Modal Logic and Invariance

Johan van Benthem Denis Bonnay

Abstract

Consider any logical system, what is its natural repertoire of logical operations? This question has been raised in particular for first-order logic and its extensions with generalized quantifiers, and various characterizations in terms of semantic invariance have been proposed. In this paper, our main concern is with modal and dynamic logics. Drawing on previous work on invariance for first-order operations, we find an abstract connection between the kind of logical operations a system uses and the kind of invariance conditions the system respects. This analysis yields (a) a characterization of invariance and safety under bisimulation as natural conditions for logical operations in modal and dynamic logics, and (b) some new transfer results between first-order logic and modal logic.

In a model-theoretic perspective, logics differ by the power they make available to describe structures. The limits to this expressive power are given by 'similarity' relations over structures: similar structures cannot be distinguished in the relevant language. The notion of bisimulation for modal logic (ML) is a case in point: bisimilar Kripke models have the same modal theory, though getting a converse involves some complications. But also other well-known types of result characterize logics in terms of similarity relations: Lindström's Theorem says that first-order logic (FOL) is the logic one gets from invariance for potential isomorphisms when adding compactness, van Benthem's characterization theorem says that ML is the logic one gets from bisimulations inside the complete logic FOL, and so on.

But one might wish to go the other way around. Consider a logic with predicate atoms that has at least existential quantification and boolean combination, what is its natural notion of similarity between structures? 'Being potentially isomorphic' is one answer, but is there more to it than some mathematical facts linking logics like FOL or the infinitary logic $L_{\infty,\omega}$ with potential isomorphisms? In Section 1, we summarize an earlier characterization of potential isomorphism as the 'coarsest' similarity relation whose invariants are closed under existential quantification. Section 2 provides a new abstract version of this result, and a characterization of bisimulation follows naturally: bisimulations are the natural match for a logic based on modal existential quantification. Section 3 analyzes the connection between these results for potential isomorphisms and bisimulations by transforming first-order structures into Kripke models and back. In Section 4, we extend the analysis to dynamic logic, and present a new characterization result for operations on accessibility relations. We prove that safety under bisimulation for operations added on top of a logic based on invariance under bisimulation provides a sufficient and necessary condition for preservation of invariance under bisimulation.

In all, these results are a sort of 'abstract model theory' characterizing structural invariance relations in more general terms, instead of taking them for granted. We see this as shedding some new light on the logical constants of first-order and modal logic, and the relations between these two perspectives.

1 Invariance and first-order operations

1.1 Invariance for logical notions

Following Frege's well-known insight, quantifiers can be viewed as second order predicates. For example, let us have a look at the satisfaction clause for existential quantification (for a model $\mathcal{M}$, a formula ϕ and an assignment σ):

$$\mathcal{M} \vDash \exists x\ \phi(x)\ \sigma$$
iff

$$\text{there is an } a \in |\mathcal{M}| \text{ s.t. } \mathcal{M} \vDash \phi(x)\ \sigma[x := a]$$
$$\text{iff}$$
$$||\phi(x)||_{\mathcal{M},\sigma} = \{a \in M / \mathcal{M} \vDash \phi(x)\ \sigma[x := a]\} \text{ is not empty}$$

where $\sigma[x := a]$ is the assignment one gets from σ by resetting the value of x to a. Reading off from the satisfaction clause, there is a natural interpretation $Q_\exists$ to give to $\exists$, namely the class of structures which represent formula interpretations $||\phi||_{\mathcal{M},\sigma}$ which are semantically fine as far as existential quantification goes. So we set $Q_\exists = \{\langle M, P\rangle \ / \ P \subseteq M \text{ and } P \neq \emptyset\}$. This licenses the following phrasing of the satisfaction clause for existential quantification:

$$\mathcal{M} \vDash \exists x\ \phi(x)\ \sigma$$
$$\text{iff}$$
$$\langle M, ||\phi(x)||_{\mathcal{M},\sigma}\rangle \in Q_\exists$$

$\exists$ is a unary monadic quantifier, so that $Q_\exists$ is a class of sets equipped with a predicate extension. The idea can be generalized to quantifiers of arbitrary syntactic type. Let us consider a class[1] Q of relational structures of the form $\langle M, R\rangle$. Q can be taken to be the interpretation of a (generalized) relational quantifier $\overline{Q}$ endowed with the following satisfaction clause:

$$\mathcal{M} \vDash \overline{Q}xy\ \phi(x, y)\ \sigma$$
$$\text{iff}$$
$$\langle M, ||\phi(x, y)||_{\mathcal{M},\sigma}\rangle \in Q$$

Propositional connectives fit in the picture if we let booleans be part of the structures. Now, the following question arises: what are the natural classes of structures Q to be used as the interpretation of logical constants (first-order quantifiers or propositional connectives)? Under different names, *invariance under isomorphism* has been widely accepted as a necessary condition. Since Q is now any class of structures of the same similarity type that is meant to interpret a logical constant, it should satisfy:

If $\mathcal{M}$ and $\mathcal{M}'$ are isomorphic, then $\mathcal{M} \in Q$ iff $\mathcal{M}' \in Q$.

This means that two structures which are isomorphic are 'logically similar' and that no logical notion should be able to distinguish them. However, other notions of similarity between structures could be considered as candidates. Some recent proposals argue that various kinds of homomorphisms would be a better fit.[2] In what follows, we consider invariance under an arbitrary 'similarity relation' S, being just an equivalence relation over structures respecting their types.

[1]In general, the structures interpreting a quantifier will not form a set, so we are implicitly working in a background theory of classes in order to describe our structures.

[2]See [6], [5] for invariance under homomorphism and [9] for a general assessment of the invariance approach to logical constants.

Definition 1. *Let S be a similarity relation and Q a class of structures. We say that Q is S-invariant iff, whenever $\mathcal{M} \; S \; \mathcal{M}'$, then $\mathcal{M} \in Q$ iff $\mathcal{M}' \in Q$.*

The restriction to equivalence relations S is harmless, since by our definition, a relation and the smallest equivalence relation containing it have the same invariants. We define the *ordering* $\leq$ on similarity relations as reverse inclusion[3], that is $S \leq S'$ iff $S' \subseteq S$.

1.2 Existential quantification and potential isomorphisms

Now assume that we want a 'first-order like' logic, endowed with existential quantification. Which invariance relation S should we pick? Given the metalogical results mentioned in the introduction, potential isomorphisms are a natural choice. We shall turn this intuition into a precise characterization. But first, we recall the following fundamental notion:[4]

Definition 2. *A potential isomorphism I between two structures $\mathcal{M}$ and $\mathcal{M}'$ (notation: $\mathcal{M} \overset{I}{\approx} \mathcal{M}'$) is a non empty set of (possibly partial) functions from $|\mathcal{M}|$ to $|\mathcal{M}'|$ such that:*
– every function $f \in I$ is a partial isomorphism, that is, an isomorphism between $\mathcal{M}$ restricted to the domain of f and $\mathcal{M}'$ restricted to its range,
– for all functions $f \in I$ and objects $a \in |\mathcal{M}|$ (resp. $a' \in |\mathcal{M}'|$), there is a function $g \in I$ with $f \subseteq g$ and $a \in dom(g)$ (resp. $a' \in rng(g)$).

We use the notation Iso_p for the similarity relation of 'being potentially isomorphic'. Let us furthermore introduce two relevant properties of similarity relations:

Definition 3. *A similarity relation S preserves atoms iff whenever $\langle M, R_1, ..., R_n, a_1, ..., a_m\rangle \; S \; \langle M', R'_1, ..., R'_n, a'_1, ..., a'_m\rangle$, then*
– $a_{j_1} = a_{j_2}$ iff $a'_{j_1} = a'_{j_2}$, for all objects a_{j_1}, a_{j_2} among $a_1, ..., a_m$ and a'_{j_1}, a'_{j_2} among $a'_1, ..., a'_m$,
– $\langle a_{j_1}, ..., a_{j_k}\rangle \in R_i$ iff $\langle a'_{j_1}, ..., a'_{j_k}\rangle \in R'_i$, for all k-tuples $\langle a_{j_1}, ..., a_{j_k}\rangle$ of objects among $a_1, ..., a_m$ and $\langle a'_{j_1}, ..., a'_{j_k}\rangle$ among $\{a'_1, ..., a'_m\}$, where k is the arity of R_i and R'_i.

Definition 4. *A similarity relation commutes with object expansions iff if $\mathcal{M} \; S \; \mathcal{M}'$, then for all $a \in |\mathcal{M}|$, there is an $a' \in |\mathcal{M}'|$ such that $\mathcal{M}, a \; S \; \mathcal{M}', a'$.*

where $\mathcal{M}, a$ is $\mathcal{M}$ expanded with the extra distinguished object a.

Note that no vice versa condition is needed in this definition since S is assumed to be symmetric. Potential isomorphisms can be uniquely defined in terms of atom preservation and object commutation:

[3]The motivation for taking a *reverse* inclusion $S \leq S'$ is to get a Galois connection between similarity relations and classes of invariants. See [3] for more.

[4]'Potential isomorphisms' are also called 'partial isomorphisms' in the literature; but we prefer to reserve the term 'partial isomorphism' for single isomorphisms between substructures.

Fact 5. *Iso_p is the smallest similarity relation which preserves atoms and commutes with objects expansions.*

Proof. Clearly, Iso_p preserves atoms and commutes with object expansions. Next, take an S that preserves atoms and commutes with object expansions. We want $Iso_p \leq S$. Let $\mathcal{M}$, $\mathcal{M}'$ be two structures with $\mathcal{M}\ S\ \mathcal{M}'$. We need to show that $\mathcal{M}\ Iso_p\ \mathcal{M}'$. By the definition of Iso_p, this amounts to finding a non-empty set I of partial isomorphisms between $\mathcal{M}$ and $\mathcal{M}'$ satisfying the back and forth properties. We set $I = \{f : |\mathcal{M}| \rightarrow |\mathcal{M}'| / \mathcal{M}, a_1, ..., a_n\ S\ \mathcal{M}', f(a_1), ..., f(a_n)\}$.

By hypothesis, $\mathcal{M}\ S\ \mathcal{M}'$, so I is non-empty, as it contains the empty function. Commutation with object expansions then yields the back and forth property. Let f be a function in I, it comes from two S-similar structures $\mathcal{M}$ and $\mathcal{M}'$ whose distinguished objects provide the arguments and values for f. Let a be an object in $|\mathcal{M}|$ which is not already a distinguished one. By commutation with object expansion, we can find an $a' \in |\mathcal{M}'|$ such that $\mathcal{M}, a$ and $\mathcal{M}', a'$ are again S-similar. Because of that, there is a function in I which extends f by sending a to a'. We can do the same starting from an $a' \in |\mathcal{M}'|$. Finally, since S preserves atoms, the f in I are partial isomorphisms. □

We would like to connect Fact 5 with a property of first-order *languages*. Let $Inv(S)$ be the class of classes of structures which are S-invariant, what is the property of $Inv(S)$ corresponding to commutation with object expansion? What does commutation with object expansions, which makes sense at the level of structures and similarity relations, amount to in terms of the associated logics, that is at the level of classes of structures and invariants ?

Definition 6. *Let Q be a class of structures of the form $\mathcal{M}, a$. The object projection of Q, $\exists(Q)$, is defined by $\mathcal{M} \in \exists(Q)$ iff there is a $b \in |\mathcal{M}|$ such that $\mathcal{M}, b \in Q$.*

Given a similarity relation S, we shall say that object projection preserves S-invariance iff, whenever $Q \in Inv(S)$, $\exists(Q) \in Inv(S)$ as well. Preservation of S-invariance under $\exists(-)$ says that existential quantification can be applied to invariant operations while staying inside the class of invariant operations. In other words, it says that existential quantification is available in the logic. In particular, if $Q_\exists \in Inv(S)$ and classes of structures definable from logical constants interpreted by classes of structures in $Inv(S)$ are again in $Inv(S)$, then $\exists(-)$ preserves invariance.[5] This is equivalent to commutation with object expansion:

Theorem 7. *S commutes with object expansions iff object projection preserves S-invariance.*

[5]For example, if Q is a class of structures of the form $\langle M, R, a\rangle$, $\overline{Q}$ applies to a formula and a term, binding two variables in the formula. $\exists(Q)$ is defined by $\exists z \overline{Q} xy \overline{R} xy, z$. It is easy to check that $\langle M, R\rangle \in \exists(Q)$ iff $\langle M, R\rangle \models \exists z \overline{Q} xy\ \overline{R} xy, z$.

Proof. For the original direct proof, see [3], Theorem 3.10 and Fact 3.14. A more general new proof will be provided in the next section. □

Putting together Fact 5 and Theorem 7, we get:

Corollary 8. *Iso_p is the smallest similarity relation S such that S preserves atoms and object projection preserves S-invariance.*

Imagine that we want to build a first-order language and that we want to base the interpretation of its logical constants on some appropriate notion of invariance. The Corollary says that if we want our language to admit of first-order existential quantification and to deal with atomic formulas in the standard way, Iso_p is the smallest similarity relation we can pick. So, in that sense, potential isomorphisms provide us with the most economic notion of similarity between structures making existential quantification available in the resulting logic.

What happens if we want second-order existential quantification as well? Then we shift from potential isomorphisms to isomorphisms. Here is the theorem which tells us this, where Iso is the similarity relation corresponding to 'being isomorphic' and 'set projection' is the same as object projection with subsets of the domain or relations over the domain replacing objects in the domain:[6]

Theorem 9. *Iso is the smallest similarity relation S such that S preserves atoms, and object and set projections preserve S-invariance.*

2 Invariance, commutation, and the modal case

2.1 Invariance and commutation: a general lemma

Theorem 7 rests upon a duality between two 'inverse' operations, (a) expanding a structure with an object, and (b) projecting a class of structures. We shall show that this idea of having two inverse operations in tandem, the second defined at a higher level, is all that is needed for commutation and preservation of invariance to be equivalent. One reason for going abstract here is that a characterization result for modal logic and bisimulation will follow automatically.

Let E be a relation over a class A of objects[7], with an associated 'inverse' function $E^{-1} : \wp(A) \rightarrow \wp(A)$ on the powerclass of A, defined for any

[6] The Theorem is due to the first author; it has been presented in the dissertation by the second author [2] with preservation of invariance under object projection and under set projection being called respectively closure under level 1 projection projection and closure under level 2 projection. Though Theorem 9 is quite analogous to Corollary 8, the proof does not rest on a commutation result like Theorem 7.

[7] We use 'class' rather than 'set' here because of the intended application to first-order structures, which constitute a proper class. The relation E echoes that of 'being an expansion with one object' over first-order structures which was at the heart of the previous section.

subclass X of A by $E^{-1}(X) = \{a \in A \mathbin{/} \exists b \in X \text{ with } aEb\}$. We shall be interested in the behavior of E and E^{-1} with respect to an equivalence relation S over A:

Definition 10. *S commutes with E iff, for all a, a', $b \in A$, if aSb and aEa', then there is a b' such that $a'Sb'$ and bEb'.*[8]

Following the notion of invariance introduced in the previous section, a subclass X of A is S-*invariant* if aSb implies $a \in X$ iff $b \in X$. We introduce preservation of invariance for E^{-1}:

Definition 11. *E^{-1} preserves S-invariance iff for any subclass X of A, if X is S-invariant, then $E^{-1}(X)$ is S-invariant.*

Now, for arbitrary S and E, commutation and preservation of invariance are equivalent – as may be shown by a little exercise in basic set theory:

Theorem 12. *S commutes with E iff E^{-1} preserves S-invariance.*

Proof. Only if. Assume that (1) S commutes with E, (2) X is S-invariant, (3) aSb and (4) $a \in E^{-1}(X)$. We want $b \in E^{-1}(X)$. By (4) and the definition of E^{-1}, there is an a' with aEa' and $a' \in X$. Hence by (3), we can apply (1) to get a b' such that $a'Sb'$ and bEb'. By (2) and $a' \in X$, $b' \in X$ as well. But bEb', therefore $b \in E^{-1}(X)$ as required.

If. Assume that (1) E^{-1} preserves S-invariance and that for some a, b and a', (2) aSb, (3) aEa'. We want a b' with $a'Sb'$ and bEb'. So, consider $[a']_S$, the S-equivalence class of a'. By definition, it is S-invariant. Hence by (1), $E^{-1}([a']_S)$ is S-invariant. By the definition of E^{-1} and (3), $a \in E^{-1}([a']_S)$. By S-invariance of $E^{-1}([a']_S)$ and (2), $b \in E^{-1}([a']_S)$. Hence by the definition of E^{-1}, there has to be a object b' with bEb' and $b' \in [a']_S$, that is $a'Sb'$. □

Now take as our class A the class of all first-order structures. Let E be the relation of 'expanding with one object', *i.e.* $\mathcal{M}\ E\ \mathcal{M}'$ iff $\mathcal{M}' = \mathcal{M}, a$ for some object $a \in |\mathcal{M}|$. Commutation with object expansion in the sense of Definition 4 is then commutation with E. And the 'inverse' E^{-1} of E is nothing but object projection $\exists(-)$. We get Theorem 7, which says that S commutes with object expansion iff object projection preserves S-invariance, as an instance of Theorem 12, which says that S commutes with E iff E^{-1} preserves S-invariance.

2.2 A modal application: diamonds and bisimulations

Moving on to our specific area of interest in this paper, Andréka, van Benthem and Németi have claimed that bisimulations are to modal logic what

[8] Note the formal similarity between this notion of relational linkage and that of a modal bisimulation, to be considered later.

potential isomorphisms are to predicate logic [1]. There are quite a number of meta-logical theorems and transfer results to back up such a claim. Is it then possible to characterize bisimulations as the good match for modal logic, just like we characterized potential isomorphisms as the good match for predicate logic, or plain isomorphisms as the good match for second-order logic? We provide a positive answer in this paragraph, by a direct application of Theorem 12.

Modal logic gets interpreted in Kripke models, that is, in the mono-modal case, structures $\langle W, R, P_1, ..., P_n\rangle$ with W a set of 'worlds', $R \subseteq W \times W$ the accessibility relation, and the $P_i \subseteq W$ interpretations for atoms. Modal formulas are evaluated at worlds in Kripke models, so their interpretation in a model is the set of worlds at which they are true. In line with our earlier general view, then, modal logical constants get interpreted by classes of pointed structures. As an example, $\Diamond$ is the unary constant which is interpreted by the class $Q_\Diamond$ of structures $\langle W, R, P, w\rangle$ with $\{w' \in W \ / \ wRw' \text{ and } w' \in P\} \neq \emptyset$. In terms of truth clauses, our earlier analysis of quantifiers then comes to look as follows:

$$\mathcal{M}, w \vDash \Diamond\phi$$
$$\text{iff}$$
$$\exists w' \ wRw' \text{ and } \mathcal{M}, w' \vDash \phi$$
$$\text{iff}$$
$$\langle W, R, ||\phi||_{\mathcal{M}}, w\rangle \in Q_\Diamond$$

where $||\phi||_{\mathcal{M}} = \{v \in |\mathcal{M}| \ / \ \mathcal{M}, v \vDash \phi\}$.

Modal similarity relations, such as bisimulations, link two worlds in two models: they are equivalence relations between *pointed* Kripke structures. As before, a class Q of pointed Kripke structures is *invariant under a similarity relation* S on such structures iff $\mathcal{M}, w \ S \ \mathcal{M}', w'$ implies $\mathcal{M}, w \in Q$ iff $\mathcal{M}', w' \in Q$. Given a modal similarity relation S, we shall note $Inv_M(S)$ the class of classes of pointed structures which are S-invariant. Note that our previous ordering on similarity relations, as well as the property of atom preservation, apply straightforwardly to the particular case of modal similarity relations.

The two essential properties of bisimulations are commutation with 'guarded object expansion' (moves along the accessibility relation) and preservation of atoms. By commutation with guarded object expansion, we mean the following:

Definition 13. *A similarity relation S commutes with guarded object expansion iff*
if $\mathcal{M}, w \ S \ \mathcal{M}', w'$, then for all $v \in |\mathcal{M}|$ with wRv, there is a $v' \in |\mathcal{M}'|$ such that $\mathcal{M}, v \ S \ \mathcal{M}', v'$ and $w'R'v'$.

Commutation with *guarded* object expansion is to bisimulation what commutation with object expansion is to potential isomorphism. This makes for the following modal version of Fact 5 (we write BiS for the relation 'being bisimilar', which holds between any two pointed structures which have a bisimulation between them):

Fact 14. *BiS is the smallest modal similarity relation which preserves atoms and commutes with guarded object expansion.*

Proof. First it is clear that BiS preserves atoms (worlds which are related by a simulation belong to the same unary predicate extensions) and commutes with guarded object expansion (by the back and forth properties of modal bisimulation).

Now let S be a similarity relation over pointed Kripke models which preserves atoms and commutes with guarded object expansion. We want $BiS \leq S$. Assume $\mathcal{M}, w\ S\ \mathcal{M}', w'$. We show that $\mathcal{M}, w\ BiS\ \mathcal{M}', w'$. S 'is' the bisimulation we need: we define a relation Z over $W \times W'$ by uZu' iff $\mathcal{M}, u\ S\ \mathcal{M}', u'$. Z contains $\langle w, w' \rangle$ by hypothesis. It preserves atoms because S does and it satisfies the back and forth condition in the definition of bisimulations precisely because S commutes with guarded object expansion. □

Guarded object expansion can be viewed as a relation E over the class of pointed structures, defined by $\mathcal{M}, w\ E\ \mathcal{M}', w'$ iff $\mathcal{M} = \mathcal{M}'$ and wRw'. Its associated inverse E^{-1} on classes of pointed structures is *guarded object projection*, defined by $E^{-1}(Q) = \{\mathcal{M}, w\ /\ \mathcal{M}, w' \in Q$ for some $w' \in |\mathcal{M}|$ with $wRw'\}$. As before with object projection and first-order existential quantification, guarded object projection is the result of applying existential modal quantification. To see this, assume that $Q_\Diamond$ is in $Inv_M(S)$ and that classes of pointed structures which are definable from modal constants interpreted by classes of pointed structures in $Inv_M(S)$ are again in $Inv_M(S)$. It follows that E^{-1} preserves S-invariance.[9]

Thus, as an instance of Theorem 12, we get the equivalence between a) the core property of bisimulations, namely commutation with moves along the accessibility relation, and b) the invariants of S being closed under applications of $\Diamond$:

Theorem 15. *S commutes with guarded object expansion iff guarded object projection preserves S-invariance.*

Now putting together Fact 14 and Theorem 15, we get:

Corollary 16. *BiS is the smallest similarity relation S such that S preserves propositional atoms, while guarded object projection preserves S-invariance.*

[9]For example, let Q be a class of structures of the form $\langle M, R, P, a \rangle$ which is the interpretation of a unary modal constant $\overline{Q}$. It is easy to define $E^{-1}(Q)$ from $\Diamond$ and $\overline{Q}$. Check that $\langle M, R, P, w \rangle \in E^{-1}(Q)$ iff $\langle M, R, P, w \rangle \vDash \Diamond\overline{Q}\, p$ where P is the interpretation of p.

As before with potential isomorphisms and first-order logic, this tells us that bisimulations are the right match for modal logic. For a logic to be a modal logic, it seems clear that it should deal with atoms in the standard way and have $\Diamond$ as a logical symbol (so that application of $\Diamond$ preserves invariance). BiS is the least demanding notion of similarity between pointed models which fits the bill.

3 Back and forth between back and forths

Our aim in this section is to provide a better understanding of the relationships between our two characterization results for FOL and ML (Corollaries 8 and 16).

3.1 Generalized assignment models

We follow the idea of 'Modal Foundations for Predicate Logic' (cf. [8]), and adopt a modal reading for first-order semantics. The starting point is the observation that key clauses of Tarskian semantics for first-order logic like:

$$\begin{gathered}\mathcal{M} \vDash \exists x\ \phi(x)\ \sigma \\ \text{iff} \\ \text{for some } a \in |\mathcal{M}|,\ \mathcal{M} \vDash \phi(x)\ \sigma[x := a]\end{gathered}$$

have a modal flavor which is revealed by the following rewriting:

$$\begin{gathered}\mathcal{M}, \sigma \vDash \exists x\ \phi(x) \\ \text{iff} \\ \text{for some } \tau,\ \sigma R_x \tau \text{ and } \mathcal{M}, \tau \vDash \phi(x)\end{gathered}$$

with assignments σ, τ viewed as abstract states and R_x as a relation over these which corresponds to updating the value of x. Thus, any classical model $\mathcal{M}$ induces a (poly-modal) assignment model $\mathcal{M}^* = \langle S, \{R_x\}_{x \in VAR}, I\rangle$ with S a set of states,[10] R_x a binary relation for each variable x, and I a valuation function which gives a truth value to each atomic formula Px in each state σ, so that $\mathcal{M}^*, \sigma \vDash Px$ iff $\sigma(x) \in P_{\mathcal{M}}$. Then there is a natural translation $(-)^*$ of FOL formulas into ML formulas such that $\mathcal{M} \vDash \phi\ \sigma$ iff $\mathcal{M}^*, \sigma \vDash \phi^*$. $(-)^*$ is defined by induction and the clause for $\exists$ turns it into a diamond. One can check that:

$$\begin{gathered}\mathcal{M} \vDash \exists x\ \phi(x)\ \sigma \\ \text{iff} \\ \mathcal{M}^*, \sigma \vDash \Diamond_x(\phi(x))^*\end{gathered}$$

[10] In general, S will be the set of all assignments on $\mathcal{M}$, but the point of 'modal foundations' is that one can just as well use only some subset of the set of all assignments. This both models interesting phenomena of 'dependence' between variables, and leads to well-behaved decidable versions of first-order logic.

This idea extends to generalized quantifiers as well. Without loss of generality, recall the satisfaction clause for a binary quantifier $\overline{Q}$ interpreted by an operation Q:

$$\mathcal{M} \vDash \overline{Q}xy\ \phi(x,y)\ \sigma$$
$$\text{iff}$$
$$\langle M, ||\phi(x,y)||_{\mathcal{M},\sigma}\rangle \in Q$$

Once the bound variables x, y are fixed, the action of Q on M consists in manipulating subsets of the set of all M-assignments $\mathbb{V}_M$. Looking at these as abstract states connected by update relations, we have a class of pointed assignment models Q^*_{xy}:

$$\{\langle \mathbb{V}_M, R_{xy}, P, \sigma\rangle \ /\ \langle M, \{\langle \tau(x), \tau(y)\rangle \ /\ \sigma R_{xy}\tau \text{ and } \tau \in P\}\rangle \in Q\}$$

where R_{xy} updates the values of both x and y in one shot. Now, Q^*_{xy} is a class of pointed Kripke structures just like those in section 2, so it interprets a generalized modal quantifier $\overline{Q^*_{xy}}$, the modal translation of $\overline{Q}xy$, and one can check that we still have what we had for $\exists x$ and $\Diamond_x$, namely:

$$\mathcal{M} \vDash \overline{Q}xy\ \phi(x,y)\ \sigma$$
$$\text{iff}$$
$$\mathcal{M}^*, \sigma \vDash \overline{Q^*_{xy}}(\phi(x,y))^*$$

Here, as before, $\mathcal{M}^*, \sigma \vDash \overline{Q^*_{xy}}(\phi(x,y))^*$ iff $\langle M, R_{xy}, ||(\phi(x,y))^*||_{\mathcal{M}}, \sigma\rangle \in Q^*_{xy}$.

Let us be more precise about the *-models. Fix a countable set of variables $VAR = \{x_1, ..., x_n, ...\}$. Any model $\mathcal{M}$ induces an *associated assignment model* $\mathcal{M}^*$:

– The domain $\mathbb{V}$ of $\mathcal{M}^*$ is the set of *finite* partial functions from VAR to $|\mathcal{M}|$.

– Each finite[11] set of variables $X \subset VAR$ induces an accessibility relation R_X with $\sigma R_X \tau$ iff τ extends σ at most on values for variables in X and differs from σ at most on variables in X. Thus, R_X corresponds to updating registers for some variables in X and creating new registers for other variables in X.

– For each n-ary relation R on the structure $\mathcal{M}$, and each n-tuple of variables x_{i_1},...,x_{i_n}, a new predicate extension $R_{i_1,...,i_n}$ on $\mathbb{V}$ is defined by setting $\sigma \in R_{i_1,...,i_n}$ iff
$\langle \sigma(x_{i_1}), ..., \sigma(x_{i_n})\rangle \in R$.

Now, a first-order formula ϕ can be evaluated with respect to any partial assignment σ defined on the *free* variables in ϕ. If, when testing for the

[11] Indexing accessibility relations with sets of variables instead of one variable at a time is necessary to handle polyadic quantifiers, which have an independent meaning in this setting.

satisfaction of ϕ w.r.t. σ, a quantifier Q occurs and binds a variable x not in the domain of σ, one considers partial assignments extending σ by giving a value to x. This is what is captured by our definition of R_X on assignment models, which corresponds either to register updating (the partial assignment was already defined for the bound variable) or register opening (the partial assignment was not defined for the bound variable). Our *-translation of first-order structures into Kripke structures is compatible with the outlined matching *-translation of a first-order language with generalized quantifiers into a modal language with generalized modal quantifiers.

3.2 From potential isomorphisms to bisimulations

$\mathcal{M}^*$-models are very special structures: their domain consists of *all* partial assignments and the accessibility relations R_X make *all* actions of updating and creating registers available. As in [8], one can explore what happens to generalized modal quantifiers when models can have 'assignment gaps'. We leave this for future research, and focus on similarity relations for the above full $\mathcal{M}^*$ models.

Exactly like quantifiers, similarity relations on classical models can be exported to assignment models. Thus, to each similarity relation S, we define its *associated similarity relation* S^* by $\mathcal{M}^*, \sigma\ S^*\ \mathcal{M}'^*, \sigma'$ iff σ and σ' have the same domain and $\mathcal{M}, \overrightarrow{\sigma(x_i)}\ S\ \mathcal{M}', \overrightarrow{\sigma'(x_i)}$ for the $x_i s$ in their domain.

Theorem 17. $Iso_p^* = BiS$

Proof. We need to prove $\mathcal{M}^*, \sigma \leftrightarroweq \mathcal{M}'^*, \sigma'$ iff $\mathcal{M}, \overrightarrow{\sigma(x_i)} \approx \mathcal{M}', \overrightarrow{\sigma'(x_i)}$.

From left to right. To match objects in the initial models, we have to look at the bisimulation and match objects according to their indexing by the variables. So assume $\mathcal{M}^*, \sigma \overset{Z}{\leftrightarroweq} \mathcal{M}'^*, \sigma'$. We set $I = \{f \ / \ \exists \rho \in |\mathcal{M}^*|, \rho' \in |\mathcal{M}'^*| \text{ s.t. } \rho Z \rho' \text{ and } f = \{\langle \rho(x), \rho'(x) \rangle \ / \ x \in Dom(\rho) \bigcap Dom(\rho')\}\}$. Since $\sigma\ Z\ \sigma'$, I contains at least the empty function (thanks to the worst case scenario in which there is no variable at which both σ and σ' are defined). The functions in I are partial isomorphisms, because Z itself respects atoms.

Now for the back and forth condition: let $f \in I$ and $a \in |\mathcal{M}|$ be an object which is not in the domain of f. We know there are two assignments τ and τ' with $\tau Z \tau'$ that gave us f. Starting from τ, we open a register for a variable x which was not in the domain of τ and give it the value a. This is a move along R_x to an assignment ρ which extends τ on x. Since $\tau Z \tau'$, the same move can be made along R'_x, and we get an assignment ρ' which extends τ' on x. From ρ, ρ' and the fact that $\rho Z \rho'$, we get a function g in I which extends f on a by $g(a) = \rho'(x)$. Therefore I is a potential isomorphism between $\mathcal{M}$ and $\mathcal{M}'$ – and, by construction, these two structures extended with objects that are indexed by the same variable according to σ and σ' are again potentially isomorphic.

From right to left. To match states in the assignment models, we have to look at the partial isomorphisms, and match assignments according to the partial isomorphisms and the indexing by the variables. Assume that $\mathcal{M} \overset{I}{\approx} \mathcal{M}'$. We define our intended bisimulation Z over $|\mathcal{M}^*| \times |\mathcal{M}'^*|$ by $\sigma Z \sigma'$ iff $Dom(\sigma) = Dom(\sigma')$ and there is an $f \in I$ s.t. $\sigma' = f \circ \sigma$. Since the f in I are partial isomorphisms, Z-related states verify exactly the same atoms.

Now for the Zig-Zag condition. Assume that $\sigma Z \sigma'$. By the definition of Z, this means that there is an $f \in I$ such that $\sigma' = f \circ \sigma$. We check the Zig-Zag condition for R_x. So assume that $\sigma R_x \rho$ for some ρ. Suppose that x is not a fresh variable, being already in the domain of σ and σ'. We look at $\rho(x)$, say this is a. By the closure condition on I, there is a function $g \in I$ such that $f \subseteq g$ and a is in the domain of g. Now consider $\rho' = \sigma'[x := g(a)]$. It is clear that $\sigma' R_x \rho'$. Since $f \subseteq g$, $\rho = \sigma[x := a]$ and $\rho' = \sigma'[x := g(a)]$, $\sigma' = f \circ \sigma$ is sufficient to guarantee that $\rho' = g \circ \rho$. Hence $\rho Z \rho'$, as desired. If x is a fresh variable, things are similar with register opening (extending assignments) replacing register updating. $\square$

The left to right direction of the proof would not go through as it stands if we were to take *total* assignments as states. From right to left, if the relations R_x only correspond to updating (and not to register opening), a bisimulation between the *-models does *not* guarantee a potential isomorphism between the first-order structures. We leave possible generalizations to a future occasion.

Theorem 17 confirms that bisimulations are a modal version of potential isomorphisms. Moreover, one can check that invariance is preserved by shifting to assignment models. So 'genuine' first-order operations, that is classes of structures invariant under potential isomorphisms, induce 'genuine' modal operations: classes of pointed assignment models invariant under bisimulation.

3.3 From bisimulations to potential isomorphisms

We can also go in the other direction and 'upgrade' bisimulations to potential isomorphisms by considering suitably richly structured Kripke models. We adapt a result from [1] that modal equivalence can be upgraded to full first-order elementary equivalence on trees with multiplied nodes. They leave a theorem for bisimulations and potential isomorphisms as an unproved claim. We prove it directly, for trees with suitably multiplied nodes.[12]

In what follows, we will work with the multiplied unraveled version $\mathcal{M}^+$ of a Kripke model $\mathcal{M}$ as in [1]. First, the standard transformation of *tree unraveling* is performed on $\mathcal{M}$. We get a model whose worlds are finite

[12] Proving the theorem about elementary equivalence involves keeping track of distances between nodes to be matched along partial isomorphisms. With potential isomorphism, in contrast with finite Ehrenfreucht-Fraïssé games, this is not necessary and the proof is simpler.

sequences of the form $w_0, ..., w_n$ with $w_0 = w$ and each w_{i+1} is an R-successor of w_i ($0 \leq i < n$), whose accessibility relation is 'immediate succession'. Worlds in the new model bisimulate with worlds in the original model via their last element. In addition, infinite 'multiplication' is applied to each node except the root, as follows, maintaining a bisimulation at each stage. First, copy each successor of w at level 1 infinitely many times and attach these disjoint copies to w. Identifying copies with originals is an obvious bisimulation. Next consider successors at level 2 on all branches of the previous stage and perform the same copying process at all level-1 worlds. Again, there is an obvious bisimulation. The intended model $\mathcal{M}^+, w$ is the result of iterating this process through all finite levels.

Theorem 18. *Two pointed Kripke models $\mathcal{M}, w$ and $\mathcal{M}', w'$ are bisimilar iff their multiplied unraveled versions $\mathcal{M}^+, w$ and $\mathcal{M}'^+, w'$ are potentially isomorphic.*

Proof. It is clear that if $\mathcal{M}^+, w$ and $\mathcal{M}'^+, w'$ are potentially isomorphic, the original models $\mathcal{M}, w$ and $\mathcal{M}', w'$ are bisimilar. For, the unconstrained exploration which corresponds to a potential isomorphism is more than we need for the constrained exploration which corresponds to a bisimulation.

In the other direction, assume that $\mathcal{M}, w$ and $\mathcal{M}', w'$ are bisimilar. Then there is a bisimulation Z between their unraveled multiplied versions. We need to show that there is a potential isomorphism as well.

Let us say that a partial isomorphism f from $\mathcal{M}^+$ to $\mathcal{M}'^+$ *follows* Z iff f relates worlds which are related by Z and the domain and range of f are closed under subsequences: if $\langle w_0, ..., w_j\rangle \in Dom(f)$, $\langle w_0, ..., w_i\rangle \in Dom(f)$ for $0 \leq i \leq j$ (and similarly for the range). We show that the set of *finite* partial isomorphisms following Z is a potential isomorphism between $\mathcal{M}^+, w$ and $\mathcal{M}'^+, w'$.

First, it is non-empty since $\langle w, w'\rangle$ is trivially a partial isomorphism following Z. Then consider any finite partial isomorphism f following Z and a new object a. We need to find a finite partial isomorphism g which extends f, follows Z, and has a in its domain. Now there is a unique path from the root to a, and on this path, a lowest node b to be in the domain of f. To get the image of a, one goes down from the image of b by following along Z the path from b to a, taking new nodes on $|\mathcal{M}'^+|$ to match the new nodes on $|\mathcal{M}^+|$. More precisely, let f be a finite partial isomorphism following Z and $a = \langle w_0, ..., w_n\rangle$ with $w_0 = w$ a world in $|\mathcal{M}^+|$ which is not in the domain of f. Since f is closed under subsequence, there is an $i \in \{0, ..., n\}$ s.t. for all $j \leq i$, $\langle w_0, ..., w_j\rangle \in Dom(f)$, and for all $k > i$, $\langle w_0, ..., w_k\rangle \notin Dom(f)$. Let $\langle w'_0, ..., w'_i\rangle = f(\langle w_0, ..., w_i\rangle)$. Since $\langle w_0, ..., w_i\rangle Z \langle w'_0, ..., w'_i\rangle$ and $\langle w_0, ..., w_i\rangle R^+ \langle w_0, ..., w_{i+1}\rangle$, there is a w'_{i+1} s.t. $\langle w'_0, ..., w'_i\rangle R'^+ \langle w'_0, ..., w'_{i+1}\rangle$ and $\langle w_0, ..., w_{i+1}\rangle Z \langle w'_0, ..., w'_{i+1}\rangle$. Moreover, since $\mathcal{M}'^+$ is a tree whose nodes have been copied infinitely many times, we can choose w'_{i+1} so that it was *not* already in the range of f. Repeating this, we get a sequence $w'_{i+1}, ..., w'_n$ matching the sequence

$w_{i+1}, ..., w_n$. We define $g = f \cup \{\langle\langle w_0, ..., w_k\rangle, \langle w'_0, ..., w'_k\rangle\rangle \,/\, i+1 \leq k \leq n\}$. It is crucial here that none of the added elements were already in the domain or in the range of f, so that g is a one-one function.

To clinch matters, we show that g is a finite partial isomorphism following Z:
– Since g extends f on a finite number of arguments and f is finite, g is finite.
– The domain and range of g are closed under subsequences, by the construction and because those of f were. Also, g relates worlds which are related by Z, by construction and because f did. Hence g respects unary predicates.
– g respects the accessibility relations as well. For, take b and c in the domain of g. We want bR^+c iff $g(b)R^{'+}g(c)$. We reason by cases, depending on whether the worlds are new. If b and c were already in the domain of f, there is nothing to show. If they are both new, so are $g(b)$ and $g(c)$ by construction, and hence the equivalence follows from the definition of g and the models being unraveled trees. Now assume (1) only one of them, say b, was in the domain of f. It follows that (2) $g(b)$ was in the range of f, but $g(c)$ was not. The idea is that bR^+c iff b is the last node on the path from the root to a which is in the domain of f and c is its successor on the path, and similarly for $g(b)$ and $g(c)$ with respect to $R^{'+}$ and the range of f. More precisely, by (1), bR^+c iff $b = \langle w_0, ..., w_i\rangle$ and $c = \langle w_0, ..., w_{i+1}\rangle$. By (2), $g(b)R^{'+}g(c)$ iff $g(b) = \langle w'_0, ..., w'_i\rangle$ and $g(c) = \langle w'_0, ..., w'_{i+1}\rangle$. Hence bR^+c iff $g(b)R^{'+}g(c)$. $\square$

So $\mathcal{M}^+, w$ and $\mathcal{M}'^+, w'$ are 'potentially' isomorphic. In fact, they are almost *isomorphic* except for the difference that the two initial models may have successor sets of different infinite cardinalities. Potential isomorphisms are blind to differences in size for infinite sets, but to get isomorphic models, we could strengthen the copying procedure, for example by making not $\aleph_0$ copies of each node but κ-copies, where $\kappa = max(card(|\mathcal{M}|), card(|\mathcal{M}'|))$. But even as we have stated things, our conclusion is that modal bisimulations naturally induce basic first-order invariance relations.

4 Dynamic logic and safety

Our final task is to extend our characterization of invariance for modal logic to *dynamic logics* which have explicit operations on accessibility relations. This makes sense from the perspective of FOL, where formulas can define both sets of objects and relations between objects (depending on how many free variables they have): the logic handles all these semantic types in a uniform way. By contrast, in ML proper, formulas only define *sets* of worlds, while accessibility relations cannot be manipulated. For the latter purpose, propositional dynamic logic (PDL) introduces a new type of expression, viz. 'programs' which define relations. How do our invariance conditions apply to the new program operations so as to make expressions of different types combine nicely? We shall first see how Theorem 12 works for dynamic operations and then connect this with the standard notion of 'safety for bisimulation'. In the last paragraph, we draw a tentative parallel with first-order logic.

4.1 Invariance and commutation

In Section 2, classes of pointed structures were used to interpret *modal* operators. What is the interpretation of *dynamic* operators, such as relational composition ; or Boolean choice $\cup$ in PDL? Their role is to make new relations available, on the basis of the accessibility relations and predicate extensions which are already there in the Kripke structures. So a dynamic operation O will associate to every set W a function O_W taking as arguments relations and predicate extensions over W and yielding a new relation over W as its value. To such an operation O will correspond a dynamic operator $\overline{O}$ with the following semantic clause for an arbitrary model $\mathcal{M}$:

$$||\overline{O}\overrightarrow{\chi}||_{\mathcal{M}} = O_{|\mathcal{M}|}(||\overrightarrow{\chi}||_{\mathcal{M}})$$

with $\overrightarrow{\chi}$ a sequence of programs and formulas matching the syntactic type of $\overline{O}$ (and the type of the function $O_{|\mathcal{M}|}$), and $||\overrightarrow{\chi}||_{\mathcal{M}}$ the sequence of relations and predicate extensions over $\mathcal{M}$ which interpret the programs and formulas $\overrightarrow{\chi}$. For instance, Boolean choice $\cup$ gets interpreted by the dynamic operation $O_\cup$ which yields for each W a function $O_{\cup,W} : \wp(W^2), \wp(W^2) \to \wp(W^2)$ defined for $R_1, R_2 \subseteq W^2$ by $O_{\cup,W}(R_1, R_2) = R_1 \bigcup R_2$.

In the well-known syntax of dynamic logic, a program π and formula ϕ combine into a new formula $\langle\pi\rangle\phi$, with $\langle\pi\rangle$ interpreted as a standard $\Diamond$ for the accessibility relation defined by π. When such operators are added to a modal language, it is natural to require that the new formulas retain the old semantic invariance. Indeed, PDL formulas are still invariant under bisimulations, and hence the increase in expressive power stays within the 'natural limits' of ML.

Thus, on classes of pointed structures which are invariant for bisimulation, dynamic operators should preserve that invariance when used to define new classes of structures. What dynamic operators $\overline{O}$ have this property? One answer is again provided by Theorem 12. In Section 2.2, we took E and E^{-1} to be "moving along the accessibility relation" and "applying $\Diamond$", respectively. Now interpret E as "moving along the relation defined by $\overline{O}-$". E^{-1} will then be "applying $\langle\overline{O}-\rangle$". More precisely, let $\overrightarrow{\chi}$ be a sequence of programs and formulas matching the type of O[13] and ϕ a formula. We use the notation $\mathcal{M} + ||\overrightarrow{\chi}||_{\mathcal{M}}, w$ for the expansion of $\mathcal{M}, w$ with the relations

[13]As before, the syntactic type of $\overline{O}-$ is arbitrary: $\overline{O}-$ applies to a certain number of formulas and programs and yields a new program. Of course, then, the semantic type of O has to match the syntactic type of $\overline{O}-$.

and sets which are the interpretations on $\mathcal{M}$ of the formulas and programs $\overrightarrow{\chi}$. Define E by $\mathcal{M} + ||\overrightarrow{\chi}||_{\mathcal{M}}, w \; E \; \mathcal{M} + ||\overrightarrow{\chi}||_{\mathcal{M}}, v$ iff $wO_{\mathcal{M}}(||\overrightarrow{\chi}||_{\mathcal{M}})v$. Then $E^{-1}(||\phi||) = ||\langle\overline{O}\overrightarrow{\chi}\rangle\phi||$, since $\mathcal{M} + ||\overrightarrow{\chi}||_{\mathcal{M}}, w \in E^{-1}(||\phi||)$ iff there is a $v \in |\mathcal{M}|$ with $wO_{|\mathcal{M}|}(||\overrightarrow{\chi}||_{\mathcal{M}})v$ and $\mathcal{M} + ||\overrightarrow{\chi}||_{\mathcal{M}}, v \in ||\phi||_{\mathcal{M}}$ iff $\mathcal{M}, w \in ||\langle\overline{O}\overrightarrow{\chi}\rangle\phi||$.

Let us say that BiS commutes with O iff it commutes with the relations one gets for all values of O.[14] We say that O preserves invariance under BiS iff applying the interpretation of $\langle\overline{O}\overrightarrow{\chi}\rangle$ preserves invariance for all possible interpretations of the $\overrightarrow{\chi}$.[15] In this setting, Theorem 12 tells us that commutation with O is precisely what guarantees that adding $\overline{O}$ does not break bisimulation invariance:

Theorem 19. *BiS commutes with O iff O preserves invariance under BiS.*

This result provides the key induction step in a proof that extending a modal logic by dynamic operators whose interpretations commute with bisimulation yields a dynamic logic whose formulas are again invariant under bisimulation. This would also work with BiS any kind of similarity relation for pointed Kripke structures.

4.2 Commutation and safety for bisimulation

The standard 'invariance condition' on PDL operations looks a bit different, however, and its is known as 'safety under bisimulation'.[16]

Definition 20. *A dynamic operation O is safe for bisimulation iff whenever $\mathcal{M}, w \overset{Z}{\leftrightarrows} \mathcal{M}', w'$, and $wO(\mathcal{M})v$ for some $v \in |\mathcal{M}|$, then there is a $v' \in |\mathcal{M}'|$ such that vZv' and $w'O(\mathcal{M}')v'$ – as well as vice versa.*

We will now show that commutation and safety are equivalent.[17] In what follows, it will be useful to work with 'quasi-transitive' bisimulations where wZw', vZw' and vZv' implies wZv'. More precisely, let W and W' be

[14] Thus, for any sequence of programs and formulas $\overrightarrow{\chi}$, if $\mathcal{M} + ||\overrightarrow{\chi}||_{\mathcal{M}}, w$ and $\mathcal{M}' + ||\overrightarrow{\chi}||_{\mathcal{M}'}, w'$ are bisimilar, $wO_{|\mathcal{M}|}(||\overrightarrow{\chi}||_{\mathcal{M}})v$ implies that there is a $v' \in |\mathcal{M}'|$ such that $\mathcal{M} + ||\overrightarrow{\chi}||_{\mathcal{M}}, v$ and $\mathcal{M}' + ||\overrightarrow{\chi}||_{\mathcal{M}'}, v'$ are bisimilar and $w'O_{|\mathcal{M}'|}(||\overrightarrow{\chi}||_{\mathcal{M}'})v'$.

[15] Let $\overrightarrow{\chi}$ be a sequence of programs and formulas and Q a bisimulation invariant class of structures of the form $\mathcal{M} + ||\overrightarrow{\chi}||_{\mathcal{M}}, w$. We want the interpretation of $\langle\overline{O}\overrightarrow{\chi}\rangle$ to be bisimulation invariant. This is exactly requiring bisimulation invariance for the class of structures $\mathcal{M} + ||\overrightarrow{\chi}||_{\mathcal{M}}, v$ having a w with $vO_{|\mathcal{M}|}(||\overrightarrow{\chi}||_{\mathcal{M}})w$ and $\mathcal{M} + ||\overrightarrow{\chi}||_{\mathcal{M}}, w \in Q$.

[16] See the 'Safety Theorem' of [8] characterizing the first-order safe operations as those operations which are definable in PDL without Kleene star.

[17] There are two differences in the discussion to follow. Commutation is not only about bisimilarity between the arguments for O (the $\overrightarrow{K}$ in the preceding footnote), but also about bisimilarity between arguments for O *plus* arbitrary extra-structure ($\mathcal{M} + \overrightarrow{K}$). On the other hand, safety is not only about mere bisimilarity but about *particular* bisimulations Z. Our results show that these two 'strengthenings' are in fact equivalent. See [7] for more on the model-theoretic relationships between safety and invariance.

two sets and Z a relation on $W \times W'$. We say that $v \in W$ and $v' \in W'$ are Z*-connected* if there is a sequence of objects $v_0, ..., v_n$ with $v_0 = v$, $v_n = v'$ such that for each $i < n$, $v_i Z v_{i+1}$ or $v_{i+1} Z v_i$. As a special case, every object is Z-connected to itself. A bisimulation Z between two Kripke structures $\mathcal{M}, w$ and $\mathcal{M}', w$ is *quasi-transitive* iff, if $v \in |\mathcal{M}|$ and $v' \in |\mathcal{M}'|$ are Z-connected, then vZv'. The following lemma shows that there is no loss of generality as far as safety is concerned:

Lemma 21. *An operation O is safe for arbitrary bisimulations iff it is safe for quasi-transitive bisimulations.*

Proof. The direction from left to right is trivial, so we take on the converse.

First, note that if Z is a bisimulation, then there is a smallest quasi-transitive relation Z^{Cl} extending Z (here two worlds, one in each model, are called Z^{Cl}*-related* iff they are Z-connected).

Let O be a dynamic operation which is safe for quasi-transitive bisimulation, $\mathcal{M}$ and $\mathcal{M}'$ two models, Z a bisimulation between them; with $w, v \in |\mathcal{M}|$ and $w' \in |\mathcal{M}'|$ three worlds such that wZw' and $wO(\mathcal{M})v$. We need to find a $v' \in |\mathcal{M}'|$ such that vZv' and $w'O(\mathcal{M}')v'$. Note that working directly with Z^{Cl} would not do. By safety for quasi-transitive bisimulation, there is a $v' \in |\mathcal{M}'|$ such that v' is Z-connected to v and $w'O(\mathcal{M}')v'$. But this v' is not necesarily Z*-related* to v, if Z is not quasi-transitive. To circumvent this difficulty, we shall define some model expansions in order to go 'step by step' from $v \in |\mathcal{M}|$ to a $v' \in |\mathcal{M}'|$ with vZv' while preserving the property of being O-related to the root. We cannot but take a roundabout way here: there are pairs of models and relations over them which are safe for quasi-transitive bisimulations, but not for arbitrary bisimulations between these models.

First, we define $\mathcal{M} + v^*$ as $\mathcal{M}$ extended by a world v^* which is a copy of v: for all predicates P, $v^* \in P_{\mathcal{M}+v^*}$ iff $v \in P_{\mathcal{M}}$, for all relations R, $wR_{\mathcal{M}+v^*}v^*$ iff $wR_{\mathcal{M}}v$, for all relations R and for all worlds $u \in |\mathcal{M}|$, $v^*R_{\mathcal{M}+v^*}u$ iff $vR_{\mathcal{M}}u$. We can assume that it *is* the case that $wO(\mathcal{M}+v^*)v^*$. To see this, consider the identity on $\mathcal{M}$ extended by $\langle v, v^* \rangle$. This is a quasi-transitive bisimulation between $\mathcal{M}$ and $\mathcal{M} + v^*$. We know that $wO(\mathcal{M})v$ and our bisimulation relates v in $\mathcal{M}$ to only two worlds in $\mathcal{M} + v^*$, namely v itself and v^*. So by safety of O for quasi-transitive bisimulation, w has to be related by $O(\mathcal{M} + v^*)$ to *at least one* of these two worlds. But if w was related by $O(\mathcal{M} + v^*)$ to *only one* of v and v^*, we would get a counter-example to safety of O for quasi-transitive bisimulation by considering the identity bisimulation between $\mathcal{M} + v^*$ and itself, with v and v^* being swapped. Therefore, w is related by $O(\mathcal{M} + v^*)$ to *both* v and v^*.

Then we build a model $\mathcal{M}' + v^*$ by adding a copy of v to the second model as well. To define this, we use the bisimulation Z we have: the copy is a world like the worlds to which v is Z-similar. Thus, let $Z(v)$ be the set of worlds in $\mathcal{M}'$ to which v is Z-related. We start from $\mathcal{M}'$ and add a world v^* such that for all predicates P, $v^* \in P_{\mathcal{M}'+v^*}$ iff $v \in P_{\mathcal{M}}$, for all relations

R, $w' R_{\mathcal{M}'+v^*} v^*$ iff $w R_{\mathcal{M}} v$ (where w' is the world bisimilar to w we introduced at the beginning), and for all relations R, for all worlds $u' \in |\mathcal{M}'|$, $v^* R_{\mathcal{M}'+v^*} u'$ iff there is a $t' \in Z(v)$ with $t' R_{\mathcal{M}'} u'$. We can use Z to build a quasi-transitive bisimulation Z^* between $\mathcal{M} + v^*$ and $\mathcal{M}' + v^*$. Consider first $Z \bigcup \{\langle v^*, v^* \rangle\}$. Because Z was a bisimulation and by the definition of our expansions, it is a bisimulation again. Now $(Z \bigcup \{\langle v^*, v^* \rangle\})^{Cl}$ is again a bisimulation and it is quasi-transitive. Moreover, it is clear that v^* in $\mathcal{M} + v^*$ is related only to v^* in $\mathcal{M}' + v^*$, so by safety for quasi-transitive bisimulation, $w' O(\mathcal{M}' + v^*) v^*$.

Finally, we compare $\mathcal{M}'$ and $\mathcal{M}' + v^*$. Consider the identity on $\mathcal{M}'$ extended by relating all worlds in $Z(v)$ to all worlds in $Z(v) \bigcup \{v^*\}$. Because of the properties we gave to v^*, this is a bisimulation between $\mathcal{M}'$ and $\mathcal{M}' + v^*$. And it is quasi-transitive by definition. Therefore, by safety of O for quasi-transitive bisimulation again, $w' O(\mathcal{M}' + v^*) v^*$ implies that there is a $u' \in |\mathcal{M}'|$ which is related by that bisimulation to v^* and which is such that $w' O(\mathcal{M}') u'$. Such a u' is in $Z(v)$. So we have a u' with $v Z u'$ and $w' O(\mathcal{M}') u'$, as required. $\square$

We are now ready to prove that safety for bisimulation and commutation coincide:

Theorem 22. *BiS commutes with O iff O is safe for bisimulation.*

Proof. From right to left. Consider any two bisimilar pointed structures $\mathcal{M} + \overrightarrow{K}, w$ and $\mathcal{M}' + \overrightarrow{K'}, w'$, where $\overrightarrow{K}$ and $\overrightarrow{K}$ are some sets and relations over $|\mathcal{M}|$ and $|\mathcal{M}'|$ (respectively) which expand two given structures $\mathcal{M}$ and $\mathcal{M}'$. Assume that (a) $w O_{|\mathcal{M}|}(\overrightarrow{K}) v$ for some $v \in |\mathcal{M}|$. We need to find an object $v' \in |\mathcal{M}'|$ such that $w' O_{|\mathcal{M}'|}(\overrightarrow{K'}) v'$, with the structures $\mathcal{M} + \overrightarrow{K}, v$ and $\mathcal{M}' + \overrightarrow{K'}, v'$ still being bisimilar. Bisimilarity is preserved when some parts of structures are left out, so there is a bisimulation Z between $\langle |\mathcal{M}|, \overrightarrow{K}, w \rangle$ and $\langle |\mathcal{M}'|, \overrightarrow{K'}, w' \rangle$. By safety for bisimulation and (a), there is a $v' \in |\mathcal{M}'|$ such that $w' O_{|\mathcal{M}'|}(\overrightarrow{K'}) v'$ and $v Z v'$. So Z itself is a bisimulation between $\mathcal{M} + \overrightarrow{K}, v$ and $\mathcal{M}' + \overrightarrow{K'}, v'$.

From left to right. Our difficulty here is that, if two structures have 'some' bisimulation, commutation with moves along O again gives us the existence of *some* bisimulation. But safety wants the *same* one. Our solution is a trick: we define a new predicate in terms of reachability using the given bisimulation. Since that predicate will be preserved by the new bisimulation, the 'commutation world' given by the new bisimulation must be a commutation world for the earlier one as well.

So let O be a dynamic operation commuting with BiS. Consider two models $\mathcal{M}$ and $\mathcal{M}'$ with Z a bisimulation between them, and three worlds $w, v \in |\mathcal{M}|$ and $w' \in |\mathcal{M}'|$ with $w Z w'$ and $w O(\mathcal{M}) v$. By Lemma 21, we can take Z to be quasi-transitive. We need to find a $v' \in |\mathcal{M}'|$ such that vZv' and $w'O(\mathcal{M}')v'$. We define a predicate $P_{\mathcal{M}}$ on $\mathcal{M}$ which holds exactly at the points which are Z-connected to v. Similarly we define another predicate $P_{\mathcal{M}'}$ on $\mathcal{M}'$ by $P_{\mathcal{M}'} = \{u' \in |\mathcal{M}'| \;/\; u' \text{ is } Z \text{ connected to } v\}$. Z is still a bisimulation between $\mathcal{M} + P_{\mathcal{M}}$ and $\mathcal{M}' + P_{\mathcal{M}'}$. By commutation with O, there is a $v' \in |\mathcal{M}'|$ and a bisimulation Z^+ between $\mathcal{M} + P_{\mathcal{M}}$ and $\mathcal{M}' + P_{\mathcal{M}'}$ such that $w'O(\mathcal{M}')v'$ and vZ^+v'. Since Z^+ is a bisimulation, it respects atomic predicates. Hence $v \in P_{\mathcal{M}}$ implies $v' \in P_{\mathcal{M}'}$: that is, v' is Z-connected to v. But Z is quasi-transitive. So vZv', as required. □

Putting Theorems 19 and 22 together, we get that a abstract dynamic operation preserves bisimulation invariance iff it is safe for bisimulation. This shows that safety, too, is a correct match for bisimulation invariance. Van Benthem 1996 only showed its sufficiency, our new results also showthe necessity of this condition. [18]

4.3 Beyond safety: definability theorems

Safety essentially says that the values of dynamic operations get 'a free ride' on bisimulations: if $\mathcal{M} \overset{Z}{\leftrightarrows} \mathcal{M}'$, then $\mathcal{M} + O(\mathcal{M}) \overset{Z}{\leftrightarrows} \mathcal{M}' + O(\mathcal{M}')$. This idea makes sense in other settings too. In the case of first-order logic, let T be a first-order theory in a language $L + P$. We say that P is *free for isomorphism* under T iff, for all models $\mathcal{M}$, $\mathcal{M}'$ of T, an L-isomorphism from $\mathcal{M} - P$ to $\mathcal{M}' - P$ is also an $L + P$-isomorphism from $\mathcal{M}$ to $\mathcal{M}'$ (where $\mathcal{M} - P$ is just the reduct of $\mathcal{M}$ to L). Beth's well-known Definability Theorem can be viewed as a result about such 'free' predicates:[19]

Theorem 23. *In first-order logic, P is free for isomorphisms under T iff T explicitly defines P.*

Proof. The right to left direction is immediate: expansion with definable predicates does not break an isomorphism. From left to right, consider the special case of two models of T, $\mathcal{M}$ and $\mathcal{M}'$, such that $\mathcal{M} - P = \mathcal{M}' - P$. The identity is an L-isomorphism between them. Therefore, since P is free, the identity is an $L + P$-isomorphism between $\mathcal{M}$ and $\mathcal{M}'$, hence $P_{\mathcal{M}} = P_{\mathcal{M}'}$. That is to say, T implicitly defines P, and therefore, by Beth theorem, it explicitly defines P as well. □

This raises many further questions. Can we connect modal safety and explicit first-order definability via Beth's Theorem? What connections exist between being free for some notion of similarity and being definable from its invariants?

[18] It may be of interest to compare our proof with that for the Safety Theorem in detail.

[19] For a standard statement and proof of Beth Theorem, including fully explicit definitions for 'implicit definability' and 'explicit definability', see e.g. [4] p. 265*sqq*.

5 Further prospects

This paper proposes pursuing a sort of 'reverse' meta-logic. Instead of characterizing a logical language in terms of general semantic properties, including invariance, we start with basic logical operations (first-order quantification, modal or dynamic operators) and identify semantic properties of invariance which match these operations best. Our main tool in doing so is the general equivalence between commutation and invariance in Theorem 12. This suggests a line of research into the duality between the 'semantic' level of structures and notions of similarity, and the 'syntactic' level of operations and notions of invariance for those operations. In particular, in future work, we intend to look at *fixed-point logics* extending ML or FOL to get a closer correspondence between definable operations and invariant ones.

In the same spirit, we think it worth investigating *generalized Lindström theorems* where the structural invariance relation is no longer given beforehand as is usually done, but considered as a parameter to be freely chosen, just as the set of sentences and the truth relation of the abstract logic.

As for first-order logic versus modal logic, our results in Sections 1 and 2 highlight some new parallels beyond the many that are already known. Even so, one would like to see a still more general 'transfer theory' between first-order languages with potential isomorphisms and modal languages with bisimulations, probably based on generalized assignment models. Are there still more general *uniform translations* between the two realms which have eluded us so far?

Finally, we see the great generality of our approach as a benefit also when it comes to more concrete systems beyond modal and first-order logic. One follow-up project, continuing the interest in dynamic logic in this paper, will be the study of the space of natural logical operations on *games*, both sequential and parallel, and the structural similarity relations appropriate to that much broader area.

References

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

[2] D. Bonnay. *Qu'est-ce qu'une constante logique ?* University Paris I, PhD Dissertation, 2006.

[3] D. Bonnay. Logicality and invariance. *Bulletin of Symbolic Logic*, 14(6):29–68, 2008.

[4] G.S. Boolos, Burgess J.P., and R.C. Jeffrey. *Computability and Logic*. Cambridge University Press, fourth edition, 2002.

[5] E. Casanovas. Logical operations and invariance. *Journal of Philosophical Logic*, 36(1):33–60, 2007.

[6] S. Feferman. Logic, logics, and logicism. *Notre Dame Journal of Formal Logic*, 40(1):31–54, 1999.

[7] M. Hollenberg. *Logic and Bisimulation*. Publications of the Zeno Institute of Philosophy, 1998.

[8] J. van Benthem. *Exploring Logical Dynamics*. Studies in Logic, Language and Information, CSLI Publications & Cambridge University Press, 1996.

[9] J. van Benthem. Logical constants, the variable fortunes of an elusive notion. In W. Sieg, R. Sommer, and C. Talcott, editors, *Reflections on the Foundations of Mathematics: Essays in Honor of Solomon Feferman*, pages 420–440. AK Peters, Ltd., 2002.

Johan van Benthem
ILLC, University of Amsterdam
Plantage Muidergracht 24
1018 TV AMSTERDAM (Netherlands)
& Stanford University
Department of Philosophy
STANFORD, California 94305 (USA)
johan@science.uva.nl,

Denis Bonnay
Département d'Études Cognitives & IHPST
École Normale Supérieure
29, rue d'Ulm
75005 PARIS (France)
denis.bonnay@ens.fr