\iff \mathfrak{U} \models \alpha[s_U] && (s_U \leftrightarrow_b t_U \text{ and } \alpha(x)\text{-beh-inv}) \\
&\iff (\mathfrak{M}_\circ + \mathfrak{N}_\circ)^\circ \models \alpha[i(s)] && (elem) \\
&\iff \mathfrak{M} \models \alpha[s] && (bm+st)
\end{aligned}$$

□

**Remark 5.7.** Note that in the proof of Theorem 5.5, we could have assumed  $\alpha(x)$  to be invariant for any of the three equivalence notions, since Proposition 3.20 tells us that also  $s_U \leftrightarrow t_U$  and  $s_U \leftrightarrow_p t_U$ .

**Remark 5.8.** An analogue of Van Benthem’s theorem for monotonic modal logic was proved by Pauly (see [36, 20]). Although the translation of monotonic modal logic and monotonic neighbourhood models is very similar to ours, Pauly’s approach is slightly different to the present one, since his result is not formulated relative to the class of first-order models which are the translation monotonic models. Rather, he defines a notion of monotonic bisimulation which applies to all first-order  $\mathcal{L}_1$ -models, and shows that translations of monotonic modal formulas are invariant under this bisimulation notion, even if the first-order models involved are not necessarily translations of monotonic models. This means his result concerns a stronger notion of invariance. The converse is shown using  $\omega$ -saturation and monotonic modal saturation, and is similar to the proof of the Van Benthem theorem. We do not get a characterisation theorem for monotonic modal logic (relative to translations of monotonic models) as a direct corollary of Theorem 5.5, but we believe it is possible to prove one using the same line of argumentation and constructions.

**Remark 5.9.** It seems straightforward to generalise Theorem 5.5 to multi-modal classical modal logic with polyadic modalities of finite arity. Multi-modal neighbourhood models are of interest in coalgebraic modal logic due to the following:

It is not always possible to find a collection of separating unary, predicate liftings for a functor  $F: \mathbf{Set} \rightarrow \mathbf{Set}$ . However, Schröder showed in [40] that any finitary functor  $F$  has a separating set of finitary, *polyadic* predicate liftings, i.e., there exists a finitary coalgebraic modal logic with polyadic modalities which is expressive for  $F$ -coalgebras. A  $k$ -ary predicate lifting  $\lambda: (2^{(-)})^k \rightarrow 2^{F(-)}$  has transposite  $\hat{\lambda}_X: F(X) \rightarrow \mathbf{N}^k(X)$ , where  $\mathbf{N}^k$  denotes the functor  $\mathbf{N}^k = 2^{(-)} \circ (2^{(-)})^k$ . Note that a map  $X \rightarrow \mathbf{N}^k(X)$  is a  $k$ -ary neighbourhood function. If  $\Lambda$  is a separating set of  $k$ -ary predicate liftings for  $F$ , then for all sets  $X$ , the source of transposites  $\{\hat{\lambda}_X: F(X) \rightarrow \mathbf{N}^k(X) \mid \lambda \in \Lambda\}$  yields a natural embedding.

$$\langle \hat{\lambda} \rangle_{\lambda \in \Lambda}: F \rightarrow \Pi_\Lambda \mathbf{N}^k, \quad (5.1)$$

where  $\Pi_\Lambda \mathbf{N}^k$  is the  $|\Lambda|$ -fold product of  $\mathbf{N}^k$ . Hence for every finitary functor  $F$ , an  $F$ -coalgebra can be transformed into a pointwise equivalent multi-modal, polyadic neighbourhood frame.

**5.3. Interpolation.** In this section we show that the results on ultrafilter extensions from the previous section can be used to prove Craig interpolation for classical modal logic. For several normal and monotonic modal logics, Craig interpolation can be proved using superamalgamation in the corresponding variety of modal algebras, see e.g. [16, 21, 28,



29, 30]. We believe similar proofs can be carried out for classical modal logic. Our proof, however, is based on the ideas used in the proof of Craig interpolation for normal modal logic presented in [6]. The proof in [6] uses first-order model-theoretic arguments similar to those employed in the proof of the Van Benthem characterisation theorem, but Theorem 4.27 allows us to prove Craig interpolation in a purely modal setting, without the use of  $\omega$ -saturated models or the explicit use of algebraic duality. All that is needed is that modal truth is invariant under ultrafilter extensions (Lemma 4.24), and that ultrafilter extensions are modally saturated (Proposition 4.25).

So far we have worked with a fixed a set  $\text{At}$  of atomic propositions, giving rise to the language  $\mathcal{L} = \mathcal{L}(\text{At})$ . In the current section we need to generalise our notions of bounded morphism and modal saturation to sublanguages  $\mathcal{L}(\text{At}')$  of  $\mathcal{L}(\text{At})$  generated by a specific subset  $\text{At}'$  of atomic propositions. We point out that all models are always models for the full language  $\mathcal{L}(\text{At})$ . This generalisation is straightforward, but in the interest of clarity we provide the details and the exact results we need. Let  $\text{At}' \subseteq \text{At}$ , and let  $\mathcal{M}_1 = \langle S_1, \nu_1, V_1 \rangle$  and  $\mathcal{M}_2 = \langle S_2, \nu_2, V_2 \rangle$  be neighbourhood  $\mathcal{L}(\text{At})$ -models. A function  $f: S_1 \rightarrow S_2$  is a *bounded  $\mathcal{L}(\text{At}')$ -morphism from  $\mathcal{M}_1$  to  $\mathcal{M}_2$*  (notation:  $f: \mathcal{M}_1 \rightarrow_{\mathcal{L}(\text{At}')} \mathcal{M}_2$ ) if  $f$  is a bounded (frame) morphism from  $\langle S_1, \nu_1 \rangle$  to  $\langle S_2, \nu_2 \rangle$ , and for all  $p \in \text{At}'$ , and all  $s \in S_1$ :  $s \in V_1(p)$  iff  $f(s) \in V_2(p)$ . An  $\mathcal{L}(\text{At}')$ -congruence is the kernel of a bounded  $\mathcal{L}(\text{At}')$ -morphism. Two states  $s_1 \in S_1$  and  $s_2 \in S_2$  are *modally  $\mathcal{L}(\text{At}')$ -equivalent* (notation:  $s_1 \equiv_{\mathcal{L}(\text{At}')} s_2$ ), if they satisfy the same  $\mathcal{L}(\text{At}')$ -formulas. Given a neighbourhood  $\mathcal{L}(\text{At})$ -model  $\mathcal{M} = \langle S, \nu, V \rangle$ , a subset  $X \subseteq S$  is *modally  $\mathcal{L}(\text{At}')$ -compact* if for all sets  $\Psi$  of modal  $\mathcal{L}(\text{At}')$ -formulas,  $\Psi$  is satisfiable in  $X$ , whenever  $\Psi$  is finitely satisfiable in  $X$ , and  $\mathcal{M}$  is *modally  $\mathcal{L}(\text{At}')$ -saturated* if for every  $\equiv_{\mathcal{L}(\text{At}')}$ -coherent neighbourhood  $X$ , both  $X$  and  $X^c$  are modally  $\mathcal{L}(\text{At}')$ -compact.

**Lemma 5.10.** *Let  $\text{At}' \subseteq \text{At}$ .*

- (1) *If  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are  $\mathcal{L}(\text{At})$ -neighbourhood models, and  $f: \mathcal{M}_1 \rightarrow_{\mathcal{L}(\text{At}')} \mathcal{M}_2$ , then for all  $s$  in  $\mathcal{M}_1$ , and all  $\varphi \in \mathcal{L}(\text{At}')$ :  $\mathcal{M}_1, s \models \varphi$  iff  $\mathcal{M}_2, f(s) \models \varphi$ .*
- (2) *If  $\mathcal{M} = \langle S, \nu, V \rangle$  is a neighbourhood  $\mathcal{L}(\text{At})$ -model, and  $R \subseteq S \times S$  is an equivalence relation, then  $R$  is an  $\mathcal{L}(\text{At}')$ -congruence on  $\mathcal{M}$  iff  $R$  is a congruence on the underlying frame  $\langle S, \nu \rangle$ , and for all  $\langle s, t \rangle \in R$ , and all  $p \in \text{At}'$ :  $s \in V(p)$  iff  $t \in V(p)$ .*
- (3) *If a neighbourhood  $\mathcal{L}(\text{At})$ -model  $\mathcal{M}$  is modally  $\mathcal{L}(\text{At}')$ -saturated, then all  $\equiv_{\mathcal{L}(\text{At}')}$ -coherent subsets are definable by an  $\mathcal{L}(\text{At}')$ -formula.*
- (4) *If a neighbourhood  $\mathcal{L}(\text{At})$ -model  $\mathcal{M}$  is modally  $\mathcal{L}(\text{At}')$ -saturated, then  $\equiv_{\mathcal{L}(\text{At}')}$  is an  $\mathcal{L}(\text{At}')$ -congruence.*
- (5) *If  $\mathcal{M}$  is neighbourhood  $\mathcal{L}(\text{At})$ -model, then its ultrafilter extension  $\mathcal{M}^u$  is modally  $\mathcal{L}(\text{At}')$ -saturated.*

*Proof.* As usual, 1 can be proved by straightforward formula induction. Item 2 is immediate. Item 3 can be proved by retracing the argument used in Lemma 4.5. Item 4 follows from item 3 and essentially the same argument used in Lemma 4.3. Item 5 can be proved in the same way as Proposition 4.25.  $\square$

For a formula  $\varphi \in \mathcal{L}$ , we denote by  $\text{At}(\varphi)$  the set of atomic propositions occurring in  $\varphi$ . Recall that for  $\Phi \cup \{\varphi\} \subseteq \mathcal{L}$ , we write  $\Phi \models \varphi$  if  $\varphi$  is a local semantic consequence of  $\Phi$  over the class of all neighbourhood models. Note that compactness of  $\models$  follows from the compactness of  $\models_{\mathbf{N}}$ , the first-order consequence relation over the class of neighbourhood models.

**Theorem 5.11** (Craig interpolation). *Let  $\varphi_1, \varphi_2 \in \mathcal{L}$ . If  $\models \varphi_1 \rightarrow \varphi_2$ , then there exists a formula  $\chi \in \mathcal{L}$  with  $\text{At}(\chi) \subseteq \text{At}(\varphi_1) \cap \text{At}(\varphi_2)$  such that  $\models \varphi_1 \rightarrow \chi$  and  $\models \chi \rightarrow \varphi_2$ .*

*Proof.* Assume that  $\models \varphi_1 \rightarrow \varphi_2$ . Let  $\text{At}_i = \text{At}(\varphi_i)$ ,  $i = 1, 2$ , and  $\text{At}_0 = \text{At}_1 \cap \text{At}_2$ . Denote by  $\text{Cons}_{\mathcal{L}(\text{At}_0)}(\varphi_1) = \{\chi \in \mathcal{L}(\text{At}_0) \mid \varphi_1 \models \chi\}$  the set of modal  $\mathcal{L}(\text{At}_0)$ -consequences of  $\varphi_1$ . It suffices to show that  $\text{Cons}_{\mathcal{L}(\text{At}_0)}(\varphi_1) \models \varphi_2$ , since then by compactness, there are  $\chi_1, \dots, \chi_n \in \text{Cons}_{\mathcal{L}(\text{At}_0)}(\varphi_1)$  such that  $\chi_1 \wedge \dots \wedge \chi_n \models \varphi_2$ , and  $\varphi_1 \models \chi_1 \wedge \dots \wedge \chi_n$ , i.e.,  $\chi = \chi_1 \wedge \dots \wedge \chi_n$  is a Craig interpolant for  $\varphi_1 \rightarrow \varphi_2$ .

So, assume  $\mathcal{M}$  is an  $\mathcal{L}(\text{At})$ -model and  $s$  is a state in  $\mathcal{M}$  such that  $\mathcal{M}, s \models \text{Cons}_{\mathcal{L}(\text{At}_0)}(\varphi_1)$ , and let  $\Psi = \{\psi \in \mathcal{L}(\text{At}_0) \mid \mathcal{M}, s \models \psi\}$ . Now  $\Psi \cup \{\varphi_1\}$  is consistent, since otherwise there would exist  $\{\psi_1, \dots, \psi_n\} \subseteq \Psi$  such that  $\models \psi_1 \wedge \dots \wedge \psi_n \rightarrow \neg\varphi_1$ , hence  $\models \varphi_1 \rightarrow \neg\psi_1 \vee \dots \vee \neg\psi_n$ , which would imply that  $\neg\psi_1 \vee \dots \vee \neg\psi_n \in \text{Cons}_{\mathcal{L}(\text{At}_0)}(\varphi_1)$  contradicting the assumption that  $\mathcal{M}, s \models \text{Cons}_{\mathcal{L}(\text{At}_0)}(\varphi_1)$ .

By definition of  $\models$ ,  $\Psi \cup \{\varphi_1\}$  is satisfiable in some neighbourhood  $\mathcal{L}(\text{At})$ -model  $\mathcal{N}$  at a state  $t$  in  $\mathcal{N}$ , i.e.,  $\mathcal{N}, t \models \Psi \cup \{\varphi_1\}$ . Then by construction  $s \equiv_{\mathcal{L}(\text{At}_0)} t$ , and as truth is preserved by the injections  $\iota: \mathcal{M} \rightarrow \mathcal{N} + \mathcal{M}$  and  $\kappa: \mathcal{N} \rightarrow \mathcal{N} + \mathcal{M}$ , and when passing to ultrafilter extensions, the principal ultrafilters generated by  $\iota(s)$  and  $\kappa(t)$  are also modally  $\mathcal{L}(\text{At}_0)$ -equivalent in  $\mathcal{U} = \langle U, \mu, V \rangle = (\mathcal{N} + \mathcal{M})^u$ , i.e.,  $\mathbf{u}_{\iota(s)} \equiv_{\mathcal{L}(\text{At}_0)} \mathbf{u}_{\kappa(t)}$ . Now since ultrafilter extensions are modally  $\mathcal{L}(\text{At}_0)$ -saturated (Lemma 5.10(5)) it follows from Lemma 5.10(4) that  $\equiv_{\mathcal{L}(\text{At}_0)}$  is an  $\mathcal{L}(\text{At}_0)$ -congruence on  $\mathcal{U}$ . For ease of notation, we denote the relation  $\equiv_{\mathcal{L}(\text{At}_0)}$  on  $\mathcal{U}$  by  $Z$  in the rest of this proof. We have, in particular,  $Z$  is a congruence on the underlying frame  $\langle U, \mu \rangle$  of  $\mathcal{U}$ , and by Proposition 3.20  $Z$  is also a  $2^2$ -bisimulation on  $\langle U, \mu \rangle$ . This means there exists a coalgebra map  $\zeta: Z \rightarrow 2^2(Z)$  such that the projections  $\pi_i: \langle Z, \zeta \rangle \rightarrow \langle U, \mu \rangle$ ,  $i = 1, 2$ , are bounded frame morphisms. We now define a valuation  $V'$  on  $\langle Z, \zeta \rangle$  to obtain a neighbourhood  $\mathcal{L}(\text{At})$ -model  $\mathcal{Z} = \langle Z, \zeta, V' \rangle$  such that  $\pi_1: \mathcal{Z} \rightarrow \mathcal{U}$  is a bounded  $\mathcal{L}(\text{At}_1)$ -morphism and  $\pi_2: \mathcal{Z} \rightarrow \mathcal{U}$  is a bounded  $\mathcal{L}(\text{At}_2)$ -morphism. Let  $p \in \text{At}$  and  $\langle u_1, u_2 \rangle \in Z$ , then we define

$$\langle u_1, u_2 \rangle \in V'(p) \iff \begin{cases} u_1 \in V(p) & \text{if } p \in \text{At}_1, \\ u_2 \in V(p) & \text{if } p \in \text{At}_2, \\ \text{never} & \text{if } p \in \text{At} \setminus (\text{At}_1 \cup \text{At}_2). \end{cases}$$

Note that  $V'$  is well-defined due to Lemma 5.10(2). The construction is illustrated below. The dashed arrow going to  $\mathcal{U}$  indicates that the principal ultrafilter map  $\mathbf{u}$  is not a bounded morphism, still  $\mathbf{u}$  does preserve modal truth (Lemma 4.24).

$$\begin{array}{c} \varphi_1 \models \mathcal{N}, t \xrightarrow{\kappa} \mathcal{N} + \mathcal{M} \xleftarrow{\iota} \mathcal{M}, s \models \text{Cons}_{\mathcal{L}(\text{At}_0)}(\varphi_1) \\ \quad \quad \quad \downarrow \mathbf{u} \equiv \\ \quad \quad \quad \mathcal{U} \\ \quad \quad \quad \uparrow \pi_1 \quad \uparrow \pi_2 \\ \quad \quad \quad \mathcal{Z} \end{array}$$

Now we have:  $\mathcal{N}, t \models \varphi_1$  implies  $\mathcal{U}, \mathbf{u}_{\kappa(t)} \models \varphi_1$ . Since  $\langle \mathbf{u}_{\kappa(t)}, \mathbf{u}_{\iota(s)} \rangle \in Z$  and  $\pi_1$  is a bounded  $\mathcal{L}(\text{At}_1)$ -morphism from  $\mathcal{Z}$  to  $\mathcal{U}$ , we have  $\mathcal{Z}, \langle \mathbf{u}_{\kappa(t)}, \mathbf{u}_{\iota(s)} \rangle \models \varphi_1$ . By the main assumption that  $\models \varphi_1 \rightarrow \varphi_2$ , we get that  $\mathcal{Z}, \langle \mathbf{u}_{\kappa(t)}, \mathbf{u}_{\iota(s)} \rangle \models \varphi_2$ , and now since  $\pi_2$  is a bounded  $\mathcal{L}(\text{At}_2)$ -morphism from  $\mathcal{Z}$  to  $\mathcal{U}$ , we get  $\mathcal{U}, \mathbf{u}_{\iota(s)} \models \varphi_2$  and hence  $\mathcal{M}, s \models \varphi_2$ .  $\square$

## 6. CONCLUSION AND RELATED WORK

In the first part of this paper we discussed and compared different notions of equivalence between neighbourhood structures. We gave back-and-forth style characterisations of  $2^2$ -bisimulations and precocongruences, and showed that, as expected, behavioural equivalence is the only one of the three notions that allows us to prove a Hennessy-Milner theorem for image-finite neighborhood models (cf. Section 4). Furthermore, we showed that for an *arbitrary* Set-functor  $F$ , precocongruences capture behavioural equivalence on a single  $F$ -coalgebra (Theorem 3.12). For functors  $F$  that weakly preserve kernel pairs, such as  $2^2$ , this is already achieved with  $F$ -bisimulations [19], but we believe that precocongruences could be an interesting alternative to  $F$ -bisimulations for functors which lack this property. A first indication of this is [23] where precocongruences are used to obtain a game-theoretic characterisation of behavioural equivalence.

After having reached a good understanding of state equivalence over neighbourhood structures, we focused on generalising two well-known model-theoretic results to the setting of neighbourhood models: the Van Benthem Characterisation Theorem (Theorem 5.5) and Craig Interpolation (Theorem 5.11). Our proof of Theorem 5.5 builds on ideas from the original proof of the Van Benthem characterisation theorem ([8]). Closely related to our work are also the invariance results by Pauly ([36]) on monotonic modal logic, and Ten Cate et al. ([11]) on topological modal logic.

A number of other model-theoretic results are worth exploring. Perhaps the most interesting one is a generalisation of the Goldblatt-Thomason Theorem (see e.g. [10]). The classic result for Kripke models can be proved using model-theoretic constructs or by using algebraic duality. The algebraic duality proof has already been generalised to the coalgebraic setting by Kurz & Rosický's [27]. Indeed, a special case of their main result is the result we are after: a Goldblatt-Thomason Theorem for neighbourhood models (cf. [27], Corollary 3.17(2) and Remark 3.18). Given the formal machinery we have developed in this paper (e.g., the ultrafilter extensions from Section 4.3), one may hope for a model-theoretic proof of this result (see e.g., Section 3.8 in [10]). Such a model-theoretic proof has been given for topological models (which are special cases of neighbourhood models) by Ten Cate et al. ([11]). However, an important ingredient in the model-theoretic proof for the Kripke case is the fact that any Kripke model is bisimilar to the disjoint union of its generated submodels. This is not true for an arbitrary neighbourhood model (cf. [18]), and at the moment, it is not clear which alternative construction could be used in its place.

A second model-theoretic issue raised by the results in this paper concerns our translation of the modal language into a two-sorted first-order language (cf. Definition 5.2). As is well-known, with respect to Kripke structures, the basic modal language can be translated into the *guarded fragment* of first-order logic (cf. [6]). This fact has been used to explain a number of the important properties of modal logic (see, for example, [5] for an extensive discussion). The question is whether classical modal logic is also contained in some kind of guarded fragment. Our translation of  $\Box\varphi$  does not fall into the guarded fragment of two-sorted first-order logic. However, it is not difficult to see that over the class  $\mathbf{N}$  of neighbourhood models viewed as first-order structures,  $st_x(\Box\varphi)$  is equivalent to the following single-sorted first-order formula:

$$\exists u(Nbhd(u) \wedge xNu \wedge \forall y(uEy \rightarrow st_y(\varphi)) \wedge \forall y(State(y) \rightarrow (\neg(st_y(\varphi)) \vee uEy)))$$

where *Nbhd* and *State* are designated predicates intended to mean "...is a neighbourhood" and "...is a state", respectively. This formula is in the (loosely) guarded fragment.

Our characterisation theorem for classical modal logic leads to a number of interesting research questions. For example, we would like to explore the possibility of proving our result using game-theoretic techniques similar to the ones exploited by Otto ([32]). Furthermore, neighbourhood structures can also be seen as a type of Chu spaces. We would like to relate our characterisation theorem to Van Benthem's characterisation of the Chu transform invariant fragment of a two-sorted first-order logic in [9].

Finally, it would be interesting to find out if our characterisation theorem can be generalised to coalgebraic modal logic for an arbitrary finitary functor  $F: \text{Set} \rightarrow \text{Set}$ , using the embedding of  $F$ -coalgebras into multi-modal,  $k$ -ary neighbourhood frames as described in Remark 5.9. It might be possible to prove that, under certain assumptions, the coalgebraic modal logic over  $F$ -coalgebras can be viewed as the bisimulation invariant fragment of some many-sorted first-order logic. Initial investigations suggest that this is possible for functors of the form  $A^{(2^k)^{(-)}}$  where  $A$  is a finite set and  $k$  is a natural number. An  $A^{(2^k)^{(-)}}$ -coalgebra can be seen as a multi-modal, polyadic neighbourhood frame  $\langle X, \{\nu_a \mid a \in A\} \rangle$  given by an  $A$ -indexed collection of  $k$ -ary neighbourhood functions  $\nu_a: X \rightarrow 2^{(2^X)^k}$  such that for each  $k$ -tuple of subsets  $\langle U_1, \dots, U_k \rangle$  and each state  $x \in X$ ,  $\langle U_1, \dots, U_k \rangle \in \nu_a(x)$  for exactly one  $a \in A$ . We must leave the details of this result as future work.

#### ACKNOWLEDGEMENTS

We would like to thank H. Peter Gumm for many fruitful discussions, and Yde Venema for initiating our cooperation on this subject. Special thanks also goes to the anonymous referees for useful comments and corrections.

#### REFERENCES

- [1] P. Aczel and N.P. Mendler. A final coalgebra theorem. In D.H. Pitt, D.E. Rydeheard, P. Dybjer, A.M. Pitts, and A. Poigné, editors, *Category Theory and Computer Science*, volume 389 of *Lecture Notes in Computer Science*, pages 357–365, 1989.
- [2] J. Adámek, H. Herrlich, and G.E. Strecker. *Abstract and Concrete Categories: The Joy of Cats*. J. Wiley and Sons, 1990. Online version: <http://katmat.math.uni-bremen.de/acc>.
- [3] J. Adámek and H.-E. Porst. On tree coalgebras and coalgebra presentations. *Theoretical Computer Science*, 311:257–283, 2004.
- [4] R. Alur, T.A. Henzinger, and O. Kupferman. Alternating-time temporal logic. *Journal of the ACM*, 49(5):672–713, 2002.
- [5] H. Andréka, J. van Benthem, and I. Németi. Back and forth between modal logic and classical logic. *Logic Journal of the IGPL*, 3:685 – 720, 1995.
- [6] H. Andréka, J. van Benthem, and I. Németi. Modal languages and bounded fragments of predicate logic. *Journal of Philosophical Logic*, 27(3):217–274, 1998.
- [7] J. van Benthem. *Modal Correspondence Theory*. PhD thesis, Mathematical Institute, University of Amsterdam, 1976.
- [8] J. van Benthem. Correspondence theory. In D. Gabbay and F. Guenther, editors, *Extensions of Classical Logic*, volume II of *Handbook of Philosophical Logic*, pages 167–247. Reidel, Dordrecht, 1984.
- [9] J. van Benthem. Information transfer across Chu spaces. *Logic Journal of the IGPL*, 8(6):719–731, 2000.
- [10] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2001.
- [11] B. ten Cate, D. Gabelaia, and D. Sustretov. Modal languages for topology: Expressivity and definability. *To appear in Annals of Pure and Applied Logic*. Preprint available at: <http://arxiv.org/abs/math/0610357>.
- [12] C. Chang and H. Keisler. *Model Theory*. North-Holland, 1973.
- [13] B.F. Chellas. *Modal Logic - An Introduction*. Cambridge University Press, 1980.
- [14] K. Došen. Duality between modal algebras and neighbourhood frames. *Studia Logica*, 48:219–234, 1989.

- [15] J. Flum and M. Ziegler. *Topological Model Theory*, volume 769 of *Lecture Notes in Mathematics*. Springer Verlag, 1980.
- [16] D.M. Gabbay and L. Maksimova. *Interpolation and Definability: Modal and Intuitionistic Logic*. Number 46 in Oxford Logic Guides. Oxford University Press, 2005.
- [17] L. Goble. Murder most gentle: The paradox deepens. *Philosophical Studies*, 64(2):217–227, 1991.
- [18] H.P. Gumm. Functors for coalgebras. *Algebra Universalis*, 45:135–147, 2001.
- [19] H.P. Gumm and T. Schröder. Types and coalgebraic structure. *Algebra universalis*, 53:229–252, 2005.
- [20] H.H. Hansen. Monotonic modal logic (Master’s thesis). Research Report PP-2003-24, Institute for Logic, Language and Computation. University of Amsterdam, 2003.
- [21] H.H. Hansen and C. Kupke. A coalgebraic perspective on monotone modal logic. In *Proceedings of the 7th Workshop on Coalgebraic Methods in Computer Science (CMCS)*, volume 106 of *Electronic Notes in Theoretical Computer Science*, pages 121–143. Elsevier Science Publishers, 2004.
- [22] H.H. Hansen, C. Kupke, and E. Pacuit. Bisimulation for neighbourhood structures. In T. Mossakowski et al., editor, *Proceedings of the 2nd Conference on Algebra and Coalgebra in Computer Science (CALCO 2007)*, volume 4624 of *Lecture Notes in Computer Science*, pages 279–293. Springer, 2007.
- [23] C. Kupke. Terminal sequence induction via games. In *Proceedings of the 7th International Tbilisi Symposium on Language, Logic and Computation*, 2008.
- [24] C. Kupke, A. Kurz, and D. Pattinson. Ultrafilter extensions for coalgebras. In J.L. Fiadero, N. Harman, M. Roggenbach, and J.J.M.M. Rutten, editors, *Proceedings of the 2nd Workshop on Algebraic and Coalgebraic Methods in Computer Science (CALCO 2005)*, volume 3629 of *Lecture Notes in Computer Science*, pages 263–277. Springer, 2005.
- [25] A. Kurz. Personal communication.
- [26] A. Kurz. *Logics for Coalgebras and Applications to Computer Science*. PhD thesis, Ludwig-Maximilians-Universität, 2000.
- [27] A. Kurz and J. Rosický. The Goldblatt-Thomason-theorem for coalgebras. In T. Mossakowski et al., editor, *Proceedings of the 2nd Conference on Algebra and Coalgebra in Computer Science (CALCO)*, volume 4624 of *Lecture Notes in Computer Science*, pages 342–355. Springer, 2007.
- [28] J. Madarász. Interpolation and amalgamation; pushing the limits. Part I. *Studia Logica*, 61:311–345, 1998.
- [29] J. Madarász. Interpolation and amalgamation; pushing the limits. Part II. *Studia Logica*, 62:1–19, 1999.
- [30] M. Marx. *Algebraic Relativization and Arrow Logic*. PhD thesis, University of Amsterdam, 1995.
- [31] R. Montague. Universal grammar. *Theoria*, 36:373–398, 1970.
- [32] M. Otto. Bisimulation invariance and finite models. In W. Pohlers Z. Chatzidakis, P. Koepke, editor, *Logic Colloquium ’02*, volume 27 of *Lecture Notes in Logic*, pages 276–298. Association for Symbolic Logic, 2006.
- [33] V. Padmanabhan, G. Governatori, and K. Su. Knowledge assesment: A modal logic approach. In *Proceedings of the 3rd Int. Workshop on Knowledge and Reasoning for Answering Questions (KRAQ)*, 2007.
- [34] D. Pattinson. An introduction to the theory of coalgebras. Lecture notes accompanying the course at NASSLLI 2003. Available at <http://www.pst.ifi.lmu.de/~pattinso/Publications/nasslli.all.ps.gz>.
- [35] D. Pattinson. Coalgebraic modal logic: Soundness, completeness and decidability of local consequence. *Theoretical Computer Science*, 309(1–3):177–193, 2003.
- [36] M. Pauly. Bisimulation for general non-normal modal logic. Manuscript (unpublished), 1999.
- [37] M. Pauly. *Logic for Social Software*. PhD thesis, University of Amsterdam, 2001.
- [38] M. Pauly. A modal logic for coalitional power in games. *Journal of Logic and Computation*, 12(1):149–166, 2002.
- [39] J.J.M.M. Rutten. Universal coalgebra: a theory of systems. *Theoretical Computer Science*, 249:3–80, 2000.
- [40] L. Schröder. Expressivity of coalgebraic modal logic: The limits and beyond. *Theoretical Computer Science*, 390:230–247, 2008.
- [41] D. Scott. Advice on modal logic. In K. Lambert, editor, *Philosophical Problems in Logic*, pages 143–173. Reidel, Dordrecht, 1970.
- [42] K. Segerberg. *An Essay in Classical Modal Logic*. Number 13 in Filosofiska Studier. Uppsala Universitet, 1971.

- [43] M.Y. Vardi. On epistemic logic and logical omniscience. In J. Halpern, editor, *Proceedings of the 1986 Conference on Theoretical Aspects of Reasoning about Knowledge (TARK)*, pages 293–305. Morgan Kaufmann, 1986.
- [44] Y. Venema. Algebras and coalgebras. In P. Blackburn, J. van Benthem, and F. Wolter, editors, *Handbook of Modal Logic*, volume 3 of *Studies in Logic and Practical Reasoning*, pages 331–426. Elsevier, 2006.