MODAL DEFINABILITY QUESTIONS IN THE FINITE

Johan van Benthem,[a,b,c] *Balder ten Cate*[a] *and Xi Yang*[a,b]

ABSTRACT. Inspired by Rob Goldblatt's foundational contributions to modal correspondence theory, this paper asks how far classical modal definability results and links between modal formulas and first-order conditions persist over finite structures. We show that several modal preservation theorems transfer to the finite, prove a finite version of the Bisimulation Safety Theorem, and discuss finite-frame analogues of the Goldblatt–Thomason Theorem.

Dedication. Rob Goldblatt is one of the architects of mathematical modal logic as we know it today. It is a pleasure to contribute this article in his honor, and we hope that he will recognize in it some of his lifelong model-theoretic interests.

1. INTRODUCTION

The model theory of first-order logic exhibits a striking divergence between the infinite and the finite. Many classic theorems — central to the semantic and structural understanding of first-order logic — fail when attention is restricted to finite structures [26; 1]. For instance, compactness breaks down, the Łoś–Tarski preservation theorem no longer characterizes the existential and universal fragments, and Craig interpolation fails once we move to the finite. Against this backdrop of failure for classical metatheory, it is all the more remarkable that a few deep results *do* survive the transition to the finite. A notable case in point is the van Benthem Characterization Theorem [7], which identifies modal logic as the bisimulation-invariant fragment of first-order logic. It was shown by Rosen [23] that this remains valid over finite structures. A second notable example is Rossman's Homomorphism Preservation Theorem [24], establishing that the positive existential fragment of first-order logic is precisely the class of formulas preserved under (not necessarily surjective) homomorphisms, even in the finite. A long standing open problem remains whether, for relational signatures, the universal Horn fragment is characterized by closure under (finite) products and induced substructures in the finite [2; 18].

Within the more restricted setting of *modal logic*, the situation is more hospitable. The basic modal language, interpreted over the class of all Kripke models, has the

2020 *Mathematics Subject Classification.* Primary: 03B45, Secondary: 03C13, 68Q19.
Keywords. Modal logic, Finite model theory, Bisimulation safety, Frame definability.

Bimbó, K. and E. Mares (eds.), *Algebraic and Intensional Logic: Essays in Honour of Rob Goldblatt*, (Tributes, vol. 56), College Publications, London, UK, 2026, pp. 61–74.

finite model property. This gives rise to a natural heuristic: results that hold in the general setting may plausibly carry over to the finite, even when their first-order analogues fail in the finite. For example, it follows quite directly from the finite model property that the Craig interpolation theorem transfers to the finite, which also implies that the Beth definability theorem remains true in the finite.

In this paper we explore several complementary themes.

(1) We review some basic model-theoretic preservation results concerning syntactic fragments of modal logic (interpreted over the class of all Kripke models) and show how they transfer to the finite. In particular, we show how the known syntactic characterizations of the semantic properties of complete additivity and continuity transfer.
(2) In the first-order setting, we show a new positive result: the Bisimulation Safety Theorem, which characterizes the first-order operations on binary relations that are safe for bisimulation, carries over to the finite.
(3) Finally, we look at the celebrated Goldblatt–Thomason Theorem about modally definable frame classes and give a version that holds in the finite, while identifying some further open problems.

2. THE MODAL LOGIC SETTING

In this section, we discuss four known examples of preservation theorems for modal formulas and show that they transfer to the finite. In the first two cases, the arguments are very easy and can be considered folklore. The third example, *complete additivity*, requires a more involved proof (and it will play an important role in the next section, where we will use it to prove that the Bisimulation Safety Theorem transfers to the finite). A recurring theme will be that we make use of the well known fact that modal logic has the finite model property.

Theorem 1 (Finite Model Property)**.** *A modal formula is valid over arbitrary Kripke models if and only if it is valid over finite Kripke models.*

Note: we assume familiarity with modal logic. We refer to [13] for a general introduction. Throughout this section, we will consider *multi-modal* formulas, i.e., modal formulas that may contain not just one pair of modal operators $\Diamond, \Box$, but a family $\langle a \rangle, [a]$ for $a \in A$. These are then interpreted on Kripke models of the form $(W, (R_a)_{a \in A}, V)$. In other words, our setting corresponds to that of the basic multi-modal logic **K**. The fact that we allow multiple modalities does not affect the results or the proofs in any way, but it facilitates the use (in Section 3) of one of the results stated in the current section.

We will now consider four important semantic properties that modal formulas can have.

2.1. MONOTONICITY.

A modal formula is said to be *monotone* in a propositional variable p if its truth is preserved when enlarging the valuation of p. Formally, ϕ is monotone in p if, for all Kripke models $M = (W, (R_a)_{a \in A}, V)$ and $w \in W$, $M, w \vDash \phi$ implies that $M[p \mapsto X], w \vDash \phi$ for all sets $X \subseteq W$ with $V(p) \subseteq X$.

Theorem 2 ([19, Theorem 3.11], cf. also [20, Appendix B])**.** *For all modal formulas ϕ and propositional variables p, the following are equivalent:*[1]

(1) *ϕ is monotone in p;*
(2) *ϕ is equivalent to a modal formula ψ that is positive in p.*

Here, we say that a modal formula ψ is *positive in p* if it is in negation normal form and does not contain any negated occurrences of p.

It turns out that the fact that a formula ϕ is monotone in a propositional variable p is correlated to the validity of an associated modal formula ϕ', both in the finite and in the infinite.

Proposition 3. *For all modal formulas ϕ and propositional variables p, the following are equivalent:*

(1) *ϕ is monotone in p.*
(2) *The following validity holds: $\vDash \phi \to \phi[p/(p \vee q)]$*

with q a fresh propositional variable. This equivalence holds both over arbitrary structures and in the restricted class of finite structures.

Indeed, clearly, any counterexample to monotonicity provides a counterexample for the given validity, and vice versa.

In light of the finite model property, it now follows that Theorem 2 holds in the finite: if ϕ is monotone in p over finite structures, then, by Proposition 3, $\phi \to \phi[p/(p \vee q)]$ is valid in the finite. Hence, by the finite model property, it is valid over arbitrary structures. Applying Proposition 3 once more, we derive that ϕ is monotone in p over arbitrary structures. Therefore, by Theorem 2, ϕ is equivalent over arbitrary structures, and hence, also over finite structures, to a formula of the required syntactic shape.

2.2. PRESERVATION UNDER INDUCED SUBSTRUCTURES. We say that a modal formula ϕ is *preserved under induced substructures* if $M, w \vDash \phi$ implies $M|X, w \vDash \phi$, where X is any subset of the domain that includes w, and $M|X$ is the restriction of M to X. The following preservation theorem is known for preservation under induced substructures; cf. also its history in [28].

Theorem 4 ([22, Theorem 6.6.5])**.** *For all modal formulas ϕ, the following two assertions are equivalent:*

(1) *ϕ is preserved under induced substructures;*
(2) *ϕ is equivalent to a modal formula ψ in negation normal form not containing any $\Diamond$ operator.*

The fact that a modal formula ϕ is preserved under induced substructures can again be correlated to the validity of an associated modal formula ϕ'. We first need to introduce some notation. For a modal formula ϕ and a propositional variable p, let ϕ^p be the result of relativizing ϕ to p (i.e., replacing every $\langle a \rangle$ by $\langle a \rangle(p \wedge \cdots)$ and replacing every $[a]$ by $[a](p \to \cdots)$).

[1] Technically speaking, the cited result applies to formulas that are monotone in *all* propositional variables rather than in a specific propositional variable. In terms of the proof, this does not make an essential difference.

Proposition 5. *For all modal formulas ϕ, the following are equivalent:*

(1) *ϕ is preserved under induced substructures;*
(2) *The following validity holds: $\models (p \wedge \phi) \rightarrow \phi^p$*

where p is a fresh propositional variable. This equivalence holds both over arbitrary structures and in the finite.

Indeed, any (finite) counterexample to preservation under induced substructures provides a (finite) counterexample for the given validity, and vice versa.

It again follows that Theorem 4 holds also in the finite, using the same line of argument as in the case of monotonicity.

2.3. COMPLETE ADDITIVITY. A modal formula ϕ is *completely additive* in a propositional variable p if ϕ is monotone in p and whenever $M, w \models \phi$, then there is an element $v \in V(p)$ such that $M[p \mapsto \{v\}], w \models \phi$.

Theorem 6 ([11])**.** *A modal formula ϕ is completely additive in a propositional variable p if and only if ϕ is equivalent to a finite disjunction of formulas θ that can each be generated by the following recursive grammar:*

$$\theta \ := \ p \wedge \phi \mid \phi \wedge \langle a \rangle \theta$$

where ϕ is a p-free modal formula and $a \in A$.

The fact that a modal formula ϕ is completely additive in a propositional variable p can again be correlated to the validity of an associated modal formula ϕ', both in the finite and in the infinite. This time, the proof, however, is less straightforward.

Proposition 7. *For all modal formulas ϕ and propositional variables p, the following are equivalent, where q_1, q_2 are fresh propositional variables:*

(1) *ϕ is completely additive in p;*
(2) *The following two validities hold: $\models \phi[p/q_1 \vee q_2] \leftrightarrow (\phi[p/q_1] \vee \phi[p/q_2])$ and $\models \neg\phi[p/\bot]$.*

The equivalence holds both over arbitrary structures and in the finite.

Proof. From 1 to 2: Suppose ϕ is completely additive in p. In particular, this means that ϕ is monotone in p. It follows from the monotonicity that $\models (\phi[p/q_1] \vee \phi[p/q_2]) \rightarrow \phi[p/q_1 \vee q_2]$. Next, suppose $M, w \models \phi[p/q_1 \vee q_2]$. Since ϕ is completely additive in p, it follows that $M[p \mapsto \{v\}], w \models \phi$ for some $v \in V(q_1) \cup V(q_2)$. If $w \in V(q_1)$, it follows by monotonicity that $M, w \models \phi[p/q_1]$, while, if $v \in V(q_2)$, it follows by monotonicity that $M, v \models \phi[p/q_2]$. It also follows directly from the complete additivity property that $\models \neg\phi[p/\bot]$. The same argument applies in the finite.

From 2 to 1: assume that the stated validities hold. We will proceed in two steps. First, we will show that ϕ is completely additive over finite structures. We prove this by contraposition: let (M, w) be any finite counterexample to the complete additivity of ϕ in p. In particular, $M = (W, (R_a)_{a \in A}, V)$ is a finite Kripke structure, $M, w \models \phi(p)$ and one of the two conditions holds:

(1) There are sets $X \subseteq Y \subseteq W$ such that $(M[p \mapsto X], w) \models \phi$ and $(M[p \mapsto Y], w) \not\models \phi$, or

(2) There is a set $X \subseteq W$ such that $(M[p \mapsto X], w) \models \phi$ and there is no $x \in X$ such that $(M[p \mapsto \{x\}], w) \models \phi$.

In the first case, we have $(M[q_1 \mapsto X, q_2 \mapsto Y], w) \not\models (\phi[p/q_1] \vee \phi[p/q_2]) \rightarrow \phi[p/q_1 \vee q_2]$. In the second case, if $X = \emptyset$ we have $(M, w) \not\models \neg\phi[p/\bot]$, while if $X \neq \emptyset$, then (by finiteness of the domain) there must exist a minimal-size non-empty subset $X \subseteq W$ satisfying condition 2. Furthermore X cannot be a singleton set, and therefore it must be the case that $X = Y \cup Z$ for disjoint non-empty sets Y and Z. It then follows that $(M[q_1 \mapsto Y, q_2 \mapsto Z], w) \not\models \phi[p/q_1 \vee q_2] \rightarrow (\phi[p/q_1] \vee \phi[p/q_2])$.

Having established already that ϕ is completely additive in p in the finite, we will now argue that the same holds over arbitrary structures. In particular ϕ is monotone in p in the finite. It follows from Proposition 3 that ϕ is monotone in p over arbitrary structures. It therefore only remains to show that the second half of the definition of complete additivity is satisfied over arbitrary structures. Assume for the sake of a contradiction that this condition is *not* satisfied. Let M, w with $M = (W, (R_a)_{a \in A}, V)$ be a witness, i.e., a pointed Kripke structure such that $M, w \models \phi$ and such that $M[p \mapsto \{v\}], w \not\models \phi$ for all $v \in V(p)$. We will show how to turn M, w into a *finite* counterexample to complete additivity.

Let $M^{unr} = (W^{unr}, (R_a)^{unr}_{a \in A}, V^{unr})$ be the tree unraveling of M, defined as usual.

Claim 1: $(M^{unr}, \langle w \rangle)$ is also a witness to the failure of complete additivity.

Indeed, $(M^{unr}, \langle w \rangle) \models \phi$ follows from the fact that (M, w) and $(M^{unr}, \langle w \rangle)$ are bisimilar. Furthermore, for any $u \in V^{unr}(p)$, we have that $(M[p \rightarrow \{\pi(u)\}], w)$ is bisimilar to $(M^{unr}[p \rightarrow X_u], \langle w \rangle)$, where π is the natural projection and where $X_u = \{u' \colon \pi(u) = \pi(u')\}$. By assumption, $(M[p \rightarrow \{\pi(u)\}], w) \not\models \phi$, and hence by bisimulation invariance, $(M^{unr}[p \rightarrow X_u], \langle w \rangle) \not\models \phi$. It follows by monotonicity (over arbitrary Kripke structures) that $(M^{unr}[p \rightarrow \{u\}], \langle w \rangle) \not\models \phi$.

Let M' be the substructure of M^{unr} containing only those worlds that are reachable from $\langle w \rangle$ in at most n steps, where n is the modal depth of ϕ. Since the truth of ϕ in a structure depends only on this substructure, Claim 1 implies:

Claim 2: $(M', \langle w \rangle)$ is also a witness to the failure of complete additivity.

Observe that M' is a tree of bounded depth. With a small amount of additional "surgery" it is possible to turn M' into a finite Kripke structure that is still a witness to the failure of complete additivity (this can be done by pruning the tree so that each node has at most two children of any given subtree isomorphism type, cf. [11]). However, as it turns out we can make life even easier by appealing to a powerful result about the monadic second order logic (MSO) theory of unranked trees, namely the fact that *the MSO-theory of the class of all trees is equal to the MSO-theory of the class of all finitely branching trees* [29].[2] Indeed, consider the MSO-sentence

$$\Psi := \exists x(ST_x(\phi) \wedge \forall Q(\exists y(Py \wedge \forall z(Q(z) \leftrightarrow z = y)) \rightarrow \neg ST_x(\phi)[P/Q])) \wedge depth_n$$

where $depth_n$ is short for $\neg \bigvee_{a_1, \ldots, a_n \in A} \exists x_1 \ldots x_{n+1}(R_{a_1}(x_1, x_2) \wedge \cdots \wedge R_{a_n}(x_n, x_{n+1}))$. By Claim 2, Ψ is satisfied in a tree. It follows that Ψ is also satisfied in a finitely branching tree. Since every finitely branching tree of finite depth is a finite tree, this concludes our proof. $\dashv$

[2] This result follows from Lemma 43 in [29], though it was not explicitly stated there.

It follows again that Theorem 6 transfers to the finite.

2.4. CONTINUITY. A modal formula ϕ is said to be *continuous* in a propositional variable p if, for all pointed Kripke structures (M,w), it holds that $M,w \vDash \phi$ if and only if there is a finite set $X \subseteq V(p)$ such that $M[p \mapsto X], w \vDash \phi$. Furthermore, ϕ said to be *strongly continuous* in p if it is n-continuous in p for some n, by which we mean that, for all pointed Kripke structures (M,w), it holds that $M,w \vDash \phi$ if and only if there is a set $X \subseteq V(p)$ of size at most n such that $M[p \mapsto X], w \vDash \phi$. For example, it is easy to see that $[a]p$ is *not* continuous in p but $\langle a \rangle p$ *is* (as is every other completely additive modal formula). This property of continuity has been identified and studied especially in the context of (modal and first-order) fixed point logics, since it guarantees stabilization of fixed-points at stage ω [9; 15; 16].

The continuous fragment of modal logic has been characterized syntactically as follows:

Theorem 8 ([15, Corollary 1])**.** *For all modal formulas ϕ and propositional variables p, the following are equivalent:*

(1) *ϕ is continuous in p;*
(2) *ϕ is equivalent to a formula in negation normal form such that no occurrence of p is in the scope of a negation or of a $\Box$ operator.*

Clearly, continuity is trivial as a property of formulas interpreted over finite structures. Therefore, the above characterization trivially does *not* transfer to the finite. However, a variant of it does.

Proposition 9. *For all modal formulas ϕ and propositional variables p, the following are equivalent:*

(1) *ϕ is continuous in p;*
(2) *ϕ is strongly continuous in p (i.e., ϕ is n-continuous in p for some $n \in \mathbb{N}$).*

Proof. The direction from 2 to 1 is trivial, while the direction from 1 to 2 follows from the compactness theorem for first-order logic, and holds in fact for every first-order formula. Indeed, let ϕ be any first-order formula containing a unary predicate P. Let $\Psi = \{\neg\psi_0, \neg\psi_1, \neg\psi_2, \ldots\}$ with $\psi_n = \exists y_1 \ldots y_n (P(y_1) \wedge \cdots \wedge P(y_n) \wedge \phi')$, with ϕ' obtained from ϕ by replacing every subformula of the form $P(z)$ by $z = y_1 \vee \cdots \vee z = y_n$. By continuity, Ψ entails $\neg\phi$. Therefore, by compactness, some finite subset $\{\neg\psi_{i_1}, \ldots, \neg\psi_{i_k}\} \subseteq \Psi$ entails $\neg\phi$. This shows that ϕ is n-continuous for $n = \max\{i_1, \ldots, i_k\} + 1$. ⊣

Next, we can apply our usual technique and correlate the fact that ϕ is n-continuous in p to the validity of an associated formula ϕ'.

Proposition 10. *For all modal formulas ϕ, propositional variables p, and $n \in \mathbb{N}$, the following are equivalent:*

(1) *ϕ is n-continuous in p;*
(2) *The following validity holds: $\vDash \phi[p/\bigvee_{i\in\{1\ldots n\}}(p_i)] \leftrightarrow \bigvee_{k\in\{1\ldots n\}} \phi[p/(\bigvee_{i\in\{1\ldots n\}\setminus\{k\}}(p_i)]$,*

where $p_1, \ldots, p_n$ are fresh propositional variables. This equivalence holds both over arbitrary structures and in the finite.

Proof. From 1 to 2: Assume that ϕ is n-continuous in p. In particular, then, ϕ is monotone in p. It follows that the right-to-left direction of the bi-implication is valid. For the left-to-right direction, suppose that $M, w \vDash \phi[p/\bigvee_{i\in\{1,\ldots,n\}}(p_i)]$. Then, by n-continuity, there is a set $X \subseteq \bigcup_{i\in\{1,\ldots,n\}} V(p_i)$ with $|X| < n$ and $M[p \mapsto X], w \vDash \phi$. For cardinality reasons, it must be the case that $X \subseteq \bigcup_{i\in\{1,\ldots,n\}\setminus\{k\}} V(p_i)$. Therefore, by monotonicity, $M, w \vDash \phi[p/\bigvee_{i\in\{1,\ldots,n\}\setminus\{k\}}(p_i)]$.

From 2 to 1, the exact same proof strategy applies as in the case of complete additivity (cf. the proof of Proposition 7). The main change is that the MSO-formula Ψ expressing failure of complete additivity is replaced by the following MSO-formula expressing failure of n-continuity:

$$\Psi' := \exists x(ST_x(\phi) \wedge \forall Q\Big(\forall y(Qy \to Py) \wedge \exists y_1 \ldots y_n \forall z(Qz \to \bigvee_{i=1\ldots n} z = y_i) \to \neg ST_x(\phi)[P/Q]\Big) \wedge depth_m \quad \dashv$$

It follows that, although Theorem 8 does not transfer to the finite as it stands, it does transfer once continuity is replaced by strong continuity.

2.5. WHERE WE STAND. We have analyzed four examples of ubiquitous and useful semantic properties of modal formulas. In the simplest cases, a syntactic characterization transfers to the finite thanks to two features. One is that the preservation property can itself be expressed in modal terms, the other that modal logic has the finite model property. In some more complex preservation cases, we found that some additional considerations were needed. All this does not add up to a concrete necessary and sufficient criterion for transfer to the finite of properties of modal logic — finding such a criterion (if one exists at all) would be interesting but is beyond our reach here.

Our investigation also has a broader context. There are many decidable extensions of the modal fragment of first-order logic, such as the Guarded Fragment [3] and the Guarded Negation Fragment [4], that retain the finite model property that we have used so extensively. We leave it as an open direction to explore the transfer-to-the-finite behavior of semantic properties of formulas for such fragments.

This concludes our modal-style exploration. In the next section, we give a similar example-based analysis of transfer to the finite for preservation properties in first-order logic.

3. THE FIRST-ORDER SETTING: BISIMULATION SAFETY IN THE FINITE

The results in Section 2 were predominantly positive. In particular, we looked at a number of semantic properties of modal formulas, such as *monotonicity* and *preservation under induced substructures*, and we saw that the corresponding preservation theorems do transfer to the finite. The picture is not as rosy for first-order logic. Indeed, it was shown in [26] that the Łoś–Tarski theorem (characterizing the first-order formulas preserved under induced substructures) fails in the finite, and it was shown in [1] that, over finite structures, monotonicity can no longer be syntactically characterized in terms of positivity either. Nevertheless, as discussed in the introduction, there are examples of preservation theorems that do hold in the finite, such as Rossman's

preservation theorem for (non-surjective) homomorphisms [24]. In this section, we add one more such positive result to the list, namely the Bisimulation Safety Theorem.

The *Bisimulation Characterization Theorem* and the *Bisimulation Safety Theorem*, stated below, characterize the first-order formulas that are invariant, respectively safe, for bisimulation. We will restrict here to first-order signatures consisting of unary relations (corresponding to propositional variables) and binary relations (corresponding to modal operators).

Recall that a first-order formula $\phi(x)$ (over the appropriate signature) is *bisimulation-invariant* if, whenever (M,w) and (M',w') are bisimilar, then $M,w \vDash \phi(x)$ iff $M',w' \vDash \phi(x)$.

Theorem 11 (Bisimulation Characterization Theorem [6])**.** *A first-order formula $\phi(x)$ is bisimulation-invariant if and only if it is logically equivalent to the standard translation of some modal formula.*

Rosen [23] showed that the Bisimulation Characterization Theorem holds also in the finite. That is, a first-order formula $\phi(x)$ is bisimulation-invariant over finite structures if and only if it is equivalent, over finite structures, to the standard translation of a modal formula.

The Bisimulation Safety Theorem is another classic result, which syntactically characterizes the first-order operations on binary relations that are safe for bisimulation. A first-order formula $\phi(x,y)$ (over the appropriate signature) is *safe for bisimulation* if every bisimulation between Kripke structures (M,w) and (M',w') also satisfies the back and forth clauses w.r.t. the binary relation defined by ϕ. To state the theorem, we first need to define *modal programs*. A modal program is an expression built by the following grammar:

$$\pi := p? \mid a \mid \pi;\pi' \mid \pi \cup \pi' \mid \sim\pi$$

where p is a propositional variable and a an atomic program. Each modal program, interpreted in a Kripke structure $M = (W,(R_a)_{a\in A},V)$, defines a binary relation:

$$\begin{array}{l}
[\![p?]\!]_M = \{\langle w,w\rangle : w \in V(p)\} \\
[\![a]\!]_M = R_a \\
[\![\pi;\pi']\!]_M = [\![\pi]\!]_M \cdot [\![\pi']\!]_M \\
[\![\pi \cup \pi']\!]_M = [\![\pi]\!]_M \cup [\![\pi']\!]_M \\
[\![\sim\pi]\!]_M = \{\langle w,w\rangle : w \in W \text{ and there is no } v \in W \text{ s.t. } \langle w,v\rangle \in [\![\pi]\!]_M\}
\end{array}$$

The standard translation from modal formulas to first-order formulas has a natural analogue for modal programs, yielding first-order formulas in two free variables:

$$\begin{array}{lcll}
ST_{x,y}(a) & = & R_a(x,y) & \text{for } a \in A \\
ST_{x,y}(p?) & = & (x=y) \wedge P(x) & \text{for } p \text{ a propositional variable} \\
ST_{x,y}(\pi;\pi') & = & \exists z(ST_{x,z}(\pi) \wedge ST_{z,y}(\pi')) & \\
ST_{x,y}(\pi \cup \pi') & = & ST_{x,y}(\pi) \vee ST_{x,y}(\pi') & \\
ST_{x,y}(\sim\pi) & = & (x=y) \wedge \neg\exists z(ST_{x,z}(\pi)) &
\end{array}$$

Theorem 12 (Bisimulation Safety Theorem [11])**.** *A first-order formula $\phi(x,y)$ is safe for bisimulation if and only if it is equivalent to the standard translation of a modal program.*

The bisimulation safety theorem can also be cast as a semantic characterization for a syntactic fragment of Tarski's relation algebra, cf. [14].

We will show that Theorem 12 transfers to the finite as well. The core ingredient of the proof is Proposition 7 above, which (in combination with the finite model property of modal logic) tells us that a modal formula is completely additive in the finite if and only if it is completely additive over arbitrary Kripke structures. By a *p-free program* we will mean a program expression that does not contain any occurrence of p. The following lemma exhibits a relationship between modal programs and completely additive formulas.

Lemma 13 ([11])**.** *For each modal formula θ of the form as described in Theorem 6, there is a p-free modal program π_θ such that, for all Kripke structures M, it holds that $M, w \vDash \theta$ iff there is a pair $\langle w, v\rangle \in [\![\pi_\theta]\!]_M$ with $M, v \vDash p$.*

Van Benthem's proof of the Bisimulation Safety Theorem in [11] uses the Bisimulation Characterization Theorem as a black box. It goes as follows: suppose $\phi(x,y)$ defines a bisimulation safe operation on binary relations. Let P be a fresh unary predicate. Then it follows that $\psi(x) := \exists y(\phi(x,y) \wedge P(y))$ is bisimulation-invariant. Therefore, by the Bisimulation Characterization Theorem, $\psi(x)$ is equivalent to a modal formula χ (which may use a propositional variable p corresponding to the unary predicate P). It follows that χ is completely additive in p (because, clearly, ψ is). Therefore, by Proposition 7 together with Lemma 13, using some straightforward syntactic manipulation, ψ can be written in the form $\langle\pi\rangle p$ for some p-free modal program π. It then follows that π defines the same relation as $\phi(x,y)$.

The bisimulation safety theorem in the finite follows using the same argument as before, but using Rosen's theorem [23] instead of the Bisimulation Characterization Theorem, and using Proposition 7. Suppose $\phi(x,y)$ defines a bisimulation safe operation on binary relations, in the finite. Let P be a fresh unary predicate. Then (by the same arguments as in [11]), $\psi(x) := \exists y(\phi(x,y) \wedge P(y))$ is bisimulation-invariant in the finite. Therefore, by Rosen's theorem, $\psi(x)$ is equivalent, in the finite, to a modal formula χ (which may use a propositional variable p corresponding to the unary predicate P). It follows that χ is completely additive in p in the finite (because, clearly, ψ is). Therefore, by Proposition 7 in combination with the finite model property, χ is completely additive in p over all Kripke structures. Hence, by Theorem 6 together with Lemma 13, ψ can be written in the form $\langle\pi\rangle p$ for some p-free modal program π. It then follows that π defines the same relation as $\phi(x,y)$, in the finite. This concludes the proof.

This concludes the main part of this paper: an exploration of transfer to the finite of the classical preservation theorems governing the fine-structure of definability in modal logic. Putting this theme in a broader context, our final section outlines how our style of analysis also applies to a celebrated classical definability result in modal model theory, due to Rob Goldblatt and Steve Thomason.

4. One more test case: Goldblatt–Thomason in the finite

The following result determines which semantic behavior is characteristic for modal formulas on relational *frames* rather than on the relational *models* studied in earlier

sections. Recall that a class K of frames is *modally definable* if there is a set Φ of modal formulas such that K consists precisely of those frames on which each modal formula in Φ is valid. Note that, as before, we restrict attention to first-order sentences over a signature consisting of binary relations only. Such sentences describe properties of (multi-modal) Kripke frames.

Theorem 14 ([17])**.** *Let* **K** *be any class of frames that is definable by a set of first-order sentences. Then the following are equivalent:*

(1) **K** *is modally definable.*
(2) **K** *is closed under taking generated subframes, disjoint unions, and bounded morphic images, and its complement is closed under taking ultrafilter extensions.*

There is also an 'absolute' version without the assumption of first-order definability, but then in terms of less appealing closure conditions. One can also focus attention on definability by single formulas rather than sets. In what follows we will not consider such variations in detail.

The original proof of the Goldblatt–Thomason theorem used algebraic methods, creating a fruitful bridge between model theory and universal algebra in modal logic. However, there is also a purely model-theoretic proof without algebraic representation methods, cf. [8]. Using this proof method, the cited paper also notes that on finite transitive frames, closure under generated subframes, disjoint unions and bounded morphic images is necessary and sufficient for modal definability. In what follows, we analyze the situation in a bit more detail, including transfer to the finite.

4.1. THE CASE OF MODELS. For a start, to see the transition from our earlier model-based perspective, let us state an analogue of the Goldblatt–Thomason theorem on models rather than frames. Here the relevant result is essentially van Benthem's theorem, whose standard proof can be adapted with minor changes to yield the following:

Theorem 15. *Let* **K** *be any class of pointed models definable by a set of first-order sentences. Then the following are equivalent:*

(1) **K** *is modally definable.*
(2) **K** *is closed under bisimulation.*

An absolute version for modal definability of arbitrary classes of models is obtained by adding familiar closure conditions from modal model theory.

Theorem 16. *Let* **K** *be any class of pointed models. Then the following are equivalent:*

(1) **K** *is modally definable.*
(2) **K** *is closed under bisimulation and under ultraproducts over countable index sets, while the complement of* **K** *is closed under ultrapowers over countable index sets.*

Again the proof follows the familiar one for van Benthem's theorem, in particular, its key lemma that modally equivalent pointed models have bisimilar ultrapowers over a countable index set.

Remark 17. This result can be improved, say, replacing ultraproducts and ultrapowers by ultrafilter extensions having just some minimal saturation properties, cf. the treatment in [27].

Next consider *finite models*. In this case a more direct approach is available by using a modal analogue of so-called 'Scott sentences' [25], which define any given model up to potential isomorphism by enumerating all existing types recursively up to some large enough ordinal.

In this setting, one relevant definability result captures models up to bisimulation, where we now employ an *extended modal language* with the basic relational modality $\Box\phi$ as well as the 'universal modality' $U\phi$, saying that ϕ is true at all points of the model.

Lemma 18. *Let (M,s) be any finite pointed model. Then there exists a modal formula $\delta(M,s)$ in the $\{\Box,U\}$ language such that the following are equivalent for all pointed models (N,t):*

(1) $(N,t) \models \delta(M,s)$.

(2) (N,t) *is connected to* (M,s) *by a total bisimulation.*

Here a total bisimulation has the complete point sets of the models connected by it for its domain and range. The definitions $\delta(M,s)$ arise by iteratively describing all 'modal types' that can occur at points in M in a recursive manner, noting that this process will terminate after finitely many steps in any fixed finite model. For a precise definition and a proof of the preceding assertion, see [10; 5].

Remark 19. The stated result is just one version. The cited papers also consider local variants with respect to standard not necessarily total bisimulations, as well as modal definitions employing the transitive closure modality, or fixed-point operators in the modal μ-calculus.

With the preceding result in hand we can prove:

Theorem 20. *Let* **K** *be any class of finite pointed models. Then the following are equivalent:*

(1) **K** *is modally definable in the modal $\{\Box,U\}$ language.*

(2) **K** *is closed under total bisimulations.*

Proof. (i) Total bisimulations preserve truth of all modal formulas in this extended language. (ii) As is easy to check, a modal definition for the class **K** is the set $\{\neg\delta(M,s)\colon (M,s)$ is not in **K** $\}$. ⊣

This is a brute force result by enumeration. More informative modal definitions arise when we analyze the situation a bit more combinatorially. For instance, the Rosen-style van Benthem theorem in the finite for the $\{\Box,U\}$ language proved in [21] implies the following. If the class **K** of models considered is defined by a single first-order sentence, then its modal definition can also make do with just one formula.

Another issue that we leave open here is the analysis of definability for classes of finite models in the pure basic modal $\Box$ language. Then complete descriptions up to bisimulation need not exist, and we can only define iterated types up to fixed finite levels.

4.2. THE CASE OF FRAMES. After these preliminaries on models, let us now move to *finite frames*, the real habitat of the Goldblatt–Thomason theorem.

We again work with an *extended modal language* containing both $\Diamond\phi$ and the *universal modality* $U\phi$. First, consider any finite frame F, and take propositional variables $p_{\{x\}}$ for each point x in F. Now expand F to a model M by interpreting these propositional variables as true only at their points. Then the following formula $\$(F)$ will be true in M:

Take one universal modality over a conjunction of

(a) every point satisfies exactly one of the $p_{\{x\}}$,

(b) for all s,t, if sRt in F, then $p_{\{s\}} \to \Diamond p_{\{t\}}$,

(c) for all s,t, if not sRt in F, then $p_{\{s\}} \to \Box\neg p_{\{t\}}$

Now the following statement holds for all models (N,t):

Lemma 21. *$(N,t) \models \$(F)$ if and only if there is a bounded morphism (i.e., a functional bisimulation) from the model N onto the model M.*

Proof. The proof uses the obvious embedding sending all points satisfying the propositional variable $p_{\{x\}}$ in N to the single point x in M, where the clauses (b), (c) in the formula $\$(F)$ enforce exactly the truth of the two back-and-forth clauses governing a bisimulation.[3] ⊣

Using this, we can formulate a version of the Goldblatt–Thomason Theorem for finite frames.

Theorem 22. *A class* **K** *of finite frames is definable by a set of modal formulas in the language $\{\Box, U\}$ if and only if* **K** *is closed under bounded morphic images.*

Proof. We show that the class **K** is defined by its own modal theory in our language. Consider any frame F which satisfies this theory, expand it to a model M as described above, and construct the formula $\$(F)$. Obviously, the formula $\neg\$(F)$ is not valid in the set **K** (since F satisfies the modal theory of **K**) , so there exists at least one frame G in **K** which refutes it for some valuation. But then, using the preceding observation, it follows that F is a bounded morphic image of G, and therefore F belongs to the class **K**. ⊣

Again, this is a brute-force method, producing in principle, a set of modal formulas defining the given class. Thus, obvious questions remain about the scope of our analysis.

One is when there is *one single modal formula* defining the given set **K**. We suspect that the assumption that **K** be definable by one first-order sentence suffices for this, but we have no proof.[4]

Another immediate question is this. Can the preceding result be extended to work for just the *basic modal language* without the universal modality? In this case, at

[3] This method is similar to an analysis of finite models up to bisimulation in the earlier-cited [10].

[4] Perhaps Kruskal's Theorem in Graph Theory may be relevant here as a technique for proving definability through the existence of finite maximal anti-chains in bounded-morphism ordering of models.

least, the Goldblatt–Thomason conditions of closure under generated subframes and disjoint unions need to be reinstated.

Going further on this downward path of fine-structure, keeping in mind our theme of preservation results for *fragments of the modal language*, what happens to the preceding analysis when we restrict the syntax to, say, positive fragments of various sorts?

Finally, returning to the main theme of this paper, one can view the preceding theorem as saying that at least *some versions of the Goldblatt–Thomason theorem transfer to the finite*. Here, its further condition of closure under ultrafilter extensions can be retained, since this is always satisfied on finite frames where the ultrafilter extension of a frame is isomorphic to that frame itself. A full study of these transfer issues is beyond the scope of this paper.[5]

This section has just offered a sample of a much larger topic, viz. transfer to the finite of major results about modal definability of frame properties. For a deeper exploration of what happens to classical results in Modal Correspondence Theory, including the hierarchy of available correspondence languages from first-order logic to fixed-point logics, see [12].

References

[1] Ajtai, M. and Gurevich, Y. (1987). Monotone versus positive, *Journal of the ACM* **34**(4): 1004–1015.

[2] Alechina, N. and Gurevich, Y. (1997). Syntax vs. semantics on finite structures, *Structures in Logic and Computer Science: A Selection of Essays in Honor of A. Ehrenfeucht*, Springer, pp. 14–33.

[3] Andréka, H., Németi, I. and van Benthem, J. (1998). Modal languages and bounded fragments of predicate logic, *Journal of Philosophical Logic* **27**(3): 217–274.

[4] Bárány, V., Cate, B. T. and Segoufin, L. (2015). Guarded negation, *Journal of the ACM* **62**(3).

[5] Barwise, J. and Moss, L. (1996). *Vicious Circles: On the Mathematics of Non-Wellfounded Phenomena*, Center for the Study of Language and Information Publication Lecture Notes, Cambridge University Press.

[6] van Benthem, J. (1976). *Modal Correspondence Theory*, PhD thesis, University of Amsterdam.

[7] van Benthem, J. (1983). *Modal Logic and Classical Logic*, number 3 in *Indices. Monographs in Philosophical Logic and Formal Linguistics*, Bibliopolis, Napoli.

[8] van Benthem, J. (1993). Modal frame classes revisited, *Fundamenta Informaticae* **18**(2–4): 307–317.

[9] van Benthem, J. (1996). *Exploring Logical Dynamics*, Center for the Study of Language and Information, Stanford, CA.

[10] van Benthem, J. (1998a). Dynamic odds and ends, *Technical Report ML1998–08*, Institute for Logic, Language and Computation, University of Amsterdam.

[11] van Benthem, J. (1998b). Program constructions that are safe for bisimulation, *Studia Logica* **60**(2): 311–330.

[5] For instance, we would then also like to know if the Goldblatt–Thomason theorem allows for a transferable version with a modal language extended with program expressions, just as we found with the Bisimulation Safety Theorem.

[12] van Benthem, J., ten Cate, B. and Yang, X. (2026). Modal definability, preservation, and correspondence in the finite. Manuscript in preparation.
[13] Blackburn, P., de Rijke, M. and Venema, Y. (2001). *Modal Logic*, number 53 in *Cambridge Tracts in Theoretical Computer Science*, Cambridge University Press, Cambridge, UK.
[14] Bogaerts, B., ten Cate, B., McLean, B. and den Bussche, J. V. (2024). Preservation theorems for Tarski's relation algebra, *Logical Methods in Computer Science* **20**(3).
[15] Fontaine, G. (2008). Continuous fragment of the mu-calculus, *in* M. Kaminski and S. Martini (eds.), *Computer Science Logic*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 139–153.
[16] Fontaine, G. and Venema, Y. (2018). Some model theory for the modal μ-calculus: syntactic characterisations of semantic properties, *Logical Methods in Computer Science* **14**(1).
[17] Goldblatt, R. I. and Thomason, S. K. (1975). Axiomatic classes in propositional modal logic, *Algebra and Logic: Papers from the 1974 Summer Research Institute of the Australian Mathematical Society, Monash University, Australia*, Springer, pp. 163–173.
[18] Ham, L. and Jackson, M. (2023). Finite model theory for pseudovarieties and universal algebra: preservation, definability and complexity.
[19] Kurtonina, N. and de Rijke, M. (1997). Simulating without negation, *Journal of Logic and Computation* **7**(4): 501–522.
[20] Kurtonina, N. and de Rijke, M. (1999). Expressiveness of concept expressions in first-order description logics, *Artificial Intelligence* **107**(2): 303–333.
[21] Otto, M. (2004). Modal and guarded characterisation theorems over finite transition systems, *Annals of Pure and Applied Logic* **130**: 173–205.
[22] de Rijke, M. (1993). *Extending Modal Logic*, PhD thesis, ILLC, University of Amsterdam.
[23] Rosen, E. (1995). *Finite Model Theory and Finite Variable Logics*, PhD thesis, University of Pennsylvania.
[24] Rossman, B. (2008). Homomorphism preservation theorems, *Journal of the ACM* **55**(3): 1–53.
[25] Scott, D. (1965). Logic with denumerably long formulas and finite strings of quantifiers, *in* J. Addison, L. Henkin and A. Tarski (eds.), *The Theory of Models*, North-Holland, Amsterdam, pp. 329–341.
[26] Tait, W. W. (1959). A counterexample to a conjecture of Scott and Suppes, *Journal of Symbolic Logic* **24**(1): 15–16.
[27] Venema, Y. (1999). Modal definability, purely modal, *in* J. Gerbrandy, M. Marx, M. de Rijke and Y. Venema (eds.), *JFAK. Essays Dedicated to Johan van Benthem on the Occasion of his 50th Birthday*, Vossiuspers (Amsterdam University Press), Amsterdam. URL: www.illc.uva.nl/j50.
[28] Visser, A., van Benthem, J., de Jongh, D. and Renardel de Lavalette, G. R. (1995). NNIL, a study in intuitionistic propositional logic, *Modal Logic and Process Algebra*, CSLI Publications, Stanford, CA, pp. 289–326.
[29] Walukiewicz, I. (2002). Monadic second-order logic on tree-like structures, *Theoretical Computer Science* **275**(1): 311–346.

[a] ILLC, University of Amsterdam, The Netherlands
[b] The Tsinghua-UvA JRC for Logic, Department of Philosophy, Tsinghua University, China
[c] Department of Philosophy, Stanford University, USA
Email: j.vanbenthem@uva.nl
Email: b.d.tencate@uva.nl
Email: x.yang4520@gmail.com